DOCUMENT RESUME

ED 443 821 TM 030 698

TITLE	Research in the Schools. Volume 6.
INSTITUTION	Mid-South Educational Research Association, MS.
PUB DATE	1999-00-00
NOTE	168p.; The individual articles are available in the "Current Index of Journals in Education."
PUB TYPE	Collected Works - Serials (022)
JOURNAL CIT	Research in the Schools; v6 n1-2 Spr-Fall 1999
EDRS PRICE	MF01/PC07 Plus Postage.
DESCRIPTORS	*Academic Achievement; *Educational Research; Educational Technology; *Elementary Secondary Education; Factor Analysis; *Higher Education; Item Response Theory; *Professional Development; Teaching Methods
IDENTIFIERS	Rasch Model

ABSTRACT

This journal publishes articles that focus on the results of applied educational research, scholarly reviews of research, descriptions of classroom techniques, studies of assessments, and other topics of interest to educational researchers. Issue number 1 of Volume 6 contains these articles: (1) "'More than Having a Vision': The Emergence of Teacher Leadership in a PDS" (Lisa DeMeulle); (2) "Preparing Leaders for Tomorrow's Schools: The Practitioners' Perspectives" (Frances K. Kochan and William A. Spencer); (3) "Schools without Principals: Do Both Management and Leadership Exist?" (Theodore B. Creighton); (4) "Theoretical Perspectives of Instructional Supervision in Contemporary Textbooks: A Content Analysis" (Barbara L. White and Larry G. Daniel); (5) "University Teachers' Perceived Usefulness of Computer-Assisted Instruction" (Yukiko Inoue); (6) "Math and Reading Tests: Dissimilar Scores Provided by Similar Tests for African-American Students" (Kevin Wickes and John R. Slate); (7) "Screening for Behavior Problems in Preschool: A Comparison of the Social Skills Rating System (SSRS) and a Structured Teacher Interview" (Ronald K. Bramblett, Kim B. Dielmann, and Maleah Smithson); and (8) "A Structural Model of Student Career Aspiration and Science Education" (Jianjun Wang). Issue 2 contains: (1) "Examining Professional Development of Principals as Adult Learners" (Diane Ricciardi); (2) "'Drafted into the Front Lines": Teacher Efficacy during School Desegregation in Columbus, Georgia, 1968-1975" (Virginia E. Causey); (3) "The Role of Technology Specialists: Case Studies of Change Agents" (Kenneth M. Proctor and Martha J. Livingston); (4) "Transforming Graduate Teacher Education Classes: Lessons from Educational Psychology" (Carol VanZile-Tamsen); (5) "Breaking the Rote Memorization Mindset of Preservice Teachers Standards-Based Instruction: An Integrated Preservice Teacher Education Model" (Regina Halpin); (6) "Evaluation of Teacher Preparation Graduates by NCATE Accredited Institutions: Techniques Used and Barriers" (J. Jackson Barnette and Keith Gorham); (7) "Person-Fit and the Rasch Model: How Seriously Model Fit Is Affected by Appropriateness Measurements in the Rasch Model" (Necati Engec); and (8) "An Introductory Primer on the Appropriate Use of Exploratory and Confirmatory Factor Analysis" (Kevin M. Kieffer). (SLD)

Research in the Schools, 1999.

Edited by James E. McLean and Alan S. Kaufman

RESEARCH IN THE SCHOOLS

A nationally refereed journal sponsored by the
Mid-South Educational Research Association
and the University of Alabama at Birmingham.

Volume 6, Number 1 **Spring 1999**

James E. McLean and Alan S. Kaufman, Editors

RESEARCH IN THE SCHOOLS

Information for Authors

Statement of Purpose

RESEARCH IN THE SCHOOLS (ISSN 1085-5300) publishes original contributions in the following areas: 1) *Research in Practice*--empirical studies focusing on the results of applied educational research including cross-cultural studies, 2) *Topical Articles*--scholarly reviews of research, perspectives on the use of research findings, theoretical articles, and related articles, 3) *Methods and Techniques*--descriptions of technology applications in the classroom, descriptions of innovative teaching strategies in research/measurement/statistics, evaluations of teaching methods, and similar articles of interest to instructors of research-oriented courses, 4) *Assessment*--empirical studies of norm-referenced, criterion-referenced, and informal tests in the areas of cognitive ability, academic achievement, personality, vocational interests, neuropsychological functioning, and the like, and 5) *Other* topics of interest to educational researchers. *RESEARCH IN THE SCHOOLS* is devoted to research conducted in *any* educational setting from a conventional elementary school or high school to a training program conducted within an industry. Likewise, there are no age restrictions on the sample, since the educational settings may include preschools, continuing education classes for adults, or adaptive skills courses in nursing homes. Studies conducted in settings such as clinics, hospitals, or prisons are ordinarily inappropriate for *RESEARCH IN THE SCHOOLS* unless they involve an educational program within such a setting. One goal of *RESEARCH IN THE SCHOOLS* is to provide a training ground for graduate students to learn effective reviewing techniques. Consequently, the journal utilizes a Graduate Student Editorial Board composed mostly of students in educational psychology and educational research. Members of this Editorial Board, each sponsored by a professor, provide supplementary reviews for a selection of submitted articles, and receive both direct and indirect feedback of the quality of these reviews.

Preparing Manuscripts

Authors should prepare manuscripts in accordance with the stylistic rules and guidelines delineated in the *Publications Manual of the American Psychological Association* (4th ed., 1994), which is available from: Order Department, American Psychological Association, PO Box 2710, Hyattsville, MD 20784. Number the pages consecutively. All manuscripts will be subject to editing for sexist language.

Author Identification

Authors should put the complete title of the article on the first text page, but they should exclude their names. Subsequent pages should include only a running head. They should prepare a separate sheet with the complete title of the article and their names and affiliations; this procedure will ensure anonymity in the review process. Authors should supply addresses and phone numbers, and electronic mail addresses and fax numbers (if available), for potential use by the editorial staff and, later, by the production staff. Unless otherwise stated, the first-named author will be sent correspondence, galley proofs, copyright forms, and so forth.

Submission of Manuscripts

Submit manuscripts in triplicate to **James E. McLean, Co-Editor, RESEARCH IN THE SCHOOLS, School of Education, 233 Educ. Bldg., The University of Alabama at Birmingham, 901 13th Street, South, Birmingham, AL 35294-1250. Please direct questions to jmclean@uab.edu.** All copies should be clear and readable; dot matrix is acceptable only if it meets these qualities of legibility. Length of the manuscripts, including references and tables, should ordinarily range from about 10 to 40 typed, double-spaced, 8-1/2 X 11-inch pages, using 11-12 point type. Abstracts are limited to 125 words. Brief reports of research are not encouraged. Authors are encouraged to keep a hard copy of the manuscript to guard against loss. It is assumed that all manuscripts submitted for publication are original material and have not been simultaneously submitted for publication elsewhere. When manuscripts are accepted for publication, authors are encouraged to submit the final version on a computer disk along with the hard copy.

Copyright and Permissions

Volume 6, Number 1 **Spring 1999**

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 1, 1-8

"More Than Having a Vision:" The Emergence of Teacher Leadership in a PDS

Lisa DeMeulle
The University of Memphis

As more schools and universities engage in developing collaborative partnerships, many teachers are being required to assume non-traditional roles such as facilitator, school community member, and leader. However, many teachers are uncomfortable with these roles, and those involved in school/university partnerships need frames of understanding that can help guide teacher leaders. One frame for understanding for the promotion of teacher leadership is the development of shared visions in professional development schools (PDS). This ethnographic case study explored the aspects of implementing an elementary professional development school over a one-year period that promoted teacher leadership. Findings report (a) the visions of the teachers in this PDS, (b) the processes that impacted the sharing of teachers' visions, and (c) the PDS efforts that encouraged and/or discouraged teachers to explicate and realize their visions.

For those involved in the world of educational change and reform, life can often be exciting, challenging, and messy. This can be especially true for teachers, because many are now being asked to serve in a variety of non-traditional roles. The traditional image of teachers as isolated disseminators of knowledge is now being replaced by visions of teachers as facilitators, collaborators, and school leaders (Darling-Hammond, 1993). However, many teachers are quite unprepared for these types of roles. One recent movement that assists teachers in receiving the type of support needed to begin to reconceptualize the role of teacher as leader has been the implementation of professional development schools (Holmes Group, 1990). This ethnographic case study explored the various aspects of implementing an elementary professional development school (PDS) that promoted teacher leadership.

Teacher Leadership in an Era of Reform

Definitions and theories of school leadership abound. While it is not a goal of this paper to provide a comprehensive overview of leadership theory (see Bolman &Deal, 1991 or Sergiovanni, 1992), it is necessary to create a theoretical frame to establish a context for this work. More recent reform efforts call for the increase of teachers' voices in creating the culture of the school and in decision making processes. Here teachers work as integral community members and leaders. Although teacher leadership may manifest itself in many different forms, Bolman and Deal (1994) discuss this type of leadership as ". . . more than having a vision, it is having a vision that can be shared with others" (p.1). To explore teacher leadership as shared vision, the following questions served as focusing tools in data collection and analysis.

1. What PDS efforts encouraged and/or discouraged teachers to explicate and realize their visions?
2. What were the visions of the teachers in this PDS?
3. What processes impacted the sharing of teachers' visions?

By examining the second wave of reform currently manifesting itself in policy, practice, and theory, we can better understand the processes of educational change and the further significance of teacher leadership.

The first wave of techno-rationale reform attempted to cure educational ills in the 1970s by running schools as bureaucracies. In the 1980s, we saw the emergence of a second wave of reform based on collaborative and critical citizenships. Here teaching and learning are viewed as processes of knowledge construction, and proponents work to promote a more holistic and democratic approach to education. They see the need for students to become critical of their worlds by having an awareness of self and others and to think in conceptual and problem solving terms. Schools become communities of learners and require teachers to work in partnership with parents, businesses, social agencies and universities (Cunningham & Gresso, 1993; Darling-Hammond, 1993).

At issue in this reform effort is who holds the power to make educational decisions (Sarason, 1990). Advocates believe that true systemic change is created through

Lisa DeMeulle, Department of Instruction and Curriculum Leadership. This study is supported in part by grants from the Center for Research on Educational Policy, The University of Memphis. The author gratefully acknowledges the faculty of this school for their input and feedback on this article. Correspondence concerning this article should be addressed to Lisa DeMeulle, who is now at Sonoma State University, Department of Education, 1801 E. Cotati Ave., Rohnert Park, California 94928. E-mail may be sent to lisa.demeulle@sonoma.edu.

critical citizenships, where the voices of those impacted the most guide the decision making process in a democratic approach (Giroux, 1992; Sarason, 1990). Teachers are at the forefront of this reform effort by having a high value placed on their knowledge and expertise. They are educational leaders who must coordinate a complex environment that supports the learning process. Such change seeks to alter the fundamental ways in which schools operate (Fullan, 1991) and requires teachers not only to assume a different mind-set toward teaching, but also new roles and relationships. We need more information about the processes that assist teachers in assuming and feeling comfortable with these changes.

Fullan (1991) informs us that most educational reforms fail because the participants lack a common understanding of the purposes of educational change and how they might best proceed. How those involved make meaning of change at a personal and sociopolitical level becomes paramount to supporting change. They advocate, "Solutions must come through the development of *shared meaning*. The interface between individual and collective meaning and action in everyday situations is where change stands or falls" (p. 5). When universities and schools enter into partnerships, it is often with different agendas. We know very little about how to arrive at a place of common goals and understandings. Leiberman (1993) prompts us to begin questioning whose world views are being promoted. She calls for developing frames of understanding that "give voice to the inner and observed lives of teachers and schools as partners rather than solely as objects of study" (p. 9).

Professional Development Schools

It is the idea of teacher leadership as a means of promoting collaborative change in an elementary professional development school that is the concern of this study. Advocated by the Holmes Group (1990) as a way to rethink the nature of schools and schooling, PDSs in their most basic nature are collaborative partnerships between a university and a school. These partnerships take a variety of forms, but all have a common focus on supporting inquiry and the professional development of the involved participants (Stallings & Kowalski, 1991). The PDS concept is a revolutionary means of school reform because it requires school and university faculty to assume new roles and relationships. This shift ultimately changes the culture of schools and universities as well as those who live within them (Leiberman, 1990; Soltis, 1987; Zeichner, 1992).

The University of Memphis began its PDS program in 1991 and is presently involved in twelve PDS partnerships. Each PDS collaboration consists of one or two faculty members who are assigned to work with a specific school as liaisons for the purposes of supervising student teachers and engaging in professional and school development with experienced teachers and administrators. Activities vary from site to site based on the needs of the partnership, but all partnerships are grounded in the six principles of the Holmes Group (1990):

1. Teaching and learning for understanding.
2. Creating a learning community.
3. Teaching and learning for understanding everybody's children.
4. Continuing learning by teachers, teacher educators, and administrators.
5. Thoughtful long-term inquiry into teaching and learning.
6. Inventing a new institution. (p. 7)

Mann Elementary, one of these schools, is the focus of this study.

Mann Elementary School. Mann Elementary School[1] is located in a middle-class, suburban neighborhood of a metropolitan area in the mid-south. It is one of the largest elementary schools in the state and the largest elementary school in this district with over 80 full-time faculty and 1,500 students in grades K-5. Of the four administrators in the school, three were female, including the principal. Seventy-nine teachers were female. The student population is 35% African-American, 64% European-American, and 1% other. During spring 1993, Mann joined the PDS program at The University of Memphis when two-thirds of the faculty voted to participate. The partnership began in the following 1993-94 school year. I served as the PDS liaison between the university and the school. My duties included: supervising student teachers, supporting the professional development of cooperating teachers, working with the entire faculty to create and implement a school development plan, and teaching a graduate course at the school site.

The school year under study began with several PDS-related activities. A group of five student teachers was at the school during the Fall semester. The student teachers and I met weekly (15 total) in a group. After these meetings, the cooperating teachers and I then met in a group. These meetings were open forums and provided both groups the opportunity to explore what it meant to learn and teach in a PDS. Beginning in September, the faculty began to write their school development plan. This took place over eight different meetings and covered a 10 week period.

In the spring semester, a new group of five student teachers started, and we continued the weekly meetings. Teachers had a one day inservice in January, and a focal session was "The Role of the Clinical Professor in the PDS." In this session, I provided definitions of PDS as

[1]All names and identifying information have been changed to protect confidentiality.

given in the literature and then facilitated dialogue about possible implications for the faculty at the school. Additionally, a graduate course entitled "Teacher Leadership in the PDS" was offered to teachers at the school site during this semester. Eleven teachers enrolled in the course, and projects ranged from grant writing to developing an art curriculum designed to involve parents. The end of the year wrapped up with a whole faculty meeting. Each committee presented their accomplishments related to the school development plan. I also shared a summary of feedback from an end-of-the-year survey that school faculty had completed a week earlier, and the teachers voted to reaffirm their commitment to the PDS program.

Method

This study employed an ethnographic case study methodology. Case study method is appropriate to answer the goal of this study. It is used "when investigators desire to (a) define topics broadly and not narrowly, (b) cover contextual conditions and not just the phenomenon of study, and (c) rely on multiple and not singular sources of evidence" (Yin, 1993, p. 1). Because I was involved in the daily lives of the participants and trying to understand the meaning of everyday school activities, I assumed the role of ethnographer (Hammersley & Atkinson, 1983).

Data Collection

Data collection began in August and continued through June of the following year. To address the goal of the study, a variety of empirical materials were collected. This diversity of data collection procedures and resulting materials allowed the voice of teachers and school faculty to be represented in multiple ways. I thereby obtained a more complete picture of their thoughts, values, and understandings. Empirical materials included: (a) field notes and a personal journal, (b) audio recordings and transcripts, (c) video tapes of whole group faculty meetings, (d) various site documents, and (e) an end-of-the-year survey of all school faculty (see Figure 1).

Because I served as both researcher and university liaison at Mann, I was the participant observer in the setting (Hammersley & Atkinson, 1983). As a part of the participant observation process, I maintained a research journal with theoretical, personal, and methodology notes (Spradley, 1979).

Semi-structured interviews lasting between 1-2 hours were held with key school faculty at the beginning and end of the school year. I asked questions such as: "What are you hoping will occur as a result of becoming a PDS?" "What is your reaction to the activities we have had to date?" "What was your most successful PDS experience?" "Why?" With the exception of two interviews, all interviews were audio-taped and transcribed.

Data Collection Procedures	Time Period of Data Collection	Resulting Empirical Materials
• Participant observation	Throughout the school year	Field notes Personal journal
• Semi-structured interviews with 8 faculty	October and May	9 audio-recordings and transcripts
• Semi-structured interviews between researcher and research assistant	October, January, February, April	6 audio-recordings and transcripts
• Video tapes of school development plan group meetings	September, October, November, May	6 video tapes
• Communication with faculty	Throughout the year	Various site documents, memos and letters
• End of the year survey given to school faculty	May	58 completed questionnaires (54 teachers, 3 student teachers, 1 unidentified)

Figure 1. Data collection procedures, time periods, and the data set.

Videotapes were made of all whole group faculty planning sessions, and copies were kept of materials and communications used in program development. Finally, an open-ended questionnaire was given to all school faculty at the end of the year to provide participants an opportunity to reflect on and share their understandings of life in a first-year PDS. The questionnaire included questions such as, "What are the long-term goals of the PDS collaboration?" "What efforts have you personally participated in this year?" and "In what ways did these efforts help the school move closer to achieving its long term goals?"

Data Analysis

Using the constant comparative method (Glaser & Strauss, 1967), I began analysis by looking at the interview transcripts and research journal to inductively generate initial categories of understanding and identify common themes. Some initial categories included mistrust, dialogue, teachers' voices, visions, and intrinsic motivation. Examples of themes included: developing a knowledge base, vehicles to voice opinion and create consensus, on-going support of leadership, and nurturing trust. These categories and themes were developed and refined by continually returning to the data. To support the findings, an effort was placed on triangulating data (Yin, 1993) through the analysis of multiple empirical materials. The interview transcripts and the research journal proved to be the major sources of data for this study.

Findings

The findings are presented in a manner that they (a) chronologically describe PDS efforts that promoted teacher leadership, (b) describe individual and collective visions of the teachers, (c) describe process issues that impacted the sharing of visions, and in keeping with the theoretical frame, (d) where possible, use the multiple voices of participants to substantiate claims. However, it must be noted here that teacher leadership cannot be examined in a vacuum. There were others in the school who provided forms of leadership that influenced the teachers' actions. These people included the principal, three curriculum coordinators, and me. These efforts will be included only as appropriate to understand the development of teacher leadership.

The First Semester: Early Efforts & Visions

During the first semester, there were two major PDS efforts that provided teachers an opportunity to reflect on and articulate their implicit visions: a) writing the school development plan, and b) weekly meetings between the five cooperating teachers and me. By writing the school development plan, the teachers were able to articulate a shared vision in the form of a mission statement. Several teachers had very specific visions for themselves or the school, while the majority of teachers contributed ideas or were willing to follow the lead of others.

The school development plan. We began writing the school development plan in September. This resulted in a 17 page document that was given to all faculty and stakeholders on December 1. It was the process of writing the plan that provided us the opportunity to begin articulating our visions of the school and our visions of ourselves as educators. My role was to lead these weekly consensus building sessions with faculty working in one of seven committees (two parents participated throughout this time). This work culminated in a mission statement that was supported with many goals and activities. The mission statement developed by all the faculty read:

> The purpose of Mann Elementary is to advocate the academic, physical, emotional, and social development of each child; thus preparing him to become a responsible citizen. This should be a cooperative effort which involves administrators, teachers, students, parents, and members of the community.

An additional vision that was shared by most of the faculty at this point was the desire to improve communication and working relationships among all participants involved in the PDS including students, parents, administrators, teachers, and university personnel. Karla, a kindergarten teacher, described this desire:

> I would like to see that we would work better together, administration and teachers and parents and community. I know that it can't be a utopia, even though we wish it could be. But we can strive to work toward a better education for the children, better working together, and not feeling all this pressure.

There were so many suggested actions for working toward this particular vision that a new committee was eventually created: The School/Community Network. Claudia, another kindergarten teacher, shared her response:

> I'm really excited about that committee, though. I was glad to hear that they were open to starting a new committee, to actually deal with all of those things rather than just skipping them, or 'you folks take care of one, you guys do one, you do one.' I appreciate that. That makes me

> feel good to know that our opinions are going to hold some weight.

Several teachers had very explicit individual visions in support of these two larger shared visions. Martha, the art teacher, wanted to revitalize the Picture Person Program at the school. This project eventually impacted every student and teacher in the school. Doris began a school beautification project. Laura and Marilyn investigated grant writing to support their efforts to build a new computer lab. These projects were carried out over the school year and are discussed in more detail later in the paper.

Weekly meetings with cooperating teachers. Part of the PDS design provided for weekly meetings between the cooperating teachers and me. Because the PDS program was designed with an open and flexible framework in mind, we spent much of the first semester coming to common understandings about what was required of student teachers and how to best meet their needs. There was a great deal of ambiguity during this period, and we were all negotiating our roles and relationships. A strong example of this ambiguity in roles was made evident in the third week. I was again asking the cooperating teachers for their ideas and input, and Billie, a first grade teacher, said, "Are you asking us these questions because you don't know or are you thinking we should decide together?" I again explained that this PDS was founded on the notion of collaborative decision making and that, indeed, one of the reasons we met weekly was to "decide together."

Although this first group of cooperating teachers never explicated a shared vision for student teachers, Laura, a first grade teacher, had the idea of creating a student teacher/new teacher handbook to help deal with some of the confusion we experienced during this first semester. Working the following semester with three of the teachers in the group, this handbook was completed to fulfill a requirement in the Teacher Leadership course.

Early process issues. It is important to note here that three issues impacted the process of developing shared visions during the first semester. First, many school faculty were mistrusting of university personnel and other teachers and administrators in the school. When interviewing Karla, she tried to warn me about mistrust among the teachers, "I know that there does sometimes seem to be, I want to say factions, because I'm going back to that hierarchy problem again. . . . And so just be aware of that." Paula, a fifth grade teacher, shared her thoughts regarding the principal, "But she does use intimidation, and she does like to silence something, and she likes to be in control, which is not an unusual characteristic of a principal . . . she has a good list keeping you on her side." These comments are reflective of the attitudes that many faculty held about the context of mistrust in their school.

Second, as the university liaison, I experienced a great deal of anxiety, excitement, and frustration while leading the school development plan sessions. In the initial meetings, many teachers seemed hesitant to get involved and several expressed their resentment at having to attend. In one incident, I spoke to the principal and asked her if she would begin the next meeting by offering a statement of support to the teachers. A segment from my journal described what followed:

> The principal told the faculty, "Lisa said that some of you resent her being here. We've committed to this program, and you *will* learn to work as a collaborative team. Dr. DeMeulle." Then she passed me the microphone. I tried to clarify the statement and felt very awkward.

The anxiety and frustration I experienced during this semester began to subside as the faculty and I developed rapport and more intimate relationships.

Finally, it must be noted that an overwhelming process issue for the faculty was the extra time required of them. On the end-of-the-year survey, over 50% of the teachers reported that "time" was the biggest drawback in their collaborative work. Many teachers were resentful of the extra hours that were required.

The Second Semester: Learning to Lead

The second semester saw a shift in understandings and attitudes about the PDS program. Major efforts this semester included a graduate course taught at the school entitled "Teacher Leadership in the PDS," teachers working independently and in their committees to complete their specified goals, and continued weekly meetings with five new cooperating teachers.

Teacher leadership course. Eleven teachers enrolled in the course which lasted a total of 13 weeks. The teachers had to pay regular graduate fees and received graduate credit through the University of Memphis. All meetings were held at the school, either in the library or individual classrooms. The two requirements for the course were to document 51 hours of school improvement time and to complete a written project. The written project was to have a "real-life" purpose that was based both on the teachers' individual interests and the school's needs which had been previously outlined in a school development plan. Finally, the project was not to be a term paper. Teachers were assisted in getting started on their projects through a variety of approaches and strategies. These approaches included: providing teachers with written guidelines that offered step-by-step assistance in

defining a problem, researching, and reporting; holding two initial group meetings where participants shared their ideas and got input from other teachers; and providing support and assistance to teachers by having on-going meetings throughout the remainder of the course with individual project teams. A final whole group meeting was held for teachers to share their completed projects.

At the beginning of the course, five of the teachers in the course were unclear about their visions for their school. They knew they wanted "a better school," and were willing to work with others, but they were unsure about where they were going or how to get started. This group expressed a great deal of anxiety about their projects. The remaining teachers had very specific ideas about projects they wanted to complete that would move their school toward meeting its long-term objectives. These teachers were excited and enthusiastic from the first class. The class projects included exploring the implementation of a whole language reading program in a kindergarten classroom, a pilot project to revitalize the "picture person" program (an art program that uses parents as teachers of fine art), writing the new teacher/student teacher handbook, two different grant writing teams (one for a computer lab, and one for the picture person program), and a teacher who documented her work with a high needs student.

Although all the teachers put in a great deal of effort during the course, those who had a vision from the beginning completed the projects in much more detail and with greater success. For example, nine additional teachers committed to join Martha's continuing efforts revitalizing the picture person program by enrolling for summer course credit through independent study. Her program went school-wide in the fall of 1994. Laura, who worked on the teacher handbook, commented on the success of the course from her perspective:

> It just made us sit down and do it. You know we think about all these things, we need to have a new teacher [hand]book. We need to give them [student teachers and new teachers] something when they first come in . . . We really sat down and did it . . . this is great. And as you're doing it, it brings out other ideas . . . you really can improve your school.

In contrast, one of the grant writing teams dropped their project at the end of the semester. One of the team members commented on her experience, "I appreciated the assistance of [university] personnel, but on more than one occasion, my lack of experience in grant writing seemed misunderstood. I felt several times that my questions were perceived as a nuisance."

Teachers' committee work. Committees were to continue working throughout the year, meeting as suggested by each committee chair. Teachers' impressions of their committees ranged from, "We didn't meet, I mean it was all individual" to "Our committee worked so wonderfully together."

Because I did not meet with the committees, I gathered the committee chairs together at the end of the year for a feedback session. Over half of the chairs reported that it would have been beneficial to meet as a group during the year. Several were uncomfortable with a leadership role, and these committees ended up either not meeting or being overrun by a senior faculty member. Denise, another first grade teacher, served on one of these lost committees and stated, "I don't know who the chair is on my committee. Somebody's got to be in charge, and I don't really know them [other teachers]." In contrast, those committees that were facilitated on an on-going basis accomplished their goals and discussed their efforts as group efforts, not as individual projects. These teachers saw their committees as working as a team toward supporting the school vision. Martha, a second grade teacher, reported:

> Our committee all got along so well. We were all very supportive of each other and really got a lot accomplished. I was very proud to be a part of a team that wants to get things done for the school.

Continued weekly meetings. Weekly meetings continued with a new group of cooperating teachers. Because I was more familiar with the routine, this second group accomplished more in terms of trying to articulate a vision for student teachers. A key incident occurred the day I suggested that we rewrite the observation form being used by the student teachers. At first, the teachers seemed little interested. I then asked, "Well, if someone came in to see you teach, what kinds of things do you think they should be looking for? What does good teaching look like?" At the prompting of these questions, the group became very engrossed in dialogue. The final result was an observation form that listed 25 questions from which student teachers could choose six to answer. Once we had articulated this "vision" of good teaching, the rest of the semester was spent examining policies and course requirements to determine how we might best work with these to support our vision of good teacher education. For example, we explored the use of peer coaching among the student teachers and cooperating teachers. The need for and appreciation of collaboration as a means to realize a vision was made evident in Betty's statement about working with a student teacher:

> Having a student teacher makes you really evaluate your own teaching, because you think about everything you're doing . . . and it's making you re-explain to yourself why you do these things. The big difference was having you as a liaison and here so much . . . There was more communication, and the three-way communication between student teacher, you and the regular teacher. I mean that needs to be there. So it helps to increase the understanding, that we're all kind of operating from the same understandings.

Process issues. The issues of general mistrust of university faculty that were apparent the first semester started to be overcome as the second semester continued. Contributing factors included: continuing to develop a knowledge base about PDS, respecting the thinking and interests of participants, and assuming commitment and responsibility. Teachers developed their theoretical knowledge by reading articles on PDS and by attending an in-service that provided an opportunity to discuss the six principles of the Holmes Group (1990). Practical knowledge was developed by day-to-day life in PDS. When we compare the definitions of PDS at the beginning and end of the year, evidence of this developing knowledge is apparent. At the beginning of the year, the majority of teachers defined a PDS as "a place to train student teachers." By the end of the year, definitions included: "school and community working together," "a school that looks within itself to find ways to improve education and communication," and "a school teaching students both young and old."

Respect of the teachers' voices was crucial to building trust and promoting shared vision. Betty talked about her understanding of the need for respecting thinking and interest of teachers:

> I think you've made the big difference in attitudes. Everything we did, you had feedback on it, you had really looked at it; and we didn't do some survey and never saw it again, which can happen many times. You always came back, you knew our responses. I guess sitting down and doing that was worthwhile.

Finally, as the year continued, many more teachers began to assume commitment and responsibility beyond their classrooms. One teacher commented, "I enjoyed working with groups of people that are cooperative and positive, ones that realize there is a job to do and get it done." At one point, 40 teachers in the school were required by the principal to interview teacher education candidates. While there was much initial complaining, on the day of the interview the teachers were all dressed very professionally, many in suits, and took the responsibility seriously. Later, Debbie commented to me:

> I enjoyed interviewing the teacher candidates. It was a good thing for our school. You know, we're the ones who should be deciding who becomes a teacher. It's our profession. I think it brought a sense of professionalism to our school.

Perhaps the greatest evidence of participants' assuming commitment was the vote taken at the end of the year--58 voted to continue the involvement in the PDS program, 5 voted no. Based on the majority vote, Mann Elementary and the university made the commitment to continue PDS involvement the following semester.

Discussion and Implications

Many of the efforts used in this initial year of PDS implementation proved to be successful in helping teachers articulate and share their visions with others. Through writing the school development plan, we saw many teachers' visions being articulated for the first time. We then saw the emergence of teacher leadership during the second semester, when teachers are provided different vehicles--such as the teacher leadership course--to begin working toward their visions.

What appeared to be crucial in promoting shared visions were opportunities for structured dialogue and reflection among participants. It was during these times that teachers were able to share their voice, begin to appreciate the voices of others, and begin to work toward the building of a critical democracy (Giroux, 1992). While many have suggested the use of teams to collaboratively effect school change (Cunningham & Gresso, 1993; Maeroff, 1993), findings from this study suggest that (a) each teacher must have individual opportunities to voice opinions while others listen, (b) a shared school vision must be achieved through total consensus of all faculty members, (c) team leaders need continued and ongoing support in leadership of their groups, (d) trust must be nurtured between all participants, and (e) a practical and theoretical knowledge base needs to be developed throughout the year.

Learning to assume leadership, or even understanding the need to do so, was a challenge for many teachers during this year. Sarason (1990) prompts us to question whether all teachers need to be leaders. This study's findings indicate that some teachers were unable to articulate visions, much less begin to share them in a way that would inspire others. However, we know that all

teachers must be leaders within their own classrooms. The PDS implementation caused many of these teachers to begin looking beyond their classrooms for the first time and let them to take responsibility for other aspects of the school based on the needs they saw there. Many teachers had individual visions, some being so strong that they created a wave of action and support within the school. Articulating these individual visions was a beginning step in teacher leadership. So, were all these teachers leaders? No. But given the opportunity to share their views of education, they have all taken the first step.

During this year, we saw many teachers willing to take responsibility and commit themselves for the greater good of the school. In viewing themselves as integral community members, they saw the need to work toward the common good. This idea of moving away from self-interest toward supporting the greater community supports Sergiovanni's (1992) notions of moral and professional leadership. Many teachers reported that their greatest reward was working with other teachers who cared about the school and who were committed. Their motivation came from an internal sense of being responsible for others, rather than from receiving personal rewards. In supporting educational change, we may be able to achieve more by promoting the notion of a caring, democratic community rather than solely focusing on rewarding individual efforts with extrinsic perks. However, teachers must be provided the time and space to engage in these efforts without it coming at the end of an already long school day or the end of a long school year.

All employees are at their best when they feel they have voice in their work (Schein, 1978). For years, teachers have been at the mercy of dictatorial principals, bureaucratic school districts, and findings from research studies. When afforded opportunities that encourage and support their professional development based on their individual understanding of their needs and the school's needs, teachers are empowered to continue improving themselves and their practice. These efforts can only improve education for our children. However, there are no easy solutions or quick fixes when attempting to promote teacher leadership. It is long and complex work fraught with many issues of affect. In defining teacher leadership as the ability to share a vision with others, and using this type of PDS structure as a vehicle to do so, a framework is provided that respects schools and universities as partners, and lets the voices of teachers be heard.

References

Bolman, L. G., & Deal, T. E. (1991). *Reframing organizations: Artistry, choice and leadership.* San Francisco: Jossey-Bass Publishers.

Bolman, L. G., & Deal, T. E. (1994). *Becoming a teacher leader: From isolation to collaboration.* Thousand Oaks, CA: Corwin Press, Inc.

Cunningham, W., & Gresso, D. (1993). *Cultural leadership: The culture of excellence in education.* Boston, MA: Allyn and Bacon.

Darling-Hammond, L. (1993). Reframing the school reform agenda: Developing capacity for school transformation. *Phi Delta Kappan, 74*(10), 752-761.

Fullan, M. G. (with Steigelbauer, S.) (1991). *The new meaning of educational change.* New York: Teachers College Press.

Giroux, H. A. (1992). Educational leadership and the crisis of democratic government. *Educational Researcher, 21*(4), 4-11.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory.* Chicago: Aldine.

Hammersley, M., & Atkinson, P. (1983). *Ethnography: Principles in practice.* New York: Routledge.

Holmes Group. (1990). *Tomorrow's schools: Principles for the design of professional development schools.* East Lansing, MI: Author.

Leiberman, A. (1990). *Schools as collaborative cultures: Creating the future now.* London: Falmer Press.

Leiberman, A. (1993). The meaning of scholarly activity and the building of community. *Equity and Choice, 10*(1), 4-10.

Maeroff, G. I. (1993). *Team building for school change: Equipping teachers for new roles.* New York: Teachers College Press.

Sarason, S. B. (1990). *The predictable failure of educational reform.* San Francisco: Jossey-Bass Publishers.

Schein, E. H. (1978). *Career dynamics: Matching individual and organizational needs.* Reading, MA: Addison-Wesley Publishing Company.

Sergiovanni, T. J. (1992). *Moral leadership.* San Francisco: Jossey-Bass.

Soltis, J. (1987). *Reforming teacher education: The impact of the Holmes Group report.* New York: Teachers College Press.

Spradley, J. P. (1979). *Participant observation.* New York: Holt, Rinehart & Winston.

Stallings, J. A., & Kowalski, T. (1991). Research on professional development schools. In W. R. Houston (Ed.), *Handbook of research on teacher education,* (pp. 251-263). New York: Macmillan.

Yin, R. K. (1993). *Applications of case study research.* Newbury Park, CA: Sage Publications.

Zeichner, K. (1992). Rethinking the practicum in the professional development school partnership. *Journal of Teacher Education, 43*(4), 296-307.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 1, 9-16

Preparing Leaders for Tomorrow's Schools: The Practitioners' Perspectives

Frances K. Kochan and William A. Spencer*
Auburn University

This study was undertaken to gain knowledge about the school principalship in Alabama. A five part survey was used to gather data from 550 principals representing 42% of the principals in the state. A number of issues related to gender and ethnicity were uncovered. The majority of principals plan on retiring within the next ten years. They indicate that the job has become tremendously demanding and complex. Those who fill these positions will need a myriad of competencies to succeed. There will be a need for principal preparation programs and for the State Department of Education in Alabama and in many states throughout the nation to develop programs and strategies to ensure that there will be an adequate supply of quality leaders to head the public schools. Thus the study has implications not only for Alabama but for all those who educate, hire, or train educational leaders.

Overview

The school principal has a powerful impact upon the lives of students and on the success or failure of a school (Bartell, 1994). The role a principal assumes with regard to instructional leadership has a direct effect upon student achievement (Wallace, 1992). The principal's leadership style and management skills impact both the culture and climate of the school (Short & Greer, 1997). The knowledge, skills, and abilities needed to lead quality schools are formidable. Due to the changes in society during the last decade and the expected changes that will occur in the future, the type of competencies required of school leaders are in need of modification (Glickman, 1992; Hanson, 1991; Hoy & Miskel, 1991; Kaiser, 1995). Numerous studies have been conducted to identify principals' perceptions of required competencies to address these changes. When administrators were given the opportunity to rank order the most important skills needed by leaders in the year 2000, they selected interpersonal skills, communication, finance, computers, and leadership (Lester, 1993). When asked to identify the skills for which they needed additional training, principals selected conflict resolution, organizational development, personnel management, consensus building, communication skills, and dealing with the change process (Alexander, 1992). A study of successful school reform leaders identified the development of self-assessment skills as a major goal of these principals (USDOE, 1996). Teachers' responses to characteristics of successful principals were somewhat different than those identified by administrators. Teachers identified the attributes of good school leaders as being honest, competent, forward-looking, inspiring, and caring (Richardson, Flanigan, Lane, & Keaster, 1992). It is obvious that to be successful, school leaders will be required to possess a broad array of complex competencies. Unfortunately there appears to be little thought being given to "helping today's leaders make the transformation to tomorrow's leaders" (Murphy, 1993, p. 30).

Frances K. Kochan is Associate Professor and William A. Spencer is Professor in the Department of Educational Foundations, Leadership, and Technology at Auburn University. Please direct all correspondence to Frances Kochan, 2195 Haley Center, Auburn University, AL 36849 or by e-mail to kochafr@mail.auburn.edu. *This article was co-authored with equal efforts by both authors.

Purpose

This study was undertaken by the Truman Pierce Institute, in the College of Education at Auburn University in partnership with the Alabama State Department of Education. The purposes of the study were threefold. First, it sought to gain an understanding of the present workforce in terms of demographics and preparation. Secondly, it examined principals' perceptions of the job and their estimated competence levels to be successful in it. The third purpose was to gain information regarding the development and restructuring of educational and training programs that prepare principals to be successful leaders.

Methodology

A five-part survey was developed around demographic questions, the state principal competencies, and information from the literature (Bartell, 1994; Nadeau & Leighton, 1996; Valentine, Clark, Irvin, Keefe, & Melton, 1993). The survey was sent to all principals in Alabama. Part I dealt with demographic and professional information. Part II included questions regarding retirement plans and reasons for making a decision to retire. Descriptive statistics were used to analyze most of the data in these sections. Reasons for retirement were counted and placed

in rank order. In Part III, respondents were asked to identify experiences which influenced their leadership style by checking options and adding any not included. Responses were counted numerically and rank ordered. Part IV contained a list of skills identified by the State Department of Education as essential principal competencies. Using a 5 point Likert-type scale, respondents were asked to rank these skills first by importance and then by their own level of competence on each of them. Mean scores were computed for each skill, and they were then rank ordered. The last part of the survey consisted of open-ended questions dealing with how the job had changed, principals' perceived competence, and the skills future administrators would need in order to succeed. Each of two researchers conducted an independent content analysis, organized responses by theme, and then shared their findings with one another. Through discussion and further joint analyses, categories were retained, renamed, developed, or deleted, and a final list was created.

Results

Demographics

Five hundred fifty, or 42%, of the principals responded to the survey. The demographic characteristics of the principals who responded are summarized in Table 1. Eighty-four percent of the principals were white, non-Hispanics, 15% were African-American (1% lower than the national average of 16%), and the remaining 1% were other minorities. Fifty-nine percent of those responding to the question on gender were males and thirty-five percent were females. The percent of female principals in Alabama is slightly higher than the national 1993–94 average (34.5%). Ninety-one percent of these leaders identified the South as their place of geographic origin, although the number born, raised, and/or educated in Alabama was not determined. Almost 90% of the principals are 40 years of age or older. Forty-three percent are 50 years of age or older. The average age is 48.3. This is slightly higher than the last reported (1993–94) national average of 47.7.

Professional Qualifications

The survey data dealing with educational preparation (Table 2) indicate that all but two of the subjects have at least a Master's degree. Three hundred twenty-five (59%) have education beyond that degree. Approximately 10% of principals hold a doctorate and 17% are working towards it. The great majority have AA certification or its equivalent and majored in education as undergraduates. It is interesting to note that the second highest major (11%) was in the "mathematics or science" category. Seventy-six percent of these principals have been in their positions for less than 10 years. Forty-eight percent have served for less than 5 years.

Most principals (47%) came to the position of principal after serving as an assistant principal. Twenty percent of them moved to their present principalship from another principal position. Nineteen percent attained the principalship from a teaching position. Only a few (2%) came from the central office to assume duties as a principal. More than 80% of the respondents became principals in systems in which they were employed. The average time spent in their previous position was 7.4 years.

Table 1
Demographic Characteristics of Alabama Principals

Principals (N = 550)		N	Percent
Gender	Males	325	59.1
	Females	191	35.0
	No Answer	34	6.2
Age	Average = 48.3 years		
Ethnicity	White Non-Hispanic	461	84.0
	African-American	81	14.7
	American Indian	4	0.73
	Puerto Rican	1	0.18
	Asian American	1	0.18
	No Answer	2	0.36
Geographic Origin	South	501	91.0
	Midwest	21	3.8
	Mid-Atlantic	11	2.0
	Southwest	9	1.6
	West	3	0.55
	New England	2	0.36
	Other	2	0.36
	No Reply	1	0.18

Retirement Plans

Retirement seems to be an important issue for principals. Within five years, almost 70% of these principals will be eligible to retire. Being eligible to retire does not, of course, mean that all of these individuals will retire. However, 40% (202) of them responded that they plan to retire within the next five years. Sixty-three percent of these principals plan to retire within the next ten years.

As depicted in Table 3, most principals (65%) cite their primary reason for considering retirement as receiving a better opportunity or promotion. However, 57% of them also identified community politics and general burnout as factors in their retirement decisions, ranking these as reasons two and three respectively. An additional 26% of the respondents would leave to obtain an out-of-state position and a new retirement plan, making this the fourth ranked reason to retire. Frustrations due to finances, external mandates, lack of fulfillment or progress, and family needs were indicated as contributing to retirement actions by 15 to 20% of those responding.

Table 2
Current Professional Characteristics of Alabama Principals

Principals (N = 550)	N	Percent
Highest Education Level Attained		
Credit towards Master's	2	
Master's Degree	217	
Credit toward AA/Sp	10	
AA or Specialist	165	
Credit toward Doctorate	92	
Doctorate	58	
Other	4	
Years in Position Average=6.8 (S.D.=5.9)		
0–4	264	48.0
5–9	151	28.0
10–14	67	12.0
15–19	41	7.0
20–24	17	3.0
25 and above	8	1.0
Highest Level of Certification		
AA or equivalent	324	
A or equivalent	198	
Other	6	
No Answer	22	
Undergraduate Preparation		
Education	358	
Science or Math	58	
Social Sciences	55	
Humanities/Arts	21	
Business	9	
Engineering	1	
Other	18	
No Answer	30	

Table 3
Reasons Given by Principals as Reason for Retiring
(Ranked by Number Choosing It)

Principals (N = 550)	N	Percent
Better opportunity or promotion	358	65.3
Community system/system politics, political stress, and/or conflict in local community	170	31.0
Burn out	142	25.8
Ability to obtain a position out of state and begin a second retirement plan	142	25.8
Frustrations and inability to accomplish my goals	131	24.0
External mandates or requirements from national, state, or other sources	116	21.1
Job is too time consuming	111	20.2
Need more time with family	95	17.3
Financial inadequacies and uncertainties which makes job difficult to perform	92	16.7
Lack of fulfillment in the job	90	16.4
Deteriorating public and community relations	63	11.5
Teachers' organizations (such as the AEA)	12	2.2
Maternity leave	4	0.73
Inadequate preparation during graduate school or certification program	3	0.55

The Principalship Position and Leadership Preparation

Questions dealing with the job of the principal and the preparation of individuals for that role are closely related, therefore the data, the analysis, and the discussion of the findings for these issues are presented together.

Leadership Style and Ranking of Competencies

A major aspect of leadership explored in this study was how administrators developed their leadership style. Respondents were a given a list of possible influences, asked to check all that applied, and to add any others. Numerical results are displayed in Table 4. Principals indicated that mentoring and professional experience with other leaders as well as on-the-job training and graduate education were the most important factors in the adoption of their leadership style. Seventy-four percent of the principals selected professional experiences under another as their most significant influence. Fifty-five percent of the respondents selected the closely related experience of having a mentor. On-the-job experience was ranked as important by 55% of the respondents, and graduate experience/education was selected by 47% of survey participants.

Table 4
Factors Given by Principals as Influencing Leadership Style
(Ranked by Number Choosing It)

Principals (N = 550)	N	Percent
Professional experience under or with other leaders	408	74.2
Mentor(s) or following a pattern set by a previous mentor	304	55.3
On-job training or prior experience in their organizations or institutions	304	55.3
Graduate experience and education	260	47.3
Advice or counsel from family members and peers	117	21.3
Religious convictions	117	21.3
Undergraduate experience in education	56	10.2
Military (habits or training styles that carry over in your work)	40	7.3
Other	30	5.5

Principals were asked to rate their level of competence on the 17 Alabama State Department of Education Skills (ASDE). Using a 5 point Likert-type scale, the mean for 10 of the 17 skills was above 4.0. The lowest mean score, assigned to three skills, was 3.17 (Table 5). Principals selected leadership and planning as their strongest skill areas. They felt less competent when dealing with legal issues, particularly those in special education, handling financial aspects of the job, especially the state's new accounting system, and in understanding the legislative process that impacts schools.

When addressing the importance of the ASDE skills, as displayed in Table 6, principals placed all but two of

the skills at a ranking of 4.0 or higher. The two skills rated of lesser importance were related to politics and funding.

Table 5
Self-Rating of ASDE Skills by Principals (5 = Highest)

Skill	Rating
Evaluates staff according to state and local policies and procedures	4.42
Demonstrates problem-solving skills	4.40
Demonstrates organizational skills	4.36
Takes a leadership role in improving education	4.36
Communicates standards of expected performance	4.36
Improves professional knowledge and skills	4.31
Demonstrates skills in the recruitment, selection, and assignment of school personnel	4.29
Manages instruction	4.21
Implements clear instructional goals and specific achievement objectives for the school	4.16
Establishes clear instructional goals and specific achievement objectives for the school	4.14
Implements evaluation strategies for improvement of instruction	3.93
Understands special education laws and requirements	3.87
Understands the state's education accountability law and requirements	3.81
Understands legislative (political) processes that impact schools	3.68
Understands impact of the New Foundation Program for funding public schools	3.52
Understands the state's education trust fund and reports to board and community on finance issues (proration, etc.)	3.31
Understands the state's new accounting system for education	3.17

Table 6
Principal's Rating of ASDE Skills by Importance (5 = Highest)

Skill	Rating
Establishes clear instructional goals and specific achievement objectives for the school	4.68
Implements clear instructional goals and specific achievement objectives for the school	4.68
Demonstrates organizational skills	4.67
Demonstrates skills in the recruitment, selection, and assignment of school personnel	4.66
Communicates standards of expected performance	4.65
Demonstrates problem-solving skills	4.65
Manages instruction	4.64
Takes a leadership role in improving education	4.61
Improves professional knowledge and skills	4.56
Evaluates staff according to state and local policies and procedures	4.52
Understands special education laws and requirements	4.51
Implements evaluation strategies for improvement of instruction	4.45
Understands the state's education accountability law and requirements	4.25
Understands the state's new accounting system for education	4.10
Understands impact of the New Foundation Program for funding public schools	4.06
Understands legislative (political) processes that impact schools	3.97
Understands the state's education trust fund and reports to board and community on finance issues (proration, etc.)	3.97

Challenges of the Job

In addition to ranking capabilities related to skills identified by the state, principals were asked to respond to three open-ended questions:

1. How has the job of principal changed?
2. What were the three greatest challenges you faced as a starting administrator in your position?
3. What are the five most important skills school leaders will need in the future?

These questions were asked to provide a more in-depth understanding of the position of the principal and the types of skills needed to be successful in this role. Principals' responses about how their job had changed varied widely and addressed many areas. While it is difficult to identify any one area as the most prominent, there were several themes that resounded again and again in their comments. One of the most consistent themes was the increased demands made upon them. Expanded external expectations by local, community, and state sources have increased the complexity of the job. There is also an enhanced focus on legal and educational issues dealing with such things as students with special needs, multicultural diversity, drug and alcohol abuse, and coping with increased violence in society. As one individual noted, the job has become "more legal and more fearful."

Issues related to accountability, dealing not only with student test scores, but also with the evaluation of teachers, expanded documentation requirements, the need to use technological tools, and additional regulations have resulted in a considerable increase in the workload for administrators and schools. Principals often wrote about this using statements such as "we are faced with tighter controls" and "more mandates than it is humanly possible to do."

Expanded external controls are coupled with a drive toward site-based management, which is part of a national trend (Short & Greer, 1997). While principals indicate that this trend gives them more freedom and flexibility, it has also increased time demands made upon them. Principals report they must now spend more time handling management tasks, paperwork, and other non-instructional duties while having less assistance in fulfilling these responsibilities. Principals report that site-based management requires new skills on the part of administrators, staff, and teachers. However, they indicate they are not receiving the money, professional development, and other support needed to effectively create and operate in this new environment. All of these elements have come together to make the job of school administrator more complicated and demanding, leading one respondent to write that the job requires someone who "has endurance and is willing to work, work, work."

Responses to the major challenges principals faced when they began in that role and the skills they believe are

required for success in the future resulted in similar categories but differences in the focus and breadth of the skills identified within them. The categories of skills needed are: relationships, communication, management, and curriculum. In the category of relationships, when principals began their jobs, they had difficulty establishing positive interpersonal relationships, not only with parents and teachers, but also with the central office. Getting parents involved with their children's education was also noted among the highest challenges. When dealing with skills principals will need in the future, respondents stressed the need to establish strong relationships not only with faculty, students, and parents, but with others inside and outside of the school community. They emphasized the need to be fair, to be a good listener, and to have the "ability to work well with a broad group of people." Establishing strong effective public relations with the community as a whole was likewise noted as essential to administrative success.

When dealing with communication, principals noted the problems they encountered in establishing a common vision for their school and developing a team focus among the teachers when they took over as the leader of the school. Motivating teachers, maintaining morale, and conducting the staff development needed to enable personnel to improve their performance were all major challenges, as principals attempted to create a strong and unified staff. This capacity to create a common vision was identified as an essential skill for future leaders by a great majority of the respondents. The principals also stressed that these visions must go beyond what is, to include innovation and transformation. A statement that captures the essence of this belief is "Be a person of vision. Know where you want to go and how to lead your school to get there."

In the area of management, principals stated that when they took over as principal they faced many challenges related to managing funds and securing facilities for students. Time management was another major issue. Principals frequently reported that they had no one to delegate work to and found themselves overcome by competing problems. The ability to organize time, prioritize tasks, understand the technical aspects of management, and handle multiple tasks with ease were the management skills considered as essential for future administrators. As one respondent stated, the principal must be able to "unstop a commode, listen to teachers' complaints, discipline a student, interpret a special education law, and eat lunch, all at the same time."

When principals became the leader of the school, one of their major challenges was in the area of curriculum. The most often noted topics of concern initially were dealing with special education, understanding curriculum matters, and dealing with student performance on standardized tests. Principals stressed the importance of having a strong knowledge base in these areas as well as being able to work with the staff to assure excellence in curriculum and instruction. They believe that future leaders must be well versed in "diverse instructional strategies" and in "research on teaching and learning," and "know what is best for children and how to help them to learn."

In addition to the specific job categories identified for the success of future administrators, principals also stressed that these individuals must be people of "courage" who possess personal attributes of integrity and fairness, have a concern for people and the needs of children, and hold strong ethical standards.

Discussion and Implications

Reflecting on the Workforce

There appear to be four major issues that need to be addressed relative to the demographic data reported. Principals in Alabama are well-educated individuals who are primarily white males from the South. Within the next ten years, most of them will retire. Their reasons for retiring relate to better opportunities and stresses of the job. A recent national study regarding what discourages or prevents qualified applicants from applying for principal positions also mentioned job responsibilities and compensation as barriers (NASSP, 1998). Consequently, it seems imperative that Alabama consider the financial aspect of this job and take steps to ensure that school administrators are paid a salary commensurate with the role they perform. Likewise, the stresses of the job must be reduced and/or steps taken to assure that leaders are equipped to deal with them. Otherwise, the state and perhaps the nation may find itself without a qualified body of individuals willing to accept and maintain these positions. As one respondent stated, "new administrators will not be willing to put in the time and dedication that it takes to lead excellent schools."

A second concern rests with the gender and racial composition of the present administrative force. Female students represent at least 50 % of the student population, yet females are underrepresented in the administrative ranks, leaving a scarcity of role models for females, which may perpetuate the disparities in these figures. This is also a problem with a national scope since less than 35% of school administrators in the nation are females.

When examining issues of ethnicity there is also cause for concern. Although 35.8% of the student population in Alabama is African-American, only 16% of the principals are from this ethnic background. In addition, there are indications that the population is becoming

more diverse leading to concerns about the future. When the ethnic composition of the student population was examined in 1986, 62% were non-Hispanic Caucasians, 37% African-American, 1% Hispanic, and .09% other minorities. In 1994 the percentages were 62.3% non-Hispanic Caucasians, 35.8% African-Americans, 4% Hispanic, and 4% other minorities. Thus, although the percent of non-Hispanic Caucasians and African-American students in Alabama has remained fairly constant, there is a small but steady increase in the percentages of other student minorities. On a national level ethnic diversity is likewise increasing, and early in the 21st century, over 50% of the population will be comprised of groups now classified as minorities. Some of these groups appear to have the greatest problems with school achievement, and schools continue to struggle with how to meet their academic and other needs. Thus it will become increasingly important for school administrators to be familiar with the diverse cultural backgrounds of the students they serve. Likewise, minority groups may need to be more widely represented in the administrative ranks in order to serve as role models and to have credibility in the communities they serve.

It might therefore be advantageous for Alabama and other states with similar demographic figures among school administrators to identify and groom female and minority leaders in a systematic way to resolve their low representation in the administrative ranks in schools (Crow, Mecklowitz, & Weekes, 1992). Scholarships and other incentives might be avenues to boost the number of women and minorities admitted into educational leadership programs and to mentor and support them once they become administrators. In addition, school systems in Alabama and other states with similar statistics should consider examining their hiring and selection practices. It might also be valuable for them to monitor the demographic characteristics of their administrators and take action to assure appropriate diversity among school leaders.

There are two other characteristics of principals which may be closely related to student achievement and the need for the reform and renewal of schools. One is that most principals assume their position from other positions within their systems. While this means they will tend to gain acceptance in the system, it may also inhibit innovation and transformation as they may follow previous practice and practitioners even if their schools are not succeeding. The finding of this study that principals develop their leadership styles primarily from other models lends credence to this possibility. In addition, most of the principals assumed their positions after being assistant principals. While such a position may make them familiar with the school, research indicates that this position does not prepare individuals to be instructional or transformational leaders (Hartzell, 1993; Michel, 1996). In fact, it may do just the opposite, encouraging them to conform to the status quo (Hartzell, 1993; Koru, 1993).

The state of Alabama has identified over 30% of its schools as being on caution or alert which means that based on state guidelines, students in these schools are not achieving at appropriate levels on standardized tests. Numerous states have also instituted accountability standards, and there is a push for national standards to raise achievement and assure a quality education for all. Improving achievement and changing the status quo in schools will require innovative leaders who can serve as change agents as noted by the principals in this study. The state, educational institutions of higher education, and school systems, working individually or in consortia, should develop programs that will assist in assuring that those who move from assistant principal to principal will have the necessary skills and experiences to appropriately fulfill their new role. School system leaders should also address this issue by examining whether the practice of hiring primarily from within their system is hindering or facilitating school improvement. They should also examine whether they are providing working environments in which forward-thinking and visionary ideas are welcomed and affirmed.

Principal Capabilities and Preparation

It appears that Alabama school leaders feel confident of their ability to lead and plan. The importance of modeling and mentoring by those in the field and the experiences they had while in graduate school appear to have been strong influences on their leadership styles. It would seem wise for those involved in leadership preparation programs, superintendents, the state, and others in the field responsible for enhancing principal success to include appropriate mentoring and modeling opportunities for those preparing for and beginning the principalship. The absence of quality mentoring in education has long been a serious problem and one which should be addressed (Scott, 1992). Since mentoring appears to be a particularly important element in the development of principals, a special problem related to the demographic data in this report needs to be considered. In the past, gender and ethnicity may have prevented the connecting of potential female and minority administrators with mentors because the majority of school administrators are white males (Galbraith & Cohen, 1995). One successful strategy in dealing with this issue is to have organizational mentoring programs which assure a mentoring relationship for all those aspiring to administrative levels. If such programs include avenues for enhancing cultural and ethnic understanding among and between all members of the organization, such programs can be very effective (Galbraith & Cohen, 1995).

When ranking the ASDE skills (Table 6) principals placed "Establishes clear instructional goals and specific achievement objectives for the school" and "Implements clear instructional goals and specific achievement objectives for the school" as the most important skills for principals with mean scores of 4.68 for both of these competencies. However, when judging their ability (Table 5) the skill dealing with instructional implementation was ranked ninth, with a mean of 4.16, and the skill dealing with establishing goals and objectives was ranked tenth, with a mean of 4.14. While these mean scores are above average, their ranking in terms of principal competency level as compared with their ranking in terms of importance, along with the fact that they identified curriculum and instruction as areas they had difficulty with when they entered the principalship, may be a cause for concern. The importance of this skill as ranked by these principals and the potential impact such competence can have upon student achievement necessitates a closer examination of principal abilities in this area.

Although principals rated themselves quite highly in organizational skills (Table 5) with a 4.36 mean and a rank of 3, time management was difficult for them when starting out as administrators. The expanded demands and myriad management responsibilities made upon administrators make the ability to manage time of vital importance. However, time management is not specifically addressed on the state of Alabama competency list. This skill area should be examined as the state continues its revision of certification standards and programs. Likewise, preparation programs should review their curricula to assure that students receive prototype and field-based experiences to develop skills in effective time management. School systems should also ensure that principals receive updated training in effective management techniques and that they are able to contact central office or other personnel from whom they can receive help and guidance. Workshops, manuals, compact discs, or mentors, and other individuals who can provide technical assistance should be considered as potential resources to support school leaders and enable them to become more competent to handle the complexities of their position.

Principals in this study reported that when they entered their position they experienced difficulty in their relationships, particularly in communicating and maintaining positive interpersonal relationships, in forming teams, and in getting people to work together to develop a team vision. Snyder and Graves (1994) claim that those who are successful in creating and having others accept such visions must be willing to be risk-takers who are not afraid of failure. Likewise, those who can transmit a vision and gain support for it must possess strong interpersonal and communication skills (Augenstein, 1988). Thus, it appears vital that those involved in leadership preparation programs consider providing experiences that involve risk-taking and the development of relational and communication skills.

The interpersonal, visionary, and personal skills and attributes identified by principals in this study as essential for future leaders to effectively deal with the leadership challenges they will face are not listed or identified as principal competencies in this state. Those responsible would do well to heed these findings, expand upon them, and incorporate the skills of team development, communication, and interpersonal relationships into any existing and proposed preparation, certification, and assessment programs. In addition, issues dealing with ethics, integrity, and equity, identified as essential for future administrators, should be integrated into principal preparation and professional development experiences. This has been recognized by the Interstate School Leaders Licensure Consortium (ISLLC) which included an ethics standard as one of the essential principal competencies (1996). This standard and methods for measuring it should be examined and considered for inclusion in the Alabama competency list. Consideration should be also be given to incorporating aspects of personal attributes such as these into the selection criteria for entry into certification programs and criteria for principal selection procedures and processes.

Conclusions

The voices of the principals in the state, recorded and reported through these surveys, provide important information about the needs of these leaders and the environment in which they must operate. Although the data reported in this study were collected from principals in Alabama and provide specific information about training and skill needs, the growing demands on school leaders is documented on a national level (Kochan, Jackson, & Duke, 1998). As one of the Alabama principals so eloquently stated, the principal of the future must have " the wisdom of Solomon, the patience of Job, the long suffering of Jeremiah, and a nervous system of iron." If this is what is expected, the job, the demands, the rewards, the preparation, and the potential pool of candidates to fill this role should be examined, discussed, and restructured. It is hoped that the voices and the findings reported herein will help to stimulate this dialogue and foster the creation of actions that will assure that there is will be an ample supply of quality school administrators available to successfully lead the schools of tomorrow in Alabama and the nation.

References

Alexander, G. C. (1992, April). *The transformation of an urban principal: Uncertain times, uncertain roles.* Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Auguenstein, J. J. (1988). The real and the ideal. *Momentum, 19*(2), 50-51.

Bartell, C. (1994, April). *Preparing future administrators: Stakeholders perceptions.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Crow, G. M., Mecklowitz, B., & Weekes, Y. N. (1992). *From teaching to administration: A preparation institute.* Lancaster, PA: Technomic.

Galbraith, M. W., & Cohen, N. H. (Eds.). (1995). *Mentoring: New strategies and challenges.* San Francisco: Jossey Bass.

Glickman, C. (Ed.). (1992). *Supervision in transition.* Alexandria, VA: ASCD.

Hanson, M. W. (1991). *Educational administration and organizational behavior.* Boston: Allyn & Bacon.

Hartzell, G. (1993). When you're not at the top. *High School Magazine, 1*(2), 16-19.

Hoy, W., & Miskel, C. (1991). *Educational administration: Theory, research and practice* (4th ed.). New York: McGraw-Hill, Inc.

Interstate School Leaders Licensure Consortium (ISLLC). (1996). *Standards for school leaders.* Washington, DC: Council of Chief State School Officers.

Kaiser, J. S. (1995). *The 21st century principal.* WI: Stylex.

Kochan, F., Jackson, B., & Duke, D. (1998). *Report of phase I of the UCEA research study, A thousand voices from the firing line: A study of educational leaders, their jobs, and the problems they face.* (Available from University Council on Educational Administration, University of Missouri, 205 Hill Hall, Columbia, MO 65211).

Koru, J. M. (1993). The assistant principal: Crisis manager, custodian, or visionary? *NASSP Bulletin, 77*(556), 67-71.

Lester, P. E. (1993, April). *Preparing administrators for the twenty-first century.* Paper presented at the annual meeting of the New England Educational Research Association, Portsmouth, NH.

Michel, G. (1996). *Socialization and career orientation of the assistant principal.* EDRS Number ED 395381.

Murphy, J. (Ed.). (1993). *Preparing tomorrow's school leaders: Alternative designs.* University Park, PA: UCEA Inc.

Nadeau, A., & Leighton, M. S. (1996). *The role of leadership in sustaining school reform: Voices from the field.* Washington, DC: US Printing Office. Contract No. EA 9478001.

National Association of Secondary School Principals (NASSP). (1998). *Is there a shortage of qualified candidates for openings in the principalship: An exploratory study.* Arlington, VA: National Research Service

Richardson, M. D., Flanigan, J. L., Lane, K. E., & Keaster, R. (1992, November). *Teacher perception of principal behaviors: A research study.* A paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN. ERIC Number ED 352 710.

Scott, M. E. (1992) Designing effective mentoring programs: Historical perspectives and current issues. *Journal of Humanistic Education and Development, 30*(4), 167-177.

Short, P. M., & Greer, J. T. (1997). *Leadership in empowered schools: Themes from innovative efforts.* Columbus, OH: Merrill.

Snyder, N. H., & Graves, M. (1994). Leadership and vision. *Business Horizons*, 1-7.

United States Department of Education (USDOE). (1996). *Schools and school staffing in the United States: A statistical profile.* Washington, DC: Author.

Valentine, J. W., Clark, D. C., Irvin, J. L., Keefe, J. W., & Melton, G. (1993). *Leadership in middle level education.* Reston, VA: National Association of Secondary School Principals.

Wallace, R. C. (1992). Leadership in schools. In S. Thomson, *School leadership* (p. 8-15). Newbury Park, CA.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 1, 17-24

Schools Without Principals: Do Both Management and Leadership Exist?

Theodore B. Creighton
Idaho State University

Successful schools have a common glue that keeps a faculty together and creates consistency among a school's various elements (Glickman, Gordon, & Ross-Gordon, 1995). The glue is the process by which some person or group of people is responsible for promoting a link between individual teachers' needs and organizational goals. Traditionally, the person responsible for linking teachers to organizational goals has been the building principal, superintendent/principal, and in some cases the vice principal. However, schools are beginning to alter the structure of administration to exclude the principal. Many schools vary with respect to who carries out leadership and supervisory responsibilities. It is crucial for us in educational administration preparation programs to look closely at not only what is happening in school leadership, but also why it is happening. The following study involves a close look at an elementary school operating without the services of a principal or vice principal.

Introduction

The effective schools research of the 70s and 80s indicates that strong administrative leadership is directly related to school improvement efforts and increased student learning (Berman & McLaughlin, 1978; Edmonds, 1979; Fullan, 1982). Though little recent research exists indicating a direct relationship between principal instructional leadership and student achievement (Hallinger, Blickman, & Davis, 1996), there is consensus and agreement that a principal has (at least) an indirect effect on student achievement through actions that shape a school's learning climate (Creighton, 1996).

Ogawa and Hart (1995) report that the principal variable accounted for 2-8% of the variance in student test scores. This finding further supports the contention that principals have the potential for directly affecting student learning. Though there is little disagreement concerning the belief that principals have an impact on the lives of teachers and students, both the nature and degree of that effort continues to be open to debate (Murphy, 1992).

Theodore Creighton is Professor and Coordinator of Master's programs in the Department of Educational Leadership at Idaho State University. Prior to joining the faculty at ISU, he served as a district superintendent in Fresno County, CA. Donald Coleman's work with the National Association of Elementary School Principals (NAESP) in improving the Administrative Diagnostic Inventory which contributed to this article is gratefully acknowledged. Correspondence regarding this article should be sent to Theodore B. Creighton, Department of Educational Leadership, Idaho State University, Campus Box 8019, Pocatello, ID 83209 or by e-mail to creitheo@isu.edu.

Fullan (1993) contends that if true learning organizations occur, the principalship as we know it may disappear. Nias (1992) found in her studies that even the collaborative leader was too strong for the development envisioned in the new work of school leaders, and that even the collaborative principal may be too authoritarian for the true learning organization to take shape. Sarason (1997), in his recent book, *How Schools Might be Governed and Why,* suggests and supports a proposal to eliminate the principal from the school governance structure.

The large number of studies on effective schools of the mid 1960s and early 1970s (Brookover, 1979; Coleman, 1966; Edmonds, 1979) clearly identified strong leadership as a characteristic of good schools. A second wave of effective schools research (Goodlad, 1982; Purkey & Smith, 1983; Steadman, 1987) found the principal's leadership crucial to school effectiveness.

In the 1990s, researchers are describing varied sources of instructional leadership (Austin & Reynolds, 1990; Chubb & Moe, 1990; Levine, 1991; Pajak & Glickman, 1989). In many cases, principals are becoming secondary leaders. Primary instructional leaders vary from system to system but include central office supervisors, assistant principals, department chairs, grade-level chairs, and teams of teachers.

Though the role of the principal is being questioned, research clearly reveals that a characteristic of effective schools is that someone, somewhere is responsible for and committed to the process, function, and tasks of school leadership (Glickman, Gordon, & Ross-Gordon, 1995).

This study began with a focus on three elementary schools in Casper, Wyoming, and Boulder, Colorado, which are changing the traditional paradigm of school

governance by providing for student learning without the usual principal or vice principal. Visits to the two schools in Colorado revealed a "principal structure," with teacher/leaders or teacher/principals performing the duties of the building principal. The leaders taught in the classroom for half a day and performed administrative duties the other half. Only the elementary school in Casper, Wyoming had a governance structure that truly reflected the absence of a principal. The author selected this school for further study.

Revision of the Administrative Diagnostic Inventory (ADI)

The Administrative Diagnostic Inventory (ADI) is the assessment tool used by the National Association of Elementary School Principals (NAESP) in their assessment center evaluation of aspiring and practicing school principals. In November 1995, the NAESP accepted a proposal to revise the ADI assessment instrument (Coleman, 1996).

A research and writing team of eight professional educators (including professors from five universities, practicing school superintendents, school principals, research consultants, assessment center directors, and the NAESP director) was invited to California State University, Fresno, to create new constructs of effective school leadership. The result was an agreement of 12 constructs (dimensions), or knowledge and skills of effective school leaders. These constructs covered both management (personal skills) and leadership (interpersonal skills) and are shown in Table 1.

Table 1
ADI Constructs of Effective School Leadership

Management Traits	*Leadership Traits*
Planning	Vision
Organizing	Communications
Problem Solving	Instructional Leadership
Creativity	Climate Development
Decisiveness	Team Building
Systems Analysis	Instructional Supervision

Further support for including these two components of educational leadership is given by the National Association of Secondary Principals (Thompson, 1993). The NASSP recommendations for principal certification include:

1. Leadership - short and long-term planning, internal and external communication, research and practice of leadership, problem analysis and decision making in organizations.
2. Management - applications of management software, human resource management, administration of facilities and services.

The research and writing team concluded by identifying eight descriptors for each of the 12 constructs. The eight descriptors for the management construct of "planning" are displayed in Table 2.

Table 2
Construct of Planning with Eight Descriptors

PLANNING - defines purpose and sets organizational goals	
01	seeks clarity in purpose
02	utilizes planning and goal setting to advance school mission
03	clarifies the purposes of various activities
04	encourages staff in planning and goal accomplishment
05	links planning to goal accomplishment
06	demonstrates a sense of purpose when setting goals
07	involves stakeholders in the planning process
08	focuses on organizational purpose and goal accomplishment

Statement of the Problem

Several schools have altered their governance structure to exclude the traditional principal or vice principal. There is limited research examining the leadership characteristics of schools without principals. The present study was designed to examine the existence or non-existence of the leadership and management constructs contained in the ADI assessment instrument. The desire of the author was to determine if a school without a principal possessed the same leadership characteristics associated with effective schools.

Research Questions

Two research questions guided the study:

1. Are leadership and management constructs of the ADI evident in the principal-less governance structure?
2. Are schools without principals maintaining the consistency, control, and effectiveness associated with effective principals?

Methodology

Any attempt to relate the character of one school to others should proceed with caution. Likewise, any attempt to suggest one particular model of school governance as having positive or negative implications for other schools must acknowledge its limitations. The study conducted represents a case study of one elementary school and its unique form of school governance.

The elementary school under study includes kindergarten through grade eight with an approximate enrollment of 200 students. Classes consist of two-grade combinations with two teachers sharing the instruction of combination classes for Grades 1-2, Grades 3-4, Grades 5-6, and Grades 7-8. The student population is predominantly white, medium to high socio-economic status, with approximately 4% of the students qualifying for free lunch.

In 1990, a team of teachers presented a proposal for a new school to their superintendent and board of trustees. The proposal included a governance structure with no principal or vice principal. Each of the identified responsibilities of a school principal was assigned to an individual, and in some cases, a pair of teachers. For example, two teachers were responsible for personnel issues, two teachers handled school finances, two teachers were in charge of discipline, etc. Early the next year, the superintendent and board accepted the proposal and doors opened for students in the fall of 1991.

The present study resulted from research conducted during the 1996-97 school year and utilized data gathered through interviews with administrators, teachers, parents, and students. A university teacher education major assisted in the administration of a parent/student survey, and additional phone interviews took place with the former superintendent and board members involved in the original proposal in 1991. Observations and interviews conducted at the school site also contributed to the research. The observations, field notes, interviews, and analysis of documents collected in the field form the empirical basis for this study. A purpose of the study was to investigate the claim that a school without a principal promotes and sustains effective learning for children.

Interviews were taped, transcribed, and entered into the Statistical Package for the Social Sciences (SPSS) for analysis. Correlation and comparison of means helped to identify significant patterns in the area of school leadership. Management and leadership constructs identified by the National Association of Elementary School Principals' Administrative Diagnostic Inventory (ADI) guided and focused the coding and interpretation of the data.

The questions used in the interviews and on the survey instrument were developed from the descriptors for each of the ADI constructs. Specifically, descriptors from each construct were used to create the items for the survey and interview. The researcher used the descriptors to determine if the construct existed in the principal-less governance structure. For instance, two descriptors (or look-fors) used for the "planning" construct were: (a) involvement of all stakeholders in the planning process, and (b) encouragement of staff in planning and goal accomplishment (see Table 2).

The population interviewed personally or by phone included one district superintendent, five board members, ten classroom teachers, thirty parents, twenty students, and three university professors familiar with the school under study.

Data Collection and Analysis

The questions used for personal interviews with teachers, parents, and administrators focused on the ADI descriptors. In addition, the questions used for the parent/student survey were developed using the descriptors. The intent was to identify the existence or non-existence of both leadership and management characteristics as defined by the NAESP and NASSP.

Respondents were asked to rate the existence of the descriptors on a 3-point Likert scale from 1(not evident) to 2 (evident) to 3 (highly evident). For example, respondents were asked if all stakeholders (i.e., parents, students, teachers, and district level personnel) were involved in the planning procedure at the school. This question relates to one of the ADI descriptors for the construct of planning. Questions from each of the 12 constructs were included in the interview process. The selection of descriptors was made according to the results of a previous construct validation study (Creighton, 1996), which subjected the entire ADI instrument to R factor analysis with Kaiser Normalization and Oblimin Rotation. This factor analysis identified the correlation of the descriptors within each construct and the relationship with the construct itself. Only descriptors highly correlated with each construct were used for survey and interview questions in this study.

The data collected were then categorized according to the appropriate constructs and entered into the Statistical Package for the Social Sciences (SPSS). Data analysis consisted of using the Pearson Product Moment Correlation and a comparison of means with standard deviations.

Results and Findings

Finding 1

Addressing the following question in the study was crucial: What was the rationale for eliminating the role of the principal in the original proposal? In discussions with the school's planning team, a recurring theme for the "principal-less" structure was the desire to use the principal's salary (i.e., $40,000 - $50,000) to accomplish a student - teacher ratio of 18 - 1. The planning team

reasoned that if the cost of a principal could be saved, the amount could be used to reduce the student – teacher ratio in each classroom. Their proposal insisted that the money saved by the elimination of the principal remain in the local school's budget, and in this case be used to place two teachers in each classroom with 36 students (student - teacher ratio of 18 - 1).

The former superintendent and two board members concurred that assigning a principal within their district to the proposed school would have been counterproductive to their proposal. They continued by saying they might jeopardize teacher/parent responsibility and ownership. The former superintendent stated:

> I did not feel that any of our current principals could have stepped in and provided the cooperation and shared decision making which were so important to the proposal. I have come to believe that a principal can be the single greatest roadblock to school improvement.

Finding 2

Though the managerial duties of a school principal (e.g., discipline, finance, and scheduling) were equally distributed across the staff, there was no evidence of any specialized training or expertise in the areas of assignment. For example, the pair of teachers assigned to finances seemed knowledgeable in math and budgeting, but had received no specific training in school finance and its unique needs and characteristics. The teachers responsible for personnel issues had no background or formal training in crisis management, group dynamics, etc.

Responding to parent interviews, some believed that when they had a specific concern about an issue, it was difficult to identify the staff in charge of that area (e.g., a problem occurring on the playground or bus). In addition, upon identifying the staff member in charge, it was difficult to talk with that person immediately, since they were in the classroom with students. The issue often had to wait until lunch or after school, or be placed on an agenda for the weekly staff meeting. A parent shared the following:

> I found it difficult to know which staff member to talk with if I had a specific concern. For example, I was recently upset about something happening on the playground and wasn't sure who to talk with about it.

Finding 3

Management as defined (the day-to-day operation of the school) seemed to be sufficiently covered by the teacher teams. However, leadership as defined (fostering purpose, passion, and imagination) was less apparent. Classes met on time, busses arrived and departed on time, and recess and cafeteria duties were carried out as scheduled. Teachers and other staff performed their duties and appropriate instruction occurred in individual classrooms.

Though teachers possessed vision and ideas for the future, a unified common purpose was not evident. There was also some disagreement among staff in regard to future direction and focus. Some felt a need to maintain the status quo from 1991, while others felt an urgency to adjust for the future. Still others felt uncomfortable in straying too far from the original proposal.

It is helpful here to recall that surveys and interviews asked respondents to respond to various statements built around the management and leadership constructs of the ADI. For example, two interview statements are shown in Table 3: one in the management domain and another in the leadership domain, along with the options for reply.

The mean response for the construct of PLANNING (management) was 2.33 with a standard deviation of .48. The mean response for the construct of INSTRUCTIONAL SUPERVISION (leadership) was 1.13 with a standard deviation of .35. A complete comparison of means and standard deviations of all ADI constructs is displayed in Table 4.

Table 3
Management and Leadership Interview Statements

1. PLANNING - Management
The staff is encouraged to participate in planning and goal accomplishment.

Not evident	evident	highly evident
1	2	3

2. INSTRUCTIONAL SUPERVISION- Leadership
Continual frequent formative teacher evaluations are conducted.

Not evident	evident	highly evident
1	2	3

Note: ADI constructs were not labeled or identified in the actual interview.

Table 4
Mean scores and standard deviations of ADI Constructs

Construct	Mean	Standard Deviation
Management		
PLANNING	2.33	.48
ORGANIZATION	2.67	.45
PROBLEM SOLVING	2.38	.48
CREATIVITY	2.33	.39
DECISIVENESS	2.12	.41
SYSTEMS ANALYSIS	2.16	.33
Leadership		
VISION	1.97	.18
COMMUNICATION	1.20	.41
INSTRUCTIONAL LEADERSHIP	1.33	.48
TEAM BUILDING	1.37	.48
CLIMATE DEVELOPMENT	1.77	.43
INSTRUCTIONAL SUPERVISION	1.13	.35

Finding 4

The existing conception of a school principal was not consistent with today's common view of the principal as instructional leader. Upon talking with many teachers, parents, and school board members, a principal image emerged characterized as a manager, rather than an instructional leader. The image also emphasized administrative control and authority. The sense was that those interviewed did not view the principal as one focused on teaching and learning. Quite the contrary, there were indications that some viewed the principal as someone who might actually impede teaching and learning. The former superintendent stated:

> Our concern was that any of our current principals would actually be counterproductive to the proposal made by the teachers. Each of them would have problems with shared decision making and empowering the teachers and parents.

A former board member echoed similar concern:

> We as a board did not view the absence of a principal as an issue. Our concern was that a traditional principal would not have the ability to relinquish individual authority and allow for total shared governance at the school. We believed the existing leadership (principals) in our district would be counterproductive to the proposal.

Finding 5

An absence of any consistent and regular system of personnel evaluation and supervision existed. Since there were no administrators in the building to conduct formative and summative evaluation of staff, the district appointed two principals from other schools to conduct teacher evaluations. This system of evaluation presented the following concerns: (a) the principals, coming from other buildings, lacked the familiarity with the day-to-day instruction in the classroom, (b) the visiting principals were not part of the culture of the building, and (c) teacher performance evaluation was based on a very limited amount of observational time.

Finding 6

An absence of any formal channel of communication was apparent, especially in regards to organizational goals. This finding was supported by the weak responses to the construct of communication (see Table 4). Communication is complex and permeates every aspect of school life. The translation of goals into concrete actions and accomplishments depends on regular, ongoing, and successful information exchanges (Hoy & Miskel, 1996). Though the staff communicated through a weekly faculty meeting, a lack of daily communication resulted in many issues needing a quick response having to wait until the Friday meeting. A need existed for someone to serve as an anchor necessary to provide guidance and assistance during times of stress and crisis.

Finding 7

The connection of school goals and objectives to district goals appeared weak and unimportant. The school site goals were fairly obvious and shared by the total staff. Though there was a common direction and focus, the relationship to the district office appeared unproductive. Little collaboration existed between the local school and district office. It was not apparent that the local education plan related to the overall vision of the district. This finding was supported by relatively weak responses to statements in the ADI constructs of Vision, Communications, and Instructional Leadership, all of which addressed the issue of working effectively with the whole district community to advance student learning (see Table 4). In addition, the interviews conducted with other parents, teachers, and administrators in the district revealed ambivalence toward the school under study. A principal stated:

> I do not believe there is effective leadership at the school. They seem to be doing their own thing, and not concerned with the overall mission and goals of the district. Though they

have a teacher who attends administrative meetings, how can that person have time to promote the direction of the district and teach also?

Discussion

The National Association of Elementary School Principals (NAESP), the National Association of Secondary Principals (NASSP), and the National Policy Board of Educational Administration (NPBEA) have identified several constructs of effective leadership (Creighton, 1996). It makes sense to assume that if the responsibilities of the principal are distributed to teams of teachers, the identified leadership construct should still be evident in the school culture. For instance, school effectiveness is determined by a principal's ability to plan appropriately. This same construct of planning should exist in the operation of a school deciding to eliminate the principal from their governance structure. The school under study exhibited a strong level of the planning construct.

In summarizing the data, a restatement of the first research question might be helpful: *Are both the leadership and management constructs of the ADI evident in the principal-less governance structure?* The data reveal that management constructs generally existed (or were perceived to exist) in the school without a principal. All six constructs displayed mean scores ranging from 2.12 to 2.67. However, only two of the leadership constructs appeared strong: (a) Vision with a mean of 1.97 and (b) Climate Development with a mean of 1.77. The implication is that the elimination of the principal did not greatly affect the management responsibilities but resulted in many of the leadership characteristics or constructs being weak or non-existent. The finding may also imply that management kinds of responsibilities lend themselves to teacher teams more than leadership constructs, which may benefit from a specific person responsible for those duties.

A significant and recurring pattern appearing in the data analysis was a common perception of a principal. The teachers, parents, and students interviewed described a principal as a person responsible for discipline, finances, and scheduling. This perception seemed based upon mechanical control and maintenance. Absent was any mention of the principal as educator: one who directs teaching and learning. Absent was any mention of the principal as a person who directs efforts to improve student learning and creates a positive school climate.

An important question here is: Why was this perception of the principal so different than the one we have talked about for two decades? The perception of the principal at the university is much different and focuses on the importance of instructional leadership, guiding a district's vision, and working collaboratively with the total staff to promote and sustain increased student performance.

The argument exists stating the relationship between university administrator preparation programs and our public schools is generally weak and in many cases, nonexistent (Murphy, 1992). Though we continually talk about the importance of school-university partnerships, in reality the relationship is more rhetoric than fact. I realize that sending administrators trained in effective leadership to public schools often results in the new leader being consumed in an environment of management and putting out fires. When the system (school site) views leadership as "managerial," it is very difficult for the newly trained instructional leader to avoid being "sucked up" into the world of managing.

Another area of responsibility exists with the university teacher preparation programs. Why would an entire staff view the principal as consisting of six or eight clearly defined tasks that can be performed by individual staff members? Very few teacher preparation programs include training (much less, dialogue) in the area of school leadership. Teachers are expected to respond to school governance but have had little training in preparation. Where in teacher preparation programs do future teachers learn about leadership? Teachers are generally "unsophisticated" in the subject of school governance (Sarason, 1993).

Should we teach leadership skills in administration preparation programs only, or should we also teach leadership skills in our teacher preparation programs? Goodlad (1991) stated undergraduate and graduate teacher education programs do not regularly include preparation in assuming leadership roles outside the classroom. Lieberman, Saxl, and Mills (1988) reported that teachers need to develop skills in building rapport, dealing with change, and managing the demands of leadership. Other authors (Gehrke, 1991; Manthei, 1992) have advised that leadership skills would be helpful to classroom teachers.

The suggestion is not to teach leadership skills to teachers for the sole purpose of eliminating the principal. The point is that a dichotomy exists in the perception of the principal's job. Unless we address leadership skills and competencies in our teacher education programs, attempts to improve schools through changes in governance structures will have little effect. Providing leadership training for our future teachers may result in more agreement regarding the meaning of effective leadership and an improvement in the common unproductive relationships between teachers and administrators.

Another pattern emerging from the data was the absence of *thinking* about the education of all students in

the school. Each teacher had a grasp and understanding of effective education in his/her classroom, but was not aware of any scope and sequence of all the learning activities occurring in the school. It is this *thinking* about the education of all students that characterizes an effective principal. A question surfaces: Without a principal or administrator, how is this overall *thinking* promoted and sustained?

Glickman, Gordon, and Ross-Gordon (1995) conclude five propositions concerning teachers' attitudes and awareness that can be promoted via leadership: (a) enhancing teacher beliefs in a cause beyond oneself, (b) promoting teachers' sense of efficacy, (c) making teachers aware of how they compliment each other in striving for common goals, (d) stimulating teachers to plan common purposes and actions, and (e) challenging teachers to think abstractly about their work.

Leadership is not something as an end in itself but must be viewed as developmental if schools are to become more successful. The new leadership of the 21st century must not only respond to teacher performance but also encourage greater involvement, autonomous thinking, and collective actions by teachers (Glickman, Gordon, & Ross-Gordon, 1995). University preparation programs must continue to revisit the way we prepare school leaders. New themes include the ability to build teams, create professional cooperation, and engage in collective action to improve instruction and promote student success.

Conclusion

The research on school based management and shared decision making reveals disappointing results to the success of these reforms (Mirel, 1994; Weise, 1995). Many new decision making structures have led to disaster (Beadie, 1996). One of the most significant studies of decision making in the schools (Weise, Cambone, & Wyeth, 1992) looked at the shared governance structures of forty-five schools across the country. The recurring theme across all schools studied was the confusion over who actually had the final word when it came to accepting and implementing a decision. In every case but one, the push toward innovation came from the principal. It was not the inclusion of teachers in the decision making process that precipitated reform, but the commitment of the principals to change the way the school operated (Weise, 1995).

This study revealed a prevalent misconception of the principalship to be similar to Senge's (1994) description of the common conception of successful managers in the business world – "being decisive, being in control, knowing what is going on, having answers, and forcefully advocating your views" (p. xvi).

Of particular note is the indication that the teacher teams are handling the management responsibilities rather successfully but are not addressing many of the leadership responsibilities of building principals. This finding does not necessarily imply that teachers are incapable of leadership, as much as it may indicate that their perception of the principalship is mainly managerial. Little evidence surfaced indicating a view of the principal as one who displays the constructs of vision building, oral and written communication, instructional leadership, and instructional supervision. The school under study continues to operate today under its principal-less governance structure. Students score at or above the district average on standardized testing. There seems to be no evidence that students encounter difficulty in adjusting to the formal principal-structure as they move on to high school. There is no conclusive evidence of a negative response to the second research question: *Are schools without principals maintaining the consistency, control, and effectiveness associated with effective principals?*

A strong recommendation is made for further study to determine if the findings of this study might be evident in other schools without principals. Many variables affect the leadership and management existing in schools. The replication of this study with other schools without principals is needed to further examine the issues of effective school leadership.

With current national attention given to schools of choice and charter schools, we will undoubtedly see more and more schools experimenting with alternative governance structures. The role of the principal will continue to be questioned and scrutinized. Those responsible for training educational leaders must continue authentic and ongoing dialogue with public schools and their communities. Unless leadership training and preparation include genuine communication with teacher education programs and our public schools, there will continue to be confusion and misconceptions about effective leadership and the role of the school principal.

References

Austin, G., & Reynolds, D. (1990). Managing for improved school effectiveness: An international survey. *School Organization, 10*(2), 167-178.

Beadie, N. (1996). From teachers as decision makers to teachers as participants in shared decision making: Reframing the purpose of social foundations in teacher education. *Teachers College Record, 98*,(1), 77-103.

Berman, P., & McLaughlin, M. (1978). Implementations of educational innovations. *Educational Forum, 40*, 347-370.

Brookover, W. (1979). *Social systems and student achievement : Schools can make a difference*. New York: Praeger.

Chubb, J., & Moe, T. (1990). *Politics, markets, and America's schools*. Washington, DC: The Brookings Institute.

Coleman, D. (1996, August). *Using symbolic-logic matrices to improve confirmatory factor analysis techniques*. Paper presented at the National Association of Professors of Educational Administration Annual Conference, Corpus Christi, TX.

Coleman, J. S. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.

Creighton, T. B. (1996). *A construct validation of the Administrative Diagnostic Inventory*. (ERIC Document Reproduction Service No. TM 027 024).

Edmonds, R. (1979). Effective schools for the urban poor. *Educational Leadership, 37*, 15-25.

Fullan, M. (1993). *Change forces: Probing the depths of educational reforms*. Bristle, PA: Falmer Press.

Fullan, M. (1982). *The meaning of educational change*. New York: Teachers College Press.

Gehrke, N. (1991). *Developing teachers' leadership skills*. (ERIC Document Reproduction Service No ED 330 691).

Glickman, C., Gordon, S., & Ross-Gordon, J. (1995). *Supervision of instruction: A developmental approach*. Needham Heights, Mass: Simon & Schuster.

Goodlad, J. (1982, January). *A study of schooling*. Paper presented to the Standford Teacher Education Project, Standford, CA.

Goodlad, J. (1991). A study of the education of educators: One year later. *Phi Delta Kappan, 73*(4), 311-316.

Hallinger, P., Blickman, L., & Davis, K. (1996). School context, principal leadership and student achievement. *Elementary School Journal*.

Hoy, W., & Miskel, C. (1996). *Educational administration: Theory, research, and practice*. New York: McGraw-Hill.

Levine, D. (1991). Creating effective schools: Findings and implications from research and practice. *Phi Delta Kappan, 72*, 389-393.

Lieberman, A., Saxl, E., & Mills, M. (1988). Teacher leadership: Ideology and practice. In A. Lieberman (ed.), *Building a professional culture in schools*. New York: Teachers College Press.

Manthei, J. (1992). *The mentor teacher as leader: The motives, characteristics, and needs of seventy-three experienced teachers who seek a new leadership role*. (ERIC Document Service No. ED 346 042).

Mirel, T. (1994). School reform unplugged: The Bensinville New American School Project. *American Educational Research Journal, 31*, 481-515.

Murphy, J. (1992). *The landscape of leadership preparation: Reframing the education of school administrators*. Newbury Park, CA: Corwin Press.

Nias, J. (1992). *Whole school curriculum in the primary school*. Bristel, PA: Falmer Press.

Ogawa, R., & Hart, A. (1995). The effect of principals on the instructional performance of schools. *Journal of Educational Administration, 22*(1), 59-72.

Pajak, E., & Glickman, C. (1989). Dimensions of school improvement. *Educational Leadership, 46*(8), 61-64.

Purkey, S., & Smith, M. (1983). Effective schools: A review. *Elementary School Journal, 83*, 427-452.

Sarason, S.B. (1993). *The case for change: Rethinking the preparation of educators*. San Francisco: Jossey-Bass.

Sarason, S. B. (1997). *How schools might be governed and why?* New York: Teachers College Press.

Senge, P. (1994). *The fifth dimension: The art and practice of the learning organization*. New York: Currency Doubleday.

Steadman, L. (1987). It's time we change the effective schools formula. *Phi Delta Kappan, 69*(3), 215-224.

Thompson, S. E. (ed.). (1993). *Principals for our changing schools: The knowledge and skill base*. Fairfax, VA: National Policy Board for Educational Administration.

Weise, C., Cambone, J., & Wyeth, A. (1992). Trouble in paradise: Teacher conflicts and shared decision making. *Educational Administration Quarterly, 28*, 350-367.

Weise, C. (1995). The four "I's" of school reform: How interests, ideologies, information, and instruction affect teachers and principals. *Harvard Educational Review, 65*(4), 571-592.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 1, 25-31

Theoretical Perspectives of Instructional Supervision in Contemporary Textbooks: A Content Analysis

Barbara L. White
The University of Southern Mississippi

Larry G. Daniel
University of North Texas

The purpose of the present study was to determine the degree to which various supervisory theories are dominant in contemporary instructional supervision textbooks. Twelve educational supervision textbooks were examined using a list of key words and concepts derived from several different supervisory theories. Results indicated that theoretical orientations consistent with evaluation-based theories were found in seven of the textbooks, whereas more clinically-oriented theories were dominant in the other five. No one theory consistently dominated the textbooks. Hence, considering that textbooks remain one of several major means for communicating a knowledge base to students, it is likely that supervisory practice will continue to be affected by varied theoretical orientations.

As a discipline, instructional supervision has been influenced by both evaluation-based and clinically-oriented theories. These theories are founded on a number of factors: prescribed orientations, taxonomies of teacher and supervisory behaviors, and descriptions delimiting the scope of the interactions between teachers and supervisors, to name a few. Theoretical content is communicated to supervisors in training in a plethora of ways; however, textbooks and other written materials remain one of several primary media for communicating this material. Hence, instructional supervision textbooks function as a means for determining instructional content and ultimately affecting supervisory practice. Considering the primacy of textbooks, the purpose of the present study was to examine the theoretical orientations utilized in contemporary instructional supervision textbooks, with the goal of determining the perspectives of supervision most likely to be dominant in the training of future instructional supervisors.

Theoretical Framework

In all professions, the development of the field rests on an established knowledge base and on newer, emerging information. Although instructional supervision has long been professionalized, the salience of theoretical orientation has been provident. Early on, the clergy controlled and supervised schools; thus, supervision reflected religious dogma (Oliva & Pawlas, 1997). In the 18th Century, under lay control, there was a dearth of mandates, and, consequently, laissez-faire attitudes prevailed (Wiles & Bondi, 1991). The mid-19th Century brought about the organization of the office of county superintendent and later the creation of state boards of education, the certification of teachers, and the institutionalization of state aid; and supervision took the form of administration "inspection" as states sought to promote school attendance and assure that educational programs met minimal standards (Anderson, 1993; Oliva & Pawlas, 1997).

In 1931, the Department of Supervisors and Directors of Instruction of the National Education Association, in an attempt to define the burgeoning profession and to delimit the roles and tasks of supervisors, proposed the following definition of instructional supervision:

> all activities by which educational officers may express leadership in the improvement of teaching. Such activities as observation of classroom instruction, conduct of teachers' meetings and of group and individual conferences are clearly within the meaning of this term. The development and execution of plans looking toward increased effectiveness in reading, arithmetic, and some other area of the school program, and the organization or reorganization of curriculum and method are still further examples of what is

Barbara L. White is Assistant Professor in Educational Leadership at the University of Southern Mississippi. Larry G. Daniel is Professor of Education at the University of North Texas. Please address correspondence regarding the article to Barbara L. White, Educational Leadership and Research, The University of Southern Mississippi, Hattiesburg, MS 39406-5027 or by e-mail to Larry Daniel at daniel@tac.coe.unt.edu.

> meant by supervisory activities. (Department of Supervisors and Directors of Instruction of the National Education Association, 1931, p. 3)

As time passed, however, and the complexity of the supervisory process was duly noted, this and other original definitions were found wanting, and the dilemma of developing a more definitive theoretical explanation was obvious. In 1972, Mosher and Purpel discussed the predicament:

> The difficulty of defining supervision in relation to education . . . stems, in large part, from unsolved theoretical problems about teaching. Quite simply, we lack sufficient understanding of the process of teaching. Our theories of learning are inadequate, the criteria for measuring teaching effectiveness are imprecise, and deep disagreement exists about what knowledge--that is, what curriculum--is most valuable to teach . . . . When we have achieved more understanding of what and how to teach, and with what special effects on students, we will be much less vague about the supervision of these processes. (p. 3)

A decade later, Harris (1985), acknowledging the complexity of both the supervisory process and the diversity of thinking in the field, noted that though several major theories had emerged, these theories often lacked focus. In this vein, Harris proffered the following insight:

> Supervision, like any complex part of an even more complex enterprise, can be viewed in various ways and inevitably is. The diversity of perceptions stems not only from organizational complexity but also from lack of information and absence of perspective. To provide perspective at least, the total school operation must be the point of departure for analyzing instructional supervision as a major function. (pp. 1-2)

Using Harris' framework, two major "total school" theoretical perspectives have been dominant in recent years. These two perspectives, noted herein as "evaluation-based" and "clinically-based" theoretical orientations, are each concerned with total school operation; however, whereas the former focuses on accountability, the latter focuses on personal and professional growth.

Evaluation-based Perspectives

By the 1970s and early 1980s, views of many of the major theorists in instructional supervision focused on measurement of teaching behavior, system accountability, and evaluation of teaching performance. This development in the field was due at least in part to the positivistic paradigm that had for quite some time dominated the field of social science research and evaluation. However, during the 1980s, numerous reports on the reform of American education were introduced, which called for specific measurable results, prompting even more so the move toward a practice of instructional supervision focused on evaluation and accountability.

In 1981, Alfonso, Firth, and Neville emphasized the officially designated behavior of supervisors as it directly affects teacher practice, which impacts student learning and achievement of organizational goals. John Lovell, in 1983, protracted the focus, noting the eminence of system accountability:

> Instructional supervision behavior is assumed to be an additional behavior system formally provided by the organization for the purpose of interacting with the teaching behavior system in such a way as to maintain, change, and improve the design and actualization of learning opportunities for students. (Lovell & Wiles, 1983, p. 43)

By the late 1980s, Krey and Burke (1989) provided a consummate definition of supervision which included several evaluation-based tenets, noting that as an accountability procedure instructional supervision consists of "instructional leadership that relates perspectives to behavior, clarifies purposes, contributes to and supports organizational actions, coordinates interactions, provides for maintenance and improvement of instructional programs, and assesses goal achievement" (p. 22).

Clinically-based Perspectives

Clinical supervision perspectives have been found in the instructional supervision literature for approximately three decades and have moved to a prominent position in the literature in recent years. Defined by Morris Cogan (1973), its leading pioneer and proponent, clinical supervision is:

> . . . the rationale and practice designed to improve the teacher's classroom performance. It takes its principal data from the events of the classroom. The analysis of these data and the relationship between the teacher and supervisor form the basis of the program, procedures, and strategies designed to improve the students' learning by improving the teachers' classroom behaviors. (p. 1)

Similarly, Acheson and Gall (1980) considered clinical supervision to be a process of supervisor-teacher

interaction that focuses on teacher classroom behavior. Likewise, Robert Goldhammer (1969) stated, "Given close observation, detailed observational data, face-to-face interaction between the supervisor and teacher, and an intensity of focus that binds the two together in an intimate professional relationship, the meaning of 'clinical' is pretty well filled out" (p. 54).

Lovell and Wiles (1983) defined clinical supervision as an objective approach that uses observation and analysis of teaching as a framework for improved performance. Tanner and Tanner (1987) provided epistemology and perspective:

> Like the concept of "internship," the term clinical is derived from the field of medicine, where it refers to practice based on the actual treatment and observation of patients as distinguished from experimental or laboratory study. However, where clinical medicine is focused on the treatment of ailments within a controlled environment, clinical supervision is conducted in the normal setting of the classroom and school, and involves the gathering of data from direct observation of actual teaching-learning events and conditions with the goal of improving classroom instruction. (p. 182)

"Developmental" supervision, more recently appearing in the literature, has been defined by its chief advocate, Glickman (Glickman, Gordon, & Ross-Gordon, 1995), as a school function whose goal--via direct supervision--is to (a) improve instruction, (b) assist curriculum development, (c) provide for professional growth, (d) promote esprit de corps, and (e) encourage action research. More narrowly, Tanner and Tanner (1987, p. 186) noted that developmental supervision involves problem-solving and curriculum improvement. Glickman's (Glickman, 1981; Glickman et al., 1995) concept of developmental supervision focuses on a sliding scale of supervisory behaviors based on the needs of any given teacher. As summarized by Tanner and Tanner (1987), "Glickman has used the developmental concept in supervision to denote different styles of supervisory leadership for the improvement of instruction. He sees the supervisor appropriately employing different leadership styles with different teachers and according to different circumstances" (p. 186). Likewise, Gordon (1990) spoke to contingencies that may affect the role of the supervisor, emphasizing instructional leadership that utilizes varying approaches to facilitate improved instruction and teacher cognitive growth.

Although the terms "clinical supervision" and "developmental supervision" have often been used interchangeably, Gonzalez-Baker and Tinagero (1983) distinguished between the two, noting that clinical supervision focuses on what and how teachers teach whereas developmental supervision is based on individual teacher stages of concern. Despite this valid epistemological distinction, developmental perspectives, like clinical perspectives, focus primarily on teacher growth; hence, developmental perspectives were considered a type of (or special case of) the clinical perspective for purposes of the present study.

Related Studies

At least three previous studies investigating some aspect of the content of textbooks for educational supervision have been conducted. Two of these studies (Bruce & Grimsley, 1987; Pajak, 1990) assessed the books for content relative to educational supervision generally whereas the remaining study (Anderson, 1993) looked specifically at content relevant to clinical supervision of instruction. Hence, there has been a relative dearth of studies on the content of supervisory textbooks specifically as regards varying theoretical perspectives of instructional leadership. Considering the importance of instructional leadership to the overall mission of schooling, it is surprising that only one study has been devoted to this purpose and that that study looked only at references to one model of instructional supervision.

In the one study addressing textbook content relative to clinical supervision, Anderson (1993) found that only about half of the 24 supervision textbooks he surveyed contained little if any mention of clinical supervision. Other books included fairly lengthy sections or chapters on the topic, although for the most part these books viewed clinical supervision as an innovative or avant-garde approach to supervision that competes with the larger work of the field. Further, although the Anderson study provided some insight into the specific content of typical supervision textbooks, it did not address the degree to which various theoretical perspectives are dominant in the books evaluated.

Method

Sample

The sample of textbooks utilized in the present study was selected based on a two-stage process. First, the two researchers examined their own personal libraries and the library at The University of Southern Mississippi to formulate an initial list of books. Books were included if they were published within the last 15 years and if they

were focused in their entirety on *instructional* supervision as opposed to educational supervision more generally. Second, colleagues who teach courses in instructional supervision at several other universities were consulted to determine what textbooks they currently used in their courses and/or books that they recognized as current leading works in the field. Based on this two-stage process, 12 instructional supervision textbooks were identified for inclusion in the sample. The selected textbooks included *Supervision of Instruction: A Developmental Approach* (Glickman, Gordon, & Ross-Gordon, 1995); *Supervision for Today's Schools* (Oliva & Pawlas, 1997); *Clinical Supervision: Coaching for Higher Performance* (Anderson & Snyder, 1993); *Supervisory Leadership: Introduction to Instructional Supervision* (Glatthorn, 1990); *Supervision: Focus on Instruction* (Beach & Reinhartz, 1989); *Evaluation-Based Leadership: School Administration in Contemporary Perspective* (Glasman, 1986); *Supervision in Education: Problems and Practices* (Tanner & Tanner, 1987); *Supervision: A Guide to Practice* (Wiles & Bondi, 1991); *School Leadership and Instructional Improvement* (Duke, 1987); *Instructional Supervision: A Behavior System* (Alfonso, Firth, & Neville, 1981); *Techniques in the Clinical Supervision of Teachers: Preservice and Inservice Applications* (Acheson & Gall, 1992); and *Handbook for Effective Supervision of Instruction* (Neagley & Evans, 1980). The sample comprised seven books with general titles of supervision, three books with the term "clinical supervision" included in the title, and one book described as a "developmental approach" to supervision. Whereas titles suggested contents, actual discourse was utilized to determine the theoretical perspectives underlying the subject matter.

Procedures

Three broad questions were used to determine the dominant supervisory orientation of the 12 textbooks included in the sample:

1. What theoretical perspectives of supervisory practice fit the specific definition of supervision as presented in each textbook?
2. What is the *dominant* theoretical perspective in each textbook?
3. To what extent does the textbook focus on teacher growth and, alternately, on teacher evaluation?

Additionally, in reviewing the content of each textbook, certain key words and phrases inherently linked in the literature to either evaluation-based and clinically-based supervision models were identified and the authors' stated theoretical approaches to supervision were acknowledged. This information was used to complement the information gained via responses to the three broad questions noted above.

In coding each textbook, the following procedures were utilized. First, a search was made throughout each text for the author's definition of supervision to determine the primary theoretical orientation underlying the definition. Second, the table of contents was examined to determine the various topics that constituted the author's conceptualization of supervision. Finally, as previously noted, each text was perused throughout to determine the vocabulary most frequently used in discussing the supervisory process. Following these procedures, in order for a text to be labeled as "evaluation-based" or "clinically-based" a very clear bent toward one direction or the other was required. It was predetermined that any work that included appreciable ratios of each orientation would be labeled as "mixed" in their theoretical orientation.

Results

Of the 12 supervisory texts included in the study, five (i.e., Acheson & Gall, 1992; Anderson & Snyder, 1993; Beach & Reinhartz, 1989; Glickman et al., 1995; Tanner & Tanner, 1987) were found to be predominantly clinically-based in their theoretical perspective, and seven (i.e., Alfonso et al., 1981; Duke, 1987; Glasman, 1986; Glatthorn, 1990; Neagley & Evans, 1980; Oliva & Pawlas, 1997; Wiles & Bondi, 1991) were more evaluation-based. None of the texts was truly "mixed" in its orientation following the criteria defined above. Two of the five clinically-based texts (Glickman et al., 1995; Tanner & Tanner, 1987) specifically targeted developmental supervision, whereas another (Beach & Reinhartz, 1989) included discussion of both clinical and evaluative strategies, with clinical strategies, obviously, predominating. The seven evaluation-based books included discussions of both clinically- and evaluation-based approaches to supervision, with an obvious tendency toward evaluation-based approaches.

The predominant theoretical orientations of two of the texts considered clinical in perspective (Beach & Reinhartz, 1989; Tanner & Tanner, 1987) were particularly difficult to determine. Although Beach and Reinhartz included a chapter on evaluation (Chapter 10), they begin the chapter with an almost apologetic rationale for its inclusion in the book: "Historically, most supervisors, unless they served in a direct line of authority (that is, as principal), have shunned the role of evaluator because they viewed this function as counterproductive to promoting staff development" (p. 209). Moreover, despite the inclusion of this chapter in their book, Beach and Reinhartz generally seemed to be much more clinical in focus. In fact, their chapter on the "mechanics of supervision" (Chapter 9) follows a highly clinical pattern, with the supervisory process described in terms of observation, data collection and analysis, and conferencing.

The Tanner and Tanner text was difficult to categorize due to the authors' painstaking attempt to offer a balanced approach and to the highly descriptive (as opposed to prescriptive) tone of the text. However, the authors make isolated statements (e.g., pp. ix, 503) that the text is developmental in focus. If beginning with the final chapter ("How to Judge a School"), it would seem there is a slight bent toward the evaluation-based perspective. However, the clear tendency throughout the text of decrying extremist production-efficiency approaches to supervising schools, along with a tacit preference for developmental strategies, throws the balance in favor of a more clinically-based (i.e., developmental) orientation.

Evaluation-based Texts

Of the seven texts characterized by this orientation, the Glasman and Alfonso et al. texts were most clearly the exemplars of the cohort. Certainly, the titles alone (*Evaluation-Based Leadership: School Administration in Contemporary Perspective* and *Instructional Supervision: A Behavior System*, respectively) indicate a tendency toward accountability and measurable behavioral outcomes. Other books, though not as clearly evaluation-based in their title or even their general tone, were included in this category based on certain underlying assumptions of the authors. For example, Neagley and Evans (1980) favor a competency-based "systems model" of supervision, even though they are also willing to consider elements of the clinical process. Duke (1987) tends with some consistency to tag together "supervision and evaluation" with some regularity, even when discussing concepts of the clinical process. Likewise, Wiles and Bondi (1991), in defining the supervisor's role as a school leader, lean toward a process-product orientation:

> Supervision leadership involves thinking, planning, organizing, and evaluating processes. The thinking and planning phases of improving instruction are most like policy formation and administration. Organizing instructional programs is most like those functions of curriculum where a translation of ideas into programs occurs. The evaluation functions of supervision are usually directed toward the instructional activities of the institution. (p. 30)

Although Oliva and Pawlas (1997) claim an "eclectic approach" (p. 12) to defining supervision, the topic of evaluation-oriented activities comes up with a good deal of frequency throughout the text even when discussing ideas that usually are not evaluative in focus. For example, Oliva and Pawlas present clinical supervision as a type of formative evaluation. Wiles and Bondi (1991), in providing a list of eight areas of supervisory competence, complete their list with the competency "Supervisors are Evaluators", which is described as follows: "The previously stated roles, collectively, place the supervisor in a constant evaluation position. Assessing teacher performance, program outcomes, texts and materials, consultant performance, and analysis of testing results--all are part of the evaluation role" (p. 23). Though more truly eclectic in nature, Glatthorn's (1990) perspective on instructional supervision also leans toward evaluation, considering, for instance, that (a) "teacher evaluation" is one of his four main foci of supervisory leadership and (b) the contextual grounding for the entire book (Chapter 1) is the effective schools movement, which, as previously noted, has frequently been linked to evaluation-based perspectives.

Clinically-based Texts

Obviously, the two texts (Acheson & Gall, 1992; Anderson & Snyder, 1993) with the term "clinical" in their titles exemplified a clinically-oriented perspective. Interestingly, however, the Anderson and Snyder text, the only edited anthology of the 12 books in the sample, includes a section (three chapters) on teacher evaluation, with a focus on how clinically-based supervision methodologies might be merged with traditional evaluation systems. Acheson and Gall's work focuses on effective data gathering and reporting techniques for use in non-evaluative clinical settings. The only mentions of evaluation in Acheson and Gall's book are for the purpose of distinguishing clinical supervision from evaluation.

As previously noted, Beach and Reinhartz have assumed a veiled clinical focus in their work whereas Glickman et al. (1995) and Tanner and Tanner (1987) are more assertive in their developmental orientations. For example, Glickman et al. concluded that the primary objective of supervision should focus on reflecting and autonomy of teachers assisted by nondirective supervision. Likewise, in explaining the orientation of their text, Tanner and Tanner noted:

> . . . when teachers see supervision as a developmental process of cooperative effort toward diagnosing and solving substantive problems in the classroom and school, they are more likely to seek supervisory assistance and their professionalism is enhanced enormously. The professionalization of supervision and of teaching hinges on the elimination of deficiency-oriented supervision in favor of a developmental approach. (pp. ix-x)

Discussion

The foregoing results suggest that the evaluation- and clinically-based theoretical orientations of instructional supervision are represented relatively equally in the textual materials presently available for training instructional supervisors, indicating that both orientations are alive and well. Moreover, considering that most of the texts are, to some degree, eclectic in nature, it is highly likely that most textbook readers will be exposed to more than one viewpoint even in those cases in which a bias toward a particular viewpoint exists. Thus, it is likely that both clinically- and evaluation-oriented perspectives will continue to have an impact on supervisory practice, and that many new educational supervisors may emerge from their training with an eclectic perspective of their supervisory role; however, it is likely that this eclectic perspective will be most likely communicated to students if they are assigned readings from several different books and/or if the instructor includes multiple perspectives in class discussions and activities.

Of the 12 texts included in the sample, seven were found to be evaluation-based. The continued emphasis on the evaluation-based perspective is healthy in that this perspective has served the profession well for many years. Every teacher is not effective in every teaching situation; however, there are certain generic teaching competencies that all teachers should possess. The evaluation-based perspective on the supervision of instruction assumes that good supervisors can coordinate the activities of instruction toward measurable individual and systemic competencies. In systems that are heavily accountability driven, these leadership skills are essential.

That five of the 12 books were clinical in their theoretical perspective is also noteworthy. In supervisory settings focusing on clinical (including developmental) perspectives, the goals of instructional supervision are mutually determined by the supervisor and the instructor and teaching may be viewed as a science with definitive behaviors to be sharpened via reflective practice (developmental approach) or as an art that promotes creativity and boundless strategies for reaching all students (traditional clinical approach). In today's climate of participatory management, supervisors would be wise to consider the inculcation of clinical strategies into their supervisory repertoire.

It will be interesting to see whether these two orientations will continue to dominate the literature used in training instructional supervisors. Wiles and Bondi (1991) have noted that as political, social, and economic forces shift, so do perceptions of instructional supervision:

> We can envision a time in the near future when social, economic, and political conditions in the United States would allow schools to redefine their mission and enter another developmental stage [of supervisory practice]. If this should occur, supervision and the role of supervisors would once again be transformed to meet the changing conditions in schools. (p. 11)

Future research studies in this area should include additional content analyses of textbooks every few years to determine whether certain supervisory orientations will continue to dominate the literature and whether newer perspectives will emerge. Further, researchers might wish to conduct interviews with instructional supervisors and their supervisees to determine the degree to which various theoretical orientations are reflected in actual supervisory practice. Finally, ethnographic inquiries into the specific supervisory activities of instructional supervisors are also warranted. Such inquiries would be useful in identifying how supervisors' theoretical perspectives are linked to specific supervisory actions.

References

Acheson, K. A., & Gall, M. D. (1980). *Techniques in the clinical supervision of teachers.* New York: Longman.

Acheson, K. A., & Gall, M. D. (1992). *Techniques in the clinical supervision of teachers: Preservice and in-service applications* (3rd ed.). New York: Longman.

Alfonso, R. J., Firth, G. R., & Neville, R. F. (1981). *Instructional supervision: A behavior system.* Boston: Allyn and Bacon.

Anderson, R. H. (1993). Clinical supervision: Its history and current context. In R. H. Anderson & K. J. Snyder (Eds.), *Clinical supervision: Coaching for higher performance* (pp. 5-18). Lancaster, PA: Technomic.

Anderson, R. H., & Snyder, K. J. (Eds.) (1993). *Clinical supervision: Coaching for higher performance.* Lancaster, PA: Technomic.

Beach, D. M., & Reinhartz, J. (1989). *Supervision: Focus on instruction.* New York: Harper & Row.

Bruce, R. E., & Grimsley, E. E. (1987). *Readings in educational supervision* (vol. 2). Alexandria, VA: Association for Supervision and Curriculum Development.

Cogan, M. (1973). *Clinical supervision.* Boston: Houghton Mifflin.

Duke, D. L. (1987). *School leadership and instructional improvement.* New York: Random House.

Glasman, N. (1986). *Evaluation-based leadership: School administration in contemporary perspective.* Albany, NY: State University of New York Press.

Glatthorn, A. A. (1990). *Supervisory leadership: Introduction to instructional supervision.* New York: Harper Collins.

Glickman, C. D. (1981). *Developmental supervision: Alternative practices for helping teachers improve instruction.* Alexandria, VA: Association for Supervision and Curriculum Development.

Glickman, C. D., Gordon, S. P., & Ross-Gordon, J. M. (1995). *Supervision of instruction: A developmental approach* (3rd ed.). Boston: Allyn and Bacon.

Goldhammer, R. (1969). *Clinical supervision: Special methods for the supervision of teachers.* New York: Holt, Rinehart and Winston.

Gonzales-Baker, M., & Tinagero, J. V. (1983, February). *Improving bilingual education through clinical and developmental supervision of bilingual education teachers and student teachers: A theoretical framework.* Paper presented at the annual meeting of the International Bilingual Bicultural Conference of the National Association for Bilingual Education, Washington. (ERIC Document Reproduction Service No. ED 228 229)

Gordon, S. P. (1990). Developmental supervision: An exploratory study of a promising model. *Journal of Curriculum and Supervision, 5,* 293-307.

Harris, B. M. (1985). *Supervisory behavior in education* (3rd ed.). Englewood Cliffs, NJ: Prentice Hall.

Krey, R. D., & Burke, P. J. (1989). *A design for instructional supervision.* Springfield, IL: Charles C. Thomas.

Lovell, J. T., & Wiles, K. (1983). *Supervision for better schools* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Mosher, R. L., & Purpel, D. E. (1972). *Supervision: The reluctant profession.* Boston: Houghton Mifflin.

National Education Association. (1931). The fourth yearbook of the department of supervisors and directors of instruction of the National Education Association. *Evaluation of supervision.* New York: Teachers College Press.

Neagley, R. L., & Evans, N. D. (1980). *Handbook for effective supervision of instruction.* Englewood Cliffs, NJ: Prentice-Hall.

Oliva, P. F., & Pawlas, G. E. (1997). *Supervision for today's schools* (5th ed.). New York: Longman.

Pajak, E. (1990, April). *Identification of dimensions of supervisory practice in education: Reviewing the literature.* Paper presented at the annual meeting of the American Educational Research Association, Boston. (ERIC Document Reproduction Service No. ED 320 285)

Tanner, D., & Tanner, L. (1987). *Supervision in education: Problems and practices.* New York: Macmillan.

Wiles, J., & Bondi, J. (1991). *Supervision: A guide to practice.* New York: Merrill.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 1, 33-40

University Teachers' Perceived Usefulness of Computer-Assisted Instruction

Yukiko Inoue
The University of Guam

Human resources and information technologies are two of the most important cornerstones in the government's developmental policy in Singapore. This study investigated the process by which university teachers arrived at the positive perception of computer-assisted instruction (CAI) for teaching. The CAI survey was distributed to 118 randomly selected (education and business) teachers at one of two leading universities in Singapore; 63 teachers (53%) responded. This study revealed that the knowledge of CAI was a dominant factor influencing the perceived usefulness of CAI : the higher the CAI knowledge level is, the more positive the perception of CAI will be. This finding may be an indication of Singapore's favorable climate for instructional technology use, yet the world's educational communities should take note of it.

Computer-assisted instruction (CAI) is not a new approach. Computers have been utilized in instructional settings since the late 1950s. Interactive educational materials, as noted by Sheldon (1995), began to evolve as soon as microcomputers became available from Apple, Commodore, and Tandy in 1977. In particular, since the IBM-PC was introduced in 1981, personal computers have developed into powerful and inexpensive machines capable of multimedia presentations using graphics, animation, audio, and interactive video. As computers have become smaller, more powerful, and more cost-effective, their use in educational settings has increased rapidly. CAI is currently being used to teach students and to train teachers as well. The field of CAI covers all kinds of instructional systems in which computers are used to support teaching and learning. If properly designed and implemented, CAI has many advantages. For instance, CAI offers an interactive response, immediate feedback, infinite patience, and motivation, as well as an ability to maintain accurate records of student progress (Sloane, Gordon, Gunn, & Mickelsen, 1989).

In this modern and technological age, CAI will be a powerful complement to traditional teaching methods. Teachers can help keep students' interest by using CAI to provide a variety of instructional methods and presentations in the class. For the past three decades, nevertheless, CAI has made only minor progress in the classroom implementation. Why? Marcinkiewicz (1993/94) believed one of the major reasons is that many teachers are reluctant to adopt CAI, even though they strongly believe that using computers can improve the quality and the quantity of their teaching. In this regard, Marcinkiewicz likened this situation to corn farmers who do not uniformly adopt a new corn seed that can improve the amount and the quality of their yields. That is to say, like any other technological innovations when used for teaching purposes, CAI needs to be accepted by teachers before it can be utilized productively and effectively (Mackowiak, 1990/91).

Selecting Singapore as a Site for the Study

The Republic of Singapore is a small country (27 million people), where almost the entire educational system is public and is under the control of government agencies. According to *A Vision of an Intelligent Island: The IT2000 Report* (NCB, 1992), over half of Singapore's teachers will have received intensive training in instructional technology by the year 2000. Human resources and information technologies (IT) are two of the most important cornerstones in the government's developmental policy. If everything goes well, Singapore will be one of the first countries in the world (probably the *first* one) to have the national information infrastructure capable of connecting computers in every home to office, school, or factory (Sponder & Hilgenfeld, 1993). Singapore is interested in IT for various reasons, such as teacher shortage, school attrition, and quality teaching, in addition to the recognition of the importance of general education. A project called the "Student Teacher Workbench" (STW) is part of the government's IT2000 plan to exploit IT for competitive advantages (Hu & Hoon, 1994). The STW is a one-stop electronic shop, which allows teachers and students at elementary and secondary schools to gain access to a wide range of information, reference

Yukiko Inoue is Assistant Professor of Educational Research in the College of Education at the University of Guam, Mangilao, Guam. Correspondence regarding this article should be sent to Yukiko Inoue at College of Education, UOG Station, Mangilao, Guam 96923 U.S.A. or by e-mail at yinoue@uog.edu.

documents, and self-study materials. It is also a tool to communicate with others located in different parts of the world. CAI will be used in the STW for enhancing self-learning in the classroom and from a distance.

Extensive experiments on the STW implementation, mainly with respect to secondary and higher education, are now being conducted by the National Computer Board (NCB) and by the National Institute of Education (NIE). The NIE is the sole college of education in Singapore. The Instructional Science Division of the NIE is involved in the ongoing development of programs and methods to improve the teaching and learning process, primarily through the use of educational technology (Sponder & Hilgenfeld, 1993). In order to prepare teachers to cope with advances in IT, the NIE established IT courses in 1986. Teachers are learning how to use computerized data banks, to extract materials from various electronic sources, and to use multimedia to create lessons by mixing words, pictures, and sounds. Pre-service teachers are required to take the thirty-hour IT courses. Software developments for CAI are now being conducted at Nanyang Technological University (NTU), the National University of Singapore (NUS), and four Polytechnic institutions.

In Singapore, of course, the use of CAI is not new. In 1986, twenty-two networked computer programs were started in a secondary school (Barker, 1988; Yip & Sim, 1990). The mission of the programs was to evaluate CAI functions (i.e., tutoring, simulation, and problem solving) and to examine the effectiveness of the network of the CAI delivery mechanism. Published studies on the effect of CAI in Singapore's schools have shown positive results in such courses as mathematics (Ong & Lee-Leck, 1986), geography (Low, 1988), and geometry (Woo-Tan, 1989). Effects of CAI have been studied in many ways, yet little attention has been given to a fundamental issue, namely, how Singapore's university teachers feel about using CAI for teaching, though it is viewed that faculty members generally have positive perceptions of CAI (Kluever, Lam, Hoffman, Green, & Swearingen, 1994). Such perceptions must have been cultivated by the belief that computers are necessary as well as useful for teaching (and faculty development) especially in technologically sophisticated societies of today. It is thus relevant to investigate which factors are significantly associated with the university teacher's perceived usefulness of CAI, in particular in Singapore's favorable academic climate for instructional technology.

Method

Model

The factors influencing the university teacher's perception of CAI were theorized in the path-analytic model. The model evaluated the factors for the teacher's perception of CAI as a function of his or her background characteristics (gender, age, and teaching discipline), teaching experience, and knowledge of CAI. Although each variable in the model (for the definitions of operational variables, see Table 1) was hypothesized to influence the perception of CAI, the manner in which this influence was exerted was expected to differ. Figure 1 depicts the structure of the model, navigating the hypothesized process by which the teacher perceives the usefulness of CAI. Each of the straight lines within the model represents the hypothesized direct influence of one variable on another with the arrow indicating the direction of influence. Each of the curved lines within the model represents that the two variables are related but no causal relationship is hypothesized between them.

Table 1
Definitions of Operational Variables

Variable	Definitions
1. Gender	A single item based on the teacher's gender: 1 = Male; 2 = Female.
2. Age	A single item based on the teacher's age (in years): 1 = 30 or less; 2 = 31 to 45; 3 = 46 or over.
3. Discipline	A single item based on the teacher's teaching discipline: 1 = Business; 2 = Education.
4. Degree	A single item measuring the teacher's highest academic degree with three categories: 1 = Doctorate; 2 = Master; 3 = Bachelor or Associate.
5. Knowledge	A single item assessing the teacher's level of CAI knowledge with five categories: 1 = know nothing about CAI; 2 = know a little about CAI; 3 = have read about CAI but have never experienced it; 4 = have seen demonstrations of CAI developed by others; 5 = have developed CAI myself.
6. Experience	A single item measuring the teacher's teaching experience (in years) at college or university levels with four categories: 1 = 5 or less; 2 = 6 to 15; 3 = 16 to 25; 4 = 26 or more.
7. User/Nonuser	A single item measuring that the teacher has been using CAI for teaching: 1 = user; 2 = nonuser.
8. Perception of CAI	A single item measuring the teacher's perceived usefulness of CAI with five categories: 5 = extremely useful; 4 = very useful; 3 = useful; 2 = somewhat useful; 1 = not useful at all.

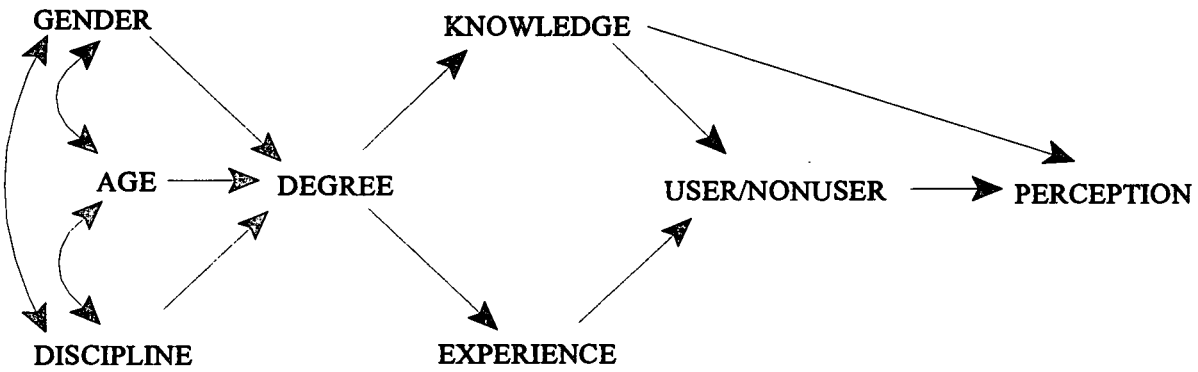

Figure 1. The Theoretical Model of University Teachers' Perceived Usefulness of CAI

Variables

The first block of variables in the model comprises exogenous variables that represent the teacher's gender, age, and teaching discipline (education versus business). While research has shown a significant difference between males and females in their attitudes toward computers, that is, males are more likely than females to have confidence in using computers (e.g., Kirk, 1992), research has shown the opposite result (e.g., Juska & Paris, 1993). In fact, there are few substantive empirical data on how male and female teachers differ regarding instructional technology. Young people are generally expected to favor CAI, but this may not be the case in technological-oriented countries such as Singapore. Singapore's faculty members are fairly young (mandatory retirement age is 55), and the government is supportive for the widespread availability of computers at schools and universities. As regards teaching discipline, it is viewed that business faculty members are technology-oriented, whereas education faculty members are behavior-oriented.

Subsequent variables in the model are endogenous, hence dependent on prior variables. The model also specifies no caused relationships between variables within the same block. The second block of variables is the highest academic degree the teacher attained. This variable came from Singapore's CAI experts, who strongly suggested that it might be interesting to investigate how academic degree would be related to the perception of computerized technologies such as CAI. Faculty members with higher academic degrees are usually tenured, thus they may be less afraid of losing their jobs because of the involvement in CAI or any innovative "teaching machines."

The third block of variables in the model considers the teacher's teaching experience (the number of years in teaching) and knowledge level of CAI. The curriculum reform (including teaching with technology) is a site of ongoing negotiation and contestation among university teachers (Tierney, 1995). Many educational innovations have failed because they have never been adopted into traditional instruction (Dynan, 1984). Perhaps teachers' resistance to change is the constraining factor for the use of CAI. In short, the more knowledge of CAI teachers have, according to Smith and Yellen (1991), the more they perceive CAI as useful for teaching.

The fourth block of variables in the model considers whether or not the university teacher is currently using CAI for teaching. Dupagne and Krendl (1992) found that CAI users perceived the usefulness of CAI as significantly greater than did nonusers. However, Singapore may be different from other cases because of the constant publicity given to the importance of instructional technology advertised through daily media. The last variable in the model represents the university teacher's perception of CAI.

Participants

Participants for this study were drawn from two of the entities at Nanyang Technological University (NTU), which is one of two leading universities in Singapore: (1) the division of Education, which is part of the National Institute of Education (NIE), and (2) the Nanyang Business School (NBS). The target population of the study was all the teachers from these two entities at NTU. The division of Education had 59 teachers (57% female; 43% male) and the NBS had 208 teachers (25% female; 75% female). All the teachers of the NIE received a copy of the survey questionnaire. In order to match the sample size of the NIE, 59 teachers of the NBS were randomly

selected to participate in the survey. Of the 118 participants, 63 (53%) responded; accordingly, the study reported here was based on the data from those 63 respondents representing the two groups: 27 were from the NIE and 36 were from the NBS.

Procedure

The preliminary survey questionnaire (consisting of determinants for the use of CAI, perceptions of CAI, open-ended comments on CAI, and demographic information) was pilot-tested through personal interviews with four teachers at NTU (two users and two non-users of CAI). Based on their comments and suggestions, the final version of the survey questionnaire was constructed. The primary issue of the survey was to identify the relative importance of facilitators and inhibitors for the use of CAI by university teachers. The secondary issue was exploratory, aiming to identify the significant relationships between the variables and the teachers' perceived usefulness of CAI. Only the results of the secondary issue are reported in this paper.

One faculty member from each entity agreed to serve as a contact person for the survey. The two contact persons received a packet containing an explanatory letter, directions for administering the survey, and 59 copies of the questionnaire. The survey was conducted with the permission of the dean of each school at NTU. Each contact person distributed, collected, and returned all questionnaires to the researcher. Table 2 presents the demographic information of the study participants: for example, (a) more than 50 percent of the respondents belong to the age group 31-45 years old, (b) education teachers have higher academic degrees than do business teachers, and (c) business teachers are more CAI-oriented than are education teachers. In addition, as seen in Table 3, the participants show positive perception of CAI. Such a positive attitude, as stated before, could have been cultivated by the belief that computers are necessary for educational instructions in the technological sophisticated world.

Analysis

Path analysis (a research method for presenting a causal model in which a series of independent variables is used to predict a series of dependent variables) was used in this study to estimate the influences on the perception of CAI. The model was estimated with ordinary least-squares procedures, using a computer program called GEMINI by Wolfle and Ethington (1985). This program computed indirect effects and their standard errors in addition to the usual regression results. All paths were estimated to test whether the paths hypothesized to be zero were nonsignificant. Direct effects are represented by regression coefficients, either standardized (beta weights) or unstandardized (b weights), interpreted in the usual manner. Indirect effects (the sums of the products from the direct effect through intervening variables in the model) represent the influences on the dependent variable that are the results of directly influencing prior causal variables in the model. Thus direct and indirect effects implied by the model were estimated from means, standard deviations, and correlations (see Table 4) among all the variables in the proposed model. Finally, a large sample is desirable when estimating path models using the GEMINI program but this study was conducted as an attempt for this particular sample size.

Table 2
Demographic Characteristics of all the Respondents

	Education $N = 27$	Business $N = 36$	Total $N = 63$
Gender:			
Male	11	25	36 (57%)
Female	16	11	27 (43%)
Age (in years):			
30 or less	4	3	7 (11%)
31 to 45	10	23	33 (52%)
46 or over	13	10	23 (37%)
Academic Degree:			
Doctorate	23	24	47 (75%)
Master	4	11	15 (24%)
Bachelor	0	1	1 (1%)
User or Nonuser of CAI:			
User	10	17	27 (43%)
Nonuser	17	19	36 (57%)
College teaching (in years):			
5 or less	7	16	23 (37%)
6 to 15	13	16	29 (46%)
16 or more	7	4	11 (17%)

Table 3
The Study Participants' Perceived Usefulness of CAI

Extremely useful	12
Very useful	21
Useful	23
Somewhat useful	6
Not useful at all	1
Total	63

Results

The estimated coefficients of the five equations defining the path model described above are given in both standardized and metric (unstandardized) forms in Table 5. The last column of the table indicates the direct effects of each variable in the model on the perception of CAI. In addition to the direct effects, the table shows that the seven variables in the model explain 19.60 percent (R-squared of .196 is not overwhelmingly large, but in a way, it might be sufficient in this study) of the variance in the perception of CAI. Figure 2 illustrates diagrammatically the significant direct effects represented in Table 5. Only standardized effects greater than .05 are considered of substantial importance (Land, 1969; Pedhazur, 1982). In this regard, all the effects that are statistically significant in this study, as seen in the table, are fairly large.

Table 4
Means, Standard Deviations, and Correlations for the Variables in the Model of University Teachers' Perceived Usefulness of CAI ($N = 63$)

	1	2	3	4	5	6	7	8
1. Gender	--							
2. Age		. 007	--					
3. Discipline	.287	.157	--					
4. Degree	.115	-.068	-.220	--				
5. Knowledge	-.053	-.064	.053	-.015	--			
6. Experience	-.224	.463	.223	.003	.173	--		
7. User/Nonuser	.102	.143	.102	.019	-.426	.013	--	
8. Perception	-.015	.102	.088	-.025	.393	.035	-.225	--
Mean	1.429	2.254	1.429	1.270	3.444	10.190	1.571	3.571
Standard Deviation	.499	.649	.499	.482	1.215	7.971	.499	.949

Table 5
Direct Effects on University Teachers' Perceived Usefulness of CAI

	Dependent Variables				
Variables	4	5	6	7	8
1. Gender	.193	-.079	-.321*	.057	-.054
	(.186)	(-.193)	(-5.131)	(.057)	(-.103)
2. Age	-.027	-.077	.430*	.084	-.199
	(-.020)	(-.145)	(5.307)	(.065)	(-.292)
3. Discipline	-.277*	.090	.276*	.093	.102
	(-.262)	(.218)	(4.411)	(.093)	(.194)
4. Degree		.009	.130	.032	.025
		(.022)	(2.150)	(.033)	(.050)
5. Knowledge				-.429*	.386*
				(-.176)	(.301)
6. Experience				.040	-.158
				(.003)	(-.019)
7. User/Nonuser					-.092
					(-.176)
8. Perception of CAI					
R-squared	.084	.333	.134	.211	.196

Note: Metric (unstandardized coefficients are given in parentheses); *$p < .05$

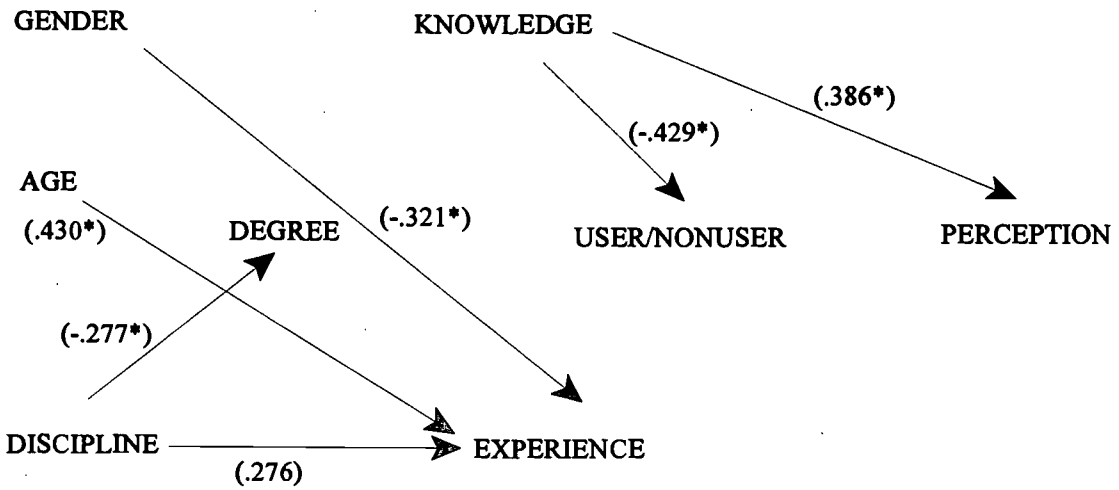

Figure 2. The Estimated Model of University Teachers' Perceived Usefulness of CAI. $*p < .05$

Direct Effects

Academic Degree. Teaching discipline (beta = -.277) was the only one that had a significant direct effect on academic degree for this sample of teachers. The negative effect indicates that teachers in education tend to have higher academic degrees than do teachers in business.

Knowledge of CAI and Teaching Experience. Contrary to the hypothesized model, academic degree had no direct influence on either knowledge of CAI or on teaching experience. Instead, all three exogenous variables indicated significant direct effects on teaching experience in the following order of magnitude: age (beta = .430), gender (beta = -.321), and teaching discipline (beta = .276). Perhaps it is a matter of course that age is the most influential effect on teaching experience. And it should be noted that the use of teaching experience (even though focused only on the college teaching experience) and age may have created a situation with multicolinearity. The negative effect in gender reveals that female teachers are more likely than male teachers to have fewer years of teaching experience at colleges and universities.

User or Nonuser of CAI. Although the theoretical model posited that the dominant influence on user versus nonuser of CAI should come from knowledge of CAI and teaching experience, knowledge (beta = -.429) was the only variable that indicated a significant direct effect as to whether the teacher was using CAI or not. This strong negative effect explains that the current user is more likely than the nonuser to have knowledge of CAI, and perhaps this is natural. Surprisingly enough, however, teaching experience (as defined by the number of years in college teaching) shows a nonsignificant direct effect on user or nonuser of CAI.

Perceived Usefulness of CAI. As hypothesized by the theoretical model, the direct influence on the perceived usefulness of CAI was knowledge of CAI (beta = .386). The theoretical model posited that dominant influences should come from the variables such as knowledge of CAI and user versus nonuser of CAI. Nevertheless, user versus nonuser had no significant direct influence on the perception of CAI. The knowledge of CAI was the only variable influencing the perception of CAI, and its effect is fairly strong, supporting the findings of other studies, for instance, by Smith and Yellen (1991) and by Spotts and Bowman (1993). Thus the higher the knowledge of CAI, the higher the positive perception of CAI will be; and, the importance of the knowledge level of CAI cannot be underestimated. Yet the model in this study is not the definitive one and, of course, other important variables may also be related to the perception of CAI. For example, the availability and speed of commercial hardware, the availability of software information, the support of higher administration, the integration of CAI with the university's goal, the suitability of appropriate classrooms, and the systematic educational planning for CAI should be explored in future studies.

Indirect Effects

Table 6 presents the indirect effects of the variables in the model on the perception of CAI. None of the variables have significant indirect effects. This is because, for this sample of teachers, there is no significant direct path (a) from academic degree to knowledge of CAI, (b) from academic degree to teaching experience, (c) from teaching experience to user versus nonuser of CAI, or (d) from user versus nonuser of CAI to perception of CAI, respectively. In short, these nonsignificant results may simply be due to the lack of power.

Table 6
Indirect Effects on University Teachers' Perceived Usefulness of CAI

	Standardized	Metric (Unstandardized)
Gender	.014	(.026)
Age	-.110	(-.162)
Discipline	-.016	(-.031)
Degree	-.020	(-.040)
Knowledge	.040	(.031)
Experience	-.004	(-.0004)

Open-Ended Comments on CAI

The study respondents' general comments on CAI include: (a) university teachers need to underpin the use of CAI with higher levels of teaching skills than they generally have, (b) CAI must be greatly improved as an instructional tool in higher education, (c) CAI is not a solution for all university courses, therefore promoting CAI implementations by using a top-down approach should be avoided, (d) CAI may be suitable for practical studies more than for theoretical studies, (e) CAI is very important since it can enhance students' critical and analytical thinking, and (f) the fact that older students tend to have less computer experience makes CAI implementation more difficult in education.

Conclusion

The results of this study have indicated that the knowledge of CAI is a dominant factor influencing the perceived usefulness of CAI for this sample of university teachers in Singapore, with no statistically significant effects coming from age, gender, or teaching discipline. This finding may be an indication of Singapore's favorable climate for instructional technology. The fact that the higher the CAI knowledge level, the more CAI is perceived to be useful is not surprising at all. One implication of this finding is that if an academic institution wants its teachers to utilize CAI, the institution should increase all the teachers' knowledge level of educational technology. This is exactly what the Ministry of Education in Singapore is doing by requiring school and university teachers to learn about computer skills in general and CAI skills in particular. Although further exploration is necessary to examine the impact of the university teacher's perception on the use of CAI, the above-mentioned finding can be treated with more confidence and used by both the School of Business and the College of Education at NTU for educational planning purposes.

In this technological age, the integration of CAI with conventional instruction is one alternative approach of teaching in higher education. Countries like Singapore are making major investments in the information superhighway toward the twenty-first century. One objective of Singapore's government is to increase the use of instructional technologies such as CAI. The results of this study may be attributed to Singapore's unique climate for the instructional technology use at schools and universities, yet the world's educational communities should take note of the finding.

It is important to note, however, that comparing users and nonusers of CAI was very important in this study, and it was necessary to ask the participants to indicate to which category they belonged. Talisayon (1990) described three kinds of computer uses: (a) as a learning tool, (b) as itself the object of study, and (c) as a planning and management tool. In addition, the terms "the use of CAI" or "computer use" have broad definitions, and each participant may have had a different definition in mind. Thus it was not clear to what extent the participants used CAI when reporting their CAI use. More important, it would have been better to have used multiple items to determine construct measures rather than single items. Accordingly, no attempt was made to generalize to the population of this study, therefore, the results of the estimation of the model for the study were considered unique to this sample of Singapore's teachers.

Finally, it is needless to say that CAI is not suitable for all college or university courses. As stated before, the integration of CAI with traditional teaching methods would be better in higher education. It should be noted, however, that the twenty-first century students must master sophisticated information age learning media (i.e., tool software such as graphics, database, spreadsheet, multimedia, and Internet) and must become accountable for demonstrating learning more directly than did students of the past. By the same token, teachers themselves must have and maintain sufficient knowledge and skills in instructional technology (though hardware and software change rapidly). Eventually, the option of mixing CAI and conventional instruction is an intriguing topic and a fruitful area for educational research.

References

Barker, P. (1988). Computer assisted learning in Singapore. *British Journal of Educational Technology, 19*(3), 193-201.

Dupagne, M., & Krendl, K. (1992). Teachers' attitudes toward computers: A review of the literature. *Journal of Research on Computing in Education, 24*(3), 420-429.

Dynan, M. B. (1984). Evaluation of implementation of curriculum innovations. In A. S. Ryan, & B. J. Fraser (Eds.). *Educational Evaluation for Program Improvement.*

Hu, R., & Hoon, P. M. (1994). Student's and teacher's workbench. *Information Technology - Journal of the Singapore Computer Society, 6*(1), 91-96.

Juska, A., & Paris, A. E. (1993). Students computer use: Its organizational structure and institutional support. *Collegiate Microcomputer, 11*(2), 42-50.

Kirk, D. (1992, April). Gender issues in information technology as found in schools: Authentic/synthetic/fantastic? *Educational Technology*, pp. 28-31.

Kluever, R. C., Lam, T. C. M., Hoffman, E. R., Green, K. E., & Swearingen, D. L. (1994). The computer attitude scale: Assessing changes in teachers attitudes toward computers. *Journal of Educational Computing Research, 11*(3), 251-261.

Land, K. C. (1969). Principle of path analysis. In E. F. Borgalt (Ed.). *Sociological Methodology.* San Francisco: Jossey-Bass.

Low, K. G. (1988). Investigation of various modes of computer-based learning for secondary schools. *Singapore Journal of Education, 9*(2), 27-32.

Mackowiak, K. (1990/91). The effects of faculty characteristics on computer applications in instruction. *Journal of Research on Computing in Education, 23*(3), 396-410.

Marcinkiewicz, H. R. (1993/94). Computers and teachers: Factors influencing computer use in the classroom. *Journal of Research on Computing in Education, 26*(2), 220-237.

National Computer Board [NCB]. (1992). *A vision of an intelligent island: The IT2000.* Singapore: Singapore National Printers.

Ong, S. T., & Lee-Leck M. K. (1986). *The effects of computer-assisted instruction on attitudes and achievement in mathematics on pre-service primary school teachers.* In Y. C. Yong, W. S. Low, & C. K. Lee (Eds.). *Selected proceedings of Educomp '86: Impact of microcomputers in education – issues and techniques* (pp. 196-201). Singapore: Ngee Ann Polytechnic.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research.* New York: Holt, Rinehart, and Winston.

Plomp, T., Pelgrum, W. J., & Steerneman, H. M. (1990). Influence of computer use on schools' curriculum: Limited integration. *Computer Educ., 14*(2), 159-171.

Sheldon, E. J. (1995). *CAI/multimedia approach to farm tractor and machinery safety certification.* Unpublished Doctor's Dissertation. Purdue University, Indiana.

Sloane, H. N., Gordon, H. M., Gunn, C., & Mickelsen, V. G. (1989). *Evaluating educational software: A guide for teachers.* Englewood Cliffs, NJ: Prentice Hall.

Smith, T. W., & Yellen, R. E. (1991). Factors inhibiting the introduction of computer-assisted instruction. *Collegiate Microcomputer, 9*(2), 117-122.

Sponder, B., & Hilgenfeld, R. M. (1993). Research in instructional technology at the National Institute of Education, Singapore. *ETR&D, 41*(4), 115-118.

Spotts, T. H., & Bowman, M. A. (1993). Increasing faculty use of instructional technology: Barriers and incentives. *Educational Media International [EMI], 30*(4), 199-204.

Talisayon, V. M. (1990). *Preparing ASEAN for the information century: A comparative study of policies and programs on computers in science and mathematics education.* University of the Philippines Press.

Tierney, W. (1995). Cultural politics and the curriculum in post-secondary education. In C. F. Conrad & J. G. Haworth (Eds.). *Re-visioning curriculum in higher education (ASHE reader series).* (pp. 35-47). Simon & Schuster.

Wolfle, L. M., & Ethington, C. A. (1985). GEMINI: Program for analysis of structural equations with standard errors of indirect effect. *Behavior Research Methods, Instrument, and Computers, 17*(5), 581-584.

Woo-Tan, L. B. (1989). *Effects of computer-assisted instruction on the learning of transformation geometry.* Unpublished Master's Thesis. National University of Singapore.

Yip, S. K., & Sim, W. K. (1990). *Evolution of educational excellence: 25 years of education in the Republic of Singapore.* Singapore: Longman Singapore.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 1, 41-45

Math and Reading Tests: Dissimilar Scores Provided by Similar Tests for African-American Students

Kevin Wickes
Taylor University

John R. Slate
Valdosta State University

Academic achievement test scores of a sample of 115 African-American students referred for a psychological evaluation or reevaluation were analyzed for similarities and dissimilarities. Despite tests that purported to measure the same math or reading constructs, statistically significant mean differences (ranging from 1.9 to 5.4 points) were present in four of six comparisons, with the Wechsler Individual Achievement Test providing higher mean scores than its KeyMath-Revised and Woodcock Reading Mastery Tests-Revised counterparts. Correlations of WIAT, KM-R, and WRMT-R scores were only moderate in nature, with math tests (average of 55%) sharing more common variance than the reading measures (average of 27%). Reading comprehension measures shared the lowest amount of common variance (18%). Findings were congruent with previous studies in which Caucasian students' test scores were analyzed. Implications of findings are discussed.

An integral part of the assessment process for special education involves the administration of achievement tests to ascertain students' skills in reading, mathematics, and writing (Individuals with Disabilities Education Act, the re-authorization of Public Law 94-142, Education for All Handicapped Children Act; United States Department of Health, Education, and Welfare [USDHEW], 1977). In an evaluation, the examiner subtracts students' scores on the achievement tests from their Full Scale IQ to determine the presence of a significant discrepancy, if a learning disability is suspected. The assessment specialist compares that score with students' IQs for congruence, if mental retardation or no disability is suspected. Regardless of the referral reason, an in-depth psychological evaluation would include at least one, and most likely two, achievement measures.

Commonly given achievement measures include the *Peabody Individual Achievement Test-Revised* (PIAT-R; Markwardt, 1989) to assess reading, math, and writing; the *Woodcock Reading Mastery Tests-Revised* (WRMT-R; Woodcock, 1987) to assess reading; the *KeyMath-Revised* (KM-R; Connolly, 1988) to assess math; or, the *Wechsler Individual Achievement Test* (WIAT; Wechsler, 1992) to assess the seven academic areas cited in special education rules and regulations. [Note. The PIAT-R is considered to be a screening test rather than an in-depth measure of academic achievement.] Not only are these tests supported in the research literature as being psychometrically sound instruments (Salvia & Ysseldyke, 1995), but, most importantly, special educators frequently use these tests, or their unrevised versions. They are the professionals most likely to administer the achievement component of the evaluation (Connelly, 1985). The WIAT (1992), a recently published achievement test, offers assessment specialists the benefit of being co-normed with the major intellectual measure, the *Wechsler Intelligence Scale for Children-Third Edition* (WISC-III, Wechsler, 1991). This co-norming reduces differences in test scores that might occur between tests solely as a function of different individuals in normative samples (Bracken, 1988).

In several recent studies (Slate, 1996; Slate & Jones, 1996; Slate & Saarnio, 1996) in which relationships of the WIAT with related achievement measures were examined, mean differences between tests with similar names that purported to measure the same achievement construct were revealed. In a study of 202 students with learning disabilities, Slate (1996) reported that the WIAT reading subscales provided statistically significant higher scores than PIAT-R and WRMT-R counterparts and that, in math, the KM-R provided higher scores than provided by the PIAT-R. Mean differences ranged from 6.0 points on basic reading skills to 5.2 points on reading comprehension and from 7.7 points on math applications to 5.6 points on overall math skills. Along with these mean differences, subtests purporting to measure the same constructs were only moderately related. Shared variance

Kevin L. Wickes is Assistant Professor in Psychology at Taylor University, Upland, Indiana. John R. Slate is Professor in the Department of Educational Leadership at Valdosta (Georgia) State University. Please address correspondence related to this article to John R. Slate, VSU, Dept. of Educational Leadership, Valdosta, GA 31698-0090 or by e-mail to jslate@valdosta.edu.

ranged from 23% to 83% for subtests purporting to measure the same constructs.

In a related study of 233 students with mental retardation, Slate and Saarnio (1996) replicated the findings of Slate (1996). Mean differences between tests with similar names were present and were larger than those differences found in the study of students with learning disabilities. For reading subtests of the PIAT-R, WRMT-R, and WIAT, mean differences ranged from 5.7 to 18.9 points whereas for math subtests of the PIAT-R, KM-R, and WIAT, mean differences ranged from 7.3 to 12.0 points. Again, the WIAT provided statistically significant higher mean scores than its counterparts. Tests purporting to measure the same reading constructs were again moderately related with reading comprehension measures having low common variance (26% to 31%). Math reasoning and math operations subtests shared between 37% and 46% of the variance.

Finally, Slate and Jones (1996) reported the presence of mean differences between the WIAT, WRMT-R, KM-R, and PIAT-R for 366 students who experienced academic difficulties. Again, mean differences on math and reading tests that purported to measure the same academic constructs were present, only moderate relationships among these measures were revealed, and reading comprehension measures had low common variance. Findings of this study were congruent with the findings of Slate (1996) for students with learning disabilities and Slate and Saarnio (1996) for students with mental retardation.

In these recent studies, as well as studies with predecessors of these instruments or studies with only one or two of the above instruments (e.g., Caskey, 1985; Caskey, Hylton, Robinson, Taylor, & Washburn, 1983; Eaves, Darch, & Haynes, 1989; Shapiro & Derr, 1987), researchers have provided extensive information that achievement tests purporting to measure the same academic constructs do not, in fact, provide similar results. Bracken (1988) provided cogent arguments why differences between tests with same/similar names should be the norm and not the exception. The strongest of his arguments involved differences, often substantial, in the quality and composition of the normative samples of the different achievement tests. Moreover, differences exist in test scores as a function of differences in publication dates among tests, with more recent test publications tending to provide lower test scores than tests with older publication dates. The fact is that different tests of the same construct will not provide identical scores. Readers are referred to Bracken (1988) for a more in-depth examination of his arguments.

Researchers, to date, have examined differences in academic achievement test scores of Caucasian students. The extent to which mean differences in achievement test scores purporting to measure the same construct are present for culturally different students such as African-American students is unclear. Because of the importance of academic achievement assessment in special education, information about the relationships of different achievement tests and whether tests provide similar test scores for students of culturally diverse populations is needed.

In this study, two research questions were addressed: (a) Are significant differences present among math and reading tests purporting to measure the same constructs?; and (b) What are the interrelationships among academic achievement tests purporting to measure the same constructs? These questions were addressed for a sample of African-American students with academic difficulties.

Methods and Procedures

Sample

Data were collected on a convenience sample of 115 African-American students (70 males, 45 females) enrolled in school districts served by three educational cooperatives in the Mississippi Delta region of northeastern Arkansas. Every African-American student on whom an evaluation (i.e., WISC-III and academic achievement measures) had been conducted in the 1995/96 school year in the school districts served by these three educational cooperatives was included in this study. The mean age of the students was 10 years 11 months (*SD* = 3 years 2 months). Ages ranged from 6 to 16. The mean WISC-III Full Scale IQ was 69.6 (*SD* = 9.8), the mean Verbal IQ was 69.7 (*SD* = 9.6), and the mean Performance IQ was 74.1 (*SD* = 11.9).

Depending on the nature of the referral question, academic achievement assessments were individualized. As a result, scores on all achievement tests and subtests were not available for every student. Thus, sample sizes for statistical analyses ranged from 45 to 74.

Results

Table 1 depicts means, standard deviations, and sample sizes on the WIAT, KM-R, and the WRMT-R. Mean scores were consistently in the 70s, with only three scores in the 80s, and these scores were obtained on the WIAT. Thus, the majority of the achievement scores were higher than the students' Full Scale IQs.

T-tests were conducted on six pairs of achievement tests with same/similar names. Because of the six comparisons, an alpha level of .008 was used as the level of statistical significance (i.e., .05 divided by the number of tests conducted). Results of four of the six *t*-tests met this .008 level of statistical significance. Scores on the WIAT Numerical Operations and the KM-R Operations tests, 5.4 points apart, were significantly different, $t(73) =$

-4.91, as were total scores, 3.5 points apart, on the KM-R and the WIAT, $t(53) = -3.82$. Scores on the WIAT Basic Reading and the WRMT-R Basic Reading, 4.4 points apart, were significantly different, $t(74) = 3.0$, as were scores, 4.3 points apart, on the WIAT Reading Comprehension and the WRMT-R Reading Comprehension tests, $t(60) = 2.74$. A comparison between the WIAT Math Reasoning and the KM-R Applications tests approached but did not reach a conventional level of statistical significance, $t(73) = -1.80$, $p < .07$. The remaining comparison, between the total reading scores on the WIAT and the WRMT-R, did not yield a difference. In all comparisons except for the total reading scores, the WIAT provided higher mean scores than the KM-R and the WRMT-R. [Note. The means and standard deviations of the tests reported in Tables 2 and 3 differ somewhat from the means and standard deviations of the tests reported in Table 1. The reason for these differences is found in the variability of the sample sizes in the *t*-test comparison. That is, only those individuals who had both sets of test scores in the *t*-test comparisons were included in the calculation of the means and standard deviations in Tables 2 and 3.]

Table 1
Means and Standard Deviations on the WIAT, KM-R, and WRMT-R.

Test Score	Mean	SD	*n*
WIAT			
Basic Reading	77.5	8.8	76
Reading Comprehension	73.7	10.4	62
Total Reading	70.4	11.6	45
Math Reasoning	80.1	10.0	76
Numerical Operations	83.6	12.2	76
Total Math	78.3	12.0	54
Spelling	77.1	9.1	76
Written Expression	71.9	8.7	43
Listening Comprehension	74.6	9.4	69
Oral Expression	81.8	9.1	69
KM-R			
Basic Concepts	76.3	11.5	94
Applications	78.7	11.4	94
Operations	78.4	12.6	94
Total Math	75.9	11.4	92
WRMT-R			
Basic Reading	73.0	15.6	94
Reading Comprehension	69.2	12.8	95
Total Reading	71.7	13.8	57

Table 2
Means, Standard Deviations, *t*-test Results, and Effect Sizes for Math Test Comparisons

	M	*SD*	Obtained Difference	*t*	*df*	*p*	effect size
KM-R Applications	78.3	11.1					
WIAT Math Reasoning	80.2	10.1	1.9	-1.80	73	>.05	.17
KM-R Operations	78.2	13.2					
WIAT Numerical Operations	83.6	12.4	5.4	-4.91	73	.0001	.42
KM-R Total Math	74.8	12.3					
WIAT Total Math	78.3	12.0	3.5	-3.82	53	.0001	.29

Table 3
Means, Standard Deviations, *t*-test Results, and Effect Sizes for Reading Test Comparisons

	M	*SD*	Obtained Difference	*t*	*df*	*p*	effect size
WIAT Basic Reading	77.6	8.9					
WRMT-R Basic Reading	73.3	15.8	4.4	3.05	74	.003	.37
WIAT Reading Comprehension	73.9	10.4					
WRMT-R Reading Comprehension	69.6	12.2	4.3	2.74	60	.008	.38
WIAT Total Reading	70.4	11.6					
WRMT-R Total Reading	72.3	14.2	1.9	-0.97	44	>.05	.14

Pearson *r*s, calculated on the math and reading comparisons shown in Tables 4 and 5, were positive and statistically significant at $p < .01$. Squared correlations, a measure of shared/common variance, provided evidence that 41% and 53% of the variance was shared for the math reasoning and operations subtests of the KM-R and WIAT and that 38% and 18% of the variance was shared for basic reading skills and the reading comprehension subscales of the WIAT and the WRMT-R. Similar to Slate (1996), Slate and Saarnio (1996), and Slate and Jones (1996), the correlations and shared variances among the math subscales were higher than the correlations among the reading subscales.

Table 4
Correlations of the WIAT and KM-R Math Subtests Purporting to Measure the Same Construct

Construct Test Scores	*r*	r^2	*n*
Math Reasoning			
WIAT with KM-R	.64	.41	74
Numerical Operations			
WIAT with KM-R	.73	.53	74
Total Math			
WIAT with KM-R	.85	.72	54

Note. All correlations were significant at the .01 level.

Table 5
Correlations of the WIAT and WRMT-R Reading Subtests Purporting to Measure the Same Construct

Construct Test Scores	*r*	r^2	*n*
Basic Reading Skills			
WRMT-R with WIAT	.62	.38	75
Reading Comprehension			
WRMT-R with WIAT	.42	.18	60
Total Reading			
WRMT-R with WIAT	.50	.25	45

Note. All correlations were significant at the .01 level.

Discussion

Similar to recent (Slate, 1996; Slate & Jones, 1996; Slate & Saarnio, 1996) and past studies (Caskey, 1985; Caskey et al., 1983; Eaves et al., 1989; Shapiro & Derr, 1987) of Caucasian students, statistically significant mean differences were present on achievement tests purporting to measure the same academic construct for a sample of African-American students. In this study, mean math differences ranged from 1.9 to 5.4 points, compared with differences of 5.6 to 7.7 points (Slate, 1996), 1.3 to 6.0 points (Slate & Saarnio, 1996), and 0.8 to 5.8 points (Slate & Jones, 1996). Mean reading differences in this study ranged from 1.9 to 4.4 points, compared to reading differences of 5.2 to 6.0 points (Slate, 1996), 4.6 to 14.0 points (Slate & Saarnio, 1996), and 0.4 to 5.5 points (Slate & Jones, 1996). Similar to these studies, higher mean scores were provided by the WIAT. Consistent with previous research, correlations among the achievement measures were moderate, with reading comprehension measures exhibiting the lowest amount of common variance (an average of 27%). Thus, substantial evidence exists to support Bracken's (1988) argument that similar tests do not provide similar results.

Findings in this study of a sample of African-American students are congruent with findings of previous studies of Caucasian students. Mean differences among math and reading subtests purporting to measure the same constructs were of the same magnitude as in previous studies. In addition, the shared variances among math and reading subtests were comparable with the shared variances reported in previous studies. Thus, the discrepancies among achievement tests purporting to measure the same academic constructs appear to be consistent for both Caucasian and African-American students.

Readers should, of course, be cautious in any generalizations they make from this study. First, the sample of African-American students was small. Second, the sample was from a geographically restricted region of the country. Despite these limitations, findings in this study were consistent with, rather than discrepant from, findings reported of larger samples of Caucasian students.

In summary, it is clear that results from standardized tests of achievement, at least those tests used in the diagnostic process, are not interchangeable. Achievement test scores of the same construct differ sufficiently that misdiagnosis can occur. Depending upon the specific achievement test administered, a child could demonstrate or not demonstrate a statistically significant discrepancy between his/her IQ and academic achievement. This discrepancy or non-discrepancy could occur as a function of the specific achievement test used. For example, if the WIAT was used as the achievement measure, then the resulting achievement test scores would be likely to be higher than if the KM-R or the WRMT-R were used. Obviously, standardized achievement tests that provide different scores for the same construct are problematic for practitioners. Practitioners are encouraged to check for alignment of the standardized achievement tests they use with the curriculum students are taught and to use the achievement test that has the highest degree of alignment with the curricula experienced by their students. Researchers are encouraged to investigate this issue in more detail with similar students as in this study as well as with students of different ethnic origins. Finally, we believe that test publishers would be well advised to inform practitioners of the relationship between the tests they publish and their competitors' tests as well as any differences present in test scores. This information would be of value to practitioners in test selection.

References

Bracken, B. (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of School Psychology, 26*, 155-166.

Caskey, W. E., Jr. (1985). The use of the Peabody Individual Achievement Test and the Woodcock Reading Mastery Tests in the diagnosis of a learning disability in reading. *Diagnostique, 11*, 14-20.

Caskey, W. E., Jr., Hylton, S., Robinson, J. B., Taylor, R. L., & Washburn, F. F. (1983). Woodcock and PIAT reading scores: A lack of equivalency. *Diagnostique, 9*, 49-54.

Connelly, J. B. (1985). Published tests-which ones do special education teachers perceive as useful? *Journal of Special Education, 19*, 149-155.

Connolly, A. (1988). *Key-Math Revised: A Diagnostic Inventory of Essential Mathematics.* Circle Pines, MN: American Guidance Service.

Eaves, R. C., Darch, C., & Haynes, M. (1989). The concurrent validity of the PIAT and WRMT among students with mild learning problems. *Psychology in the Schools, 26*, 261-266.

Markwardt, F. (1989). *Peabody Individual Achievement Test-Revised.* Circle Pines, MN: American Guidance Service.

Salvia, J., & Ysseldyke, J. E. (1995). *Assessment 6th ed.* Boston: Houghton Mifflin Company.

Shapiro, E. S., & Derr, T. F. (1987). An examination of overlap between reading curricula and standardized achievement tests. *The Journal of Special Education, 21*, 59-67.

Slate, J. R. (1996). Interrelationships of frequently administered achievement measures in the determination of Specific Learning Disabilities. *Learning Disabilities Research and Practice, 11*(1), 86-89.

Slate, J. R. (1995). Relationship of the WISC-III to the WRAT-R and the PPVT-R for students with academic difficulties. *Assessment in Rehabilitation and Exceptionality, 2*(4), 251-256.

Slate, J. R. (1994). WISC-III correlations with the WIAT. *Psychology in the Schools, 31*, 278-285.

Slate, J. R., & Jones, C. H. (1996). Differences in reading and math achievement test scores for students experiencing academic difficulty. *RESEARCH IN THE SCHOOLS, 3* (2), 15-21.

Slate, J. R., & Saarnio, D. A. (1996). Differences in reading and math achievement test scores for students with Mental Retardation. *B. C. Journal of Special Education, 20*, 34-46.

United States Department of Health, Education, and Welfare (USDHEW). (1977). Education of handicapped children: Implementation of Part B of the Education of the Handicapped Act. *Federal Register, 42*, 42474-42518.

Wechsler Individual Achievement Test (WIAT). (1992). San Antonio, TX: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children-III.* San Antonio: The Psychological Corporation.

Woodcock, R. W. (1987). *Woodcock Reading Mastery Tests - Revised.* Circle Pines, MN: American Guidance Service.

Footnote

[1]These data were collected as part of a larger project designed to explore multiple aspects of testing in special education diagnosis. Additional data regarding students' IQs, and the relationship of their IQs to achievement test scores, can be found in Slate (1994), Slate (1995), and Slate (1996).

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 1, 47-52

Screening for Behavior Problems in Preschool: A Comparison of the Social Skills Rating System (SSRS) and a Structured Teacher Interview

Ronald K. Bramlett and Kim B. Dielmann
University of Central Arkansas

Maleah Smithson
Cabot Public Schools, Cabot, Arkansas

The present study examined the classification agreement of risk status across alternative behavior rating methods with 80 preschool children and their teachers. In addition, the focus of intervention or reliability of target behavior selection was examined. The Social Skills Ratings System (SSRS) (Teacher) was compared to a structured teacher interview. Computer generated target behaviors were compared to teacher generated ones with moderate agreement found. Overall, the classification agreement in each comparison was low. Implications for school mental health professionals related to the identification of children with behavior problems as well as foci for intervention are discussed.

In recent years, considerable attention has been given to the assessment of preschool children for early intervention efforts (e.g., Bagnato & Neisworth, 1991; Barnett & Carey, 1992). However, the psychometric characteristics (e.g., item gradients, floor, ceiling, etc.) of formal assessment tools has proven to be limited when used with preschool children (Bagnato & Neisworth, 1994; Bracken, 1987; Bracken, Keith, & Walker, 1994; Flanagan & Alfonso, 1995). While intelligence tests have received the bulk of the criticism, tests that measure other constructs such as developmental delays, language ability, and behavior rating scales, are all subject to these criticisms (Bracken, 1987; Bracken & Rhymer, 1991). These psychometric issues are critical when one considers the possible outcomes of assessment findings. For example, scales measuring similar constructs may produce very different decisions (Bracken, 1988).

Several researchers have demonstrated that when children are assessed with different instruments even though they purportedly measure the same constructs (e.g., cognitive or language ability), diverse decisions are made such as which children will receive early intervention versus those who will not (Barnett & Macmann, 1992; Bracken, 1988; Hall & Barnett, 1991; Hall, Bramlett, Barnett, & Cox, 1994). Studies examining classification and screening decisions have indicated that very different decisions can be made depending on which instrument is selected (Scale A vs. Scale B). In a study by Hall and Barnett (1991) two tests designed to measure educational risk in preschoolers, the DIAL-R and the SCREEN, found little agreement in identifying at-risk preschool students. Even scales with high reliability and validity may lead to very different outcomes (Barnett & Macmann, 1992).

As a result of these criticisms, many early childhood educators and psychologists have recommended ecologically-based assessment procedures (e.g., Bagnato & Neisworth, 1991; Barnett & Carey, 1992; Bracken, Bagnato, & Barnett, 1991; Neisworth, 1990). Assessment procedures should demonstrate treatment and social validity. Treatment validity refers to the extent to which decisions about intervention needs, services, goals, and outcomes are directly guided by assessment. Social validity refers to the acceptability of the methods utilized, the importance of the data obtained, and ecological characteristics of the data (Neisworth, 1990). There is a lack of research demonstrating that ecologically-based procedures provide reliable and valid decisions (Macmann, Barnett, Allen, Bramlett, Hall, & Erhardt, 1996).

As a part of the overall consultation process, problem-solving interviews with teachers, parents, and others in the child's environment hold promise as an effective tool for the behavioral assessment of preschool children, particularly with social and behavioral concerns (Bagnato & Neisworth, 1994; Barnett & Carey, 1992; Gresham, 1992). Much of the research with interviews as assessment procedures has been conducted in an informal manner despite its conceptual and practical importance in target behavior identification (Kratochwill, 1985). A structured problem-solving interview can assist in

Ronald K. Bramlett is Associate Professor and Kim Dielmann is an instructor in the Department of Psychology and Counseling at the University of Central Arkansas. Maleah Smithson is with the Cabot Public Schools in Cabot, Arkansas. This research was supported by a grant from the University of Central Arkansas. Please address correspondence to Ronald K. Bramlett, Department of Psychology and Counseling, Box 4915, University of Central Arkansas, Conway, AR 72035 (Ronkb@mail.uca.edu).

defining problem behaviors, determining when and in what settings these behaviors occur, antecedents and consequences of these behaviors, and information needed to develop effective interventions. Interviews can also provide information about how others react to the child which can be useful in establishing a more complete picture of the child's functioning. However, the technical characteristics of most interviews are unknown (Macmann et al., 1996).

The problem-solving interview is one of the primary techniques in conducting a functional analysis of the child's behavior (e.g, Gresham, 1991). This is of particular importance given IDEA's current emphasis on functional behavioral assessments. The problem-solving interview provides information that will assist in defining the specific problems of concern and the function of those behaviors (Daly, Witt, Marten, & Dool, 1997). In two studies comparing the problem-solving interview to standardized instruments, the intervention plans already developed by the problem-solving interview did not change as a result of information added by the standardized instruments (Carey, 1989; Vedder-Dubocq, 1990). Therefore, the problem-solving interview can provide a method of understanding a child's behavior, how that behavior affects others, and suggest ways to intervene. More research is needed on the technical adequacy of problem-solving interviews and their relationship to other assessment methods (Macmann et al., 1996).

Purpose of this Study

Social functioning is an important area for directing early intervention efforts. Deficits in social skills and peer related problems have been predictive of later adjustment problems and/or psychopathology (e.g., Cowen, Pederson, Babigian, Izzo, & Trost, 1973; Landau & Milich, 1990). One current standardized method to assess social skills and problem behaviors in preschool children is the Social Skills Rating System (SSRS) (Gresham & Elliott, 1990). This instrument is among the most technically sound social skills assessment instruments available (Bracken, Keith, & Walker, 1994). In this study, the SSRS (Teacher Form) and a structured, problem-solving teacher interview were compared to determine their agreement on the classification of preschool children in need of some type of social/behavioral intervention. A second purpose of the study was to compare the agreement between these two procedures on the focus of the intervention for these children (e.g., the comparison of computer generated behavioral objectives of the SSRS-Teacher Form to a structured teacher interview).

Methods

Participants and Setting

Five teachers of eighty preschool children from three Head Start centers in a rural setting participated in the study. The children served by the Head Start centers came from lower income families. The children who were rated consisted of 38 males and 42 females with a mean age of 4 years, 6 months (range 3-10 to 4-11, S.D. = 5 months). The sample was racially characterized as 80% Caucasian and 20% African American. The average class size was 16 with two of the Head Start centers having two classrooms each.

Instruments

Social Skills Rating System (SSRS). The Social Skills Rating System (SSRS) is a standardized, norm-referenced instrument designed to provide professionals with a means to screen and/or classify student social and problematic behaviors in educational and family settings. The SSRS offers three methods of evaluating students' social behaviors: (a) the Parent Form, (b) the Teacher Form, and (c) the Student Form. The Parent and Teacher versions are divided into three developmental levels: preschool, kindergarten through sixth grade, and seventh through twelfth grades. The Student Form is available for students who are able to read at approximately the third grade level. The number of items on the Parent Form ranges from 49 at the Preschool Level to 55 at the Elementary Level. The Parent Form yields four subscale scores: Cooperation, Assertion, Responsibility, and Self-Control. In addition a total social skills score is obtained (mean = 100; S.D. = 15). Finally, two Problem Behavior Subscale scores (Internalizing and Externalizing) are derived from a total of 10 items. Social skills ratings are based on the frequency of the social behavior (never, sometimes, or very often) and on the importance of the behavior (not important, important, or critical). Problem behavior ratings are based only on the frequency of the behavior.

The SSRS standardization sample included 4,170 student self-ratings, 1,027 parents, and 259 teachers. Included in this sample were 212 teacher and 193 parent ratings of preschool children. The sample was drawn from 18 states in the Northeast, North Central, South, and Western regions of the United States. Minority students made up approximately 27% of the normative sample.

Internal reliability is reported to be in the .83 to .94 range for the Social Skills scale and from .73 to .88 for the Problem Behavior scale. Test-retest reliability at the 4-week interval is reported to be .85 for Social Skills and .84 for Problem Behaviors. Inter-rater reliability of the SSRS is low. Coefficients for the Elementary Level ranged from .03 to .41. The median correlation between teacher and parent ratings was .30. The authors note that

one would not expect a high level of agreement between raters because they typically possess different views of the student. Validity evidence is supported by correlational studies comparing the SSRS to other established social behavior instruments (Gresham & Elliott, 1990).

ASSIST, a computer scoring system for the SSRS, was developed to score and generate behavioral objectives based on the responses of the teachers. This program was used in this study to generate behavioral objectives for comparison purposes.

The Problem-Solving Interview. The Problem-Solving Interview was adapted from Peterson's (1968) scanning operation and Alessi's (1977) problem-solving interview. Similar interviews are found frequently in consultation research and practice (e.g., Bergan & Kratochwill, 1990). The interview allows teachers to define problem behaviors exhibited by a child and the child's strengths. The interview uses a sequence of questions to determine (a) what behaviors are problematic, (b) in what settings they occur, and (c) under what conditions. As the interviewer guides the teacher through the interview, plans for interventions are identified to facilitate problem-solving. By the end of the interview, the teacher provides a systematic list of problem behaviors in terms of frequency and importance for intervention purposes. No psychometric characteristics of the interview are available.

Procedures

All students were rated by their classroom teacher with the SSRS Teacher Form in the social skills and problem behaviors domains. Children scoring one standard deviation below the mean (standard score of 85 and below) on Social Skills and one standard deviation above the mean (standard score of 115 and above) on Problem Behaviors were considered to be at-risk and in need of intervention. This was the criterion recommended by Gresham and Elliott (1990).

During the structured interview, teachers were asked to identify students from their classrooms with social/behavioral problems and/or about whom they were concerned. Teachers were interviewed by school psychology graduate students trained in using the structured problem-solving interview. Teachers were given the SSRS and the interview in a counterbalanced order. That is, two of the teachers completed the SSRS prior to being interviewed while three teachers were interviewed prior to completing the SSRS. The interviews and SSRS ratings occurred over a two-week period.

Data Analysis

Classification of risk status was evaluated three ways: overall proportion of agreement, specific agreement, and Kappa. Proportion of agreement (PO) measures agreement on the presence and absence of risk status but is markedly affected by rates of agreement. Specific agreement (S) represents agreement on the presence of risk status. *S* may be interpreted to mean that a participant identified as at-risk on one procedure will also be identified as at-risk on the other procedure. Kappa (Cohen, 1960) was computed to represent the proportion of agreement between children classified on the interview and the SSRS in social skills and problem behavior domains on both the presence and absence of risk status, corrected for chance agreement. It generally is accepted for establishing nominal scale agreement (McDermott, 1988).

In order to determine the agreement between the two instruments on targets for intervention, inter-rater agreement was calculated. For the children selected by the teacher as "at-risk," the behaviors identified on the interview were compared to the behaviors generated by the SSRS. Three school psychology graduate students and a school psychology faculty member served as judges in rating the behaviors generated by both instruments. We used the interview as the criterion for judging agreement because generally teachers are the ones most responsible for defining concerns in most evaluations of students. Thus, if the behavior identified by the interview was also found on the SSRS, an agreement was counted. If the behavior was found on the interview and not on the SSRS, a disagreement was counted. All agreements and disagreements were added and divided to obtain an agreement coefficient (agreements divided by agreements + disagreements). Occasionally, semantic differences made comparisons difficult. For example, the SSRS would use "the student will follow the teacher's directions," while on the interview, the teacher might want to target the student "should do as is asked," etc. On these occasions, if the behavior was determined to be very similar by the three judges, an agreement was counted. On items where disagreements occurred between the judges, consensus was reached prior to categorizing the item as an agreement or disagreement.

Results

The SSRS Teacher Form identified twelve children as having deficits in the social skills domain and thirteen children as having problem behaviors. On the Problem-Solving Interview, teachers also selected twelve children as having social or behavioral concerns. Tables 1 and 2 illustrate the cross-classification of subjects based on results obtained from the SSRS and the Interview. The two tables include: (a) children identified as at-risk by both the interview and the SSRS social skills domain and

(b) children identified as at-risk by both the interview and the SSRS problem behavior domain.

Table 1
Cross-classification of Risk Status on the Interview and the SSRS Social Skills Domain (n = 80)

		SSRS Social Skills	
		At-risk	Not at-risk
Interview	At-risk	3 (25%)	9 (75%)
	Not at-risk	8 (12%)	60 (88%)

Table 2
Cross-classification of Risk Status on the Interview and the SSRS Problem Behavior Domain (n = 80)

		SSRS Problem Behaviors	
		At-risk	Not at-risk
Interview	At-risk	2 (17%)	10 (83%)
	Not at-risk	10 (15%)	58 (85%)

There was relatively little agreement between the SSRS and the Interview on the classification of which children were actually at-risk. Although the overall proportion of agreement on the presence or absence of risk was moderate (PO = .64, Social Skills; PO = .61, Problem Behaviors), the PO statistic is likely to be inflated because of the agreement on the absence of risk (Fleiss, 1981). The proportion of specific agreement was low (S = .26, Social Skills; S = .17, Problem Behaviors) but presents a more accurate view of agreement because it is not affected by rates of agreement on the "not at-risk" category. Kappa, which represents agreement beyond chance expectations, was also low and suggests that the relationship between the interview and the SSRS on the classification of risk status did not exceed chance expectations (kappa = .17, Social Skills; kappa = .20, Problem Behaviors). For example, when the Interview was compared to the Social Skills domain, a total of 20 children were identified as at-risk by one of the procedures. Yet, the procedures agreed on only 3 of those children. They were much more likely to agree on those who were not at-risk. A similar finding was made when the Interview was compared to the SSRS Problem Behavior domain. The Interview and the SSRS agreed on 2 out of the total 22 that were identified as at-risk on both procedures.

Table 3 lists the agreement between the Interview and the SSRS on the selection of target behaviors for intervention. The SSRS averaged 12 behavioral objectives per identified child compared to only three for the interview. It should be noted, however, that the interview was designed to prioritize target behaviors, thus, a smaller number of behaviors was the goal. If the behavior identified by the interview was also found on the SSRS, an agreement was counted. If the behavior was found on the interview and not on the SSRS, a disagreement was counted. Agreement on specific target behaviors was low to moderate. The mean inter-rater agreement scores for target behavior selection for the interview were .51 for social skills (range = .00 to 1.00) and .32 for problem behaviors (range =.00 to .67). For example, if all three target behaviors identified on the Interview were also found on the SSRS, then 100% agreement was noted even though the SSRS had 9 additional behavioral objectives. If 2 of the 3 target behaviors on the Interview were found on the SSRS, then we considered that 67% agreement.

Table 3
Agreement between the Interview and the SSRS on Target Behavior Selection

	Agreement with SSRS	
Subject	Social Skills	Problem Behaviors
1	.67	.33
2	.50	.00
3	1.00	.00
4	.25	.33
5	.50	.00
6	.50	.50
7	.67	.67
8	1.00	.00
9	.00	.67
10	.50	.00
11	.00	.33
12	.50	1.00
Mean	.51	.32

Discussion and Implications for School Mental Health Practitioners

These results are consistent with other studies comparing the agreement (or lack of agreement) of decisions when utilizing different assessment procedures (e.g., Hall & Barnett, 1991; Hall, Bramlett, Barnett, & Cox, 1994). The agreement between the instruments on targets for intervention was low to moderate which shows some evidence that the methods converge on some important key areas. Perhaps, these areas of agreement might assist practitioners in prioritizing intervention goals. These goals are probably best determined on an individual child basis.

The implications for school mental health practitioners are important. It is clear from this study and

similar previous studies (Hall & Barnett, 1991; Hall et al., 1994), that practitioners should be aware of potential divergence about intervention decisions depending on the type of instrument selected as well as the informant selected (e.g., parent, teacher). Depending on the assessment procedure selected, different children may be identified for early intervention. False positives and false negatives translate into potential financial and human costs. Also, after children are identified, the focus of intervention can be greatly affected by the instrument(s) selected.

Bracken et al. (1994) point out that multiple-raters, multiple-instrument, and multiple-context assessment should improve the reliability of professional judgment about young children's functioning. By utilizing multiple procedures, school mental health professionals may not be lulled into relying exclusively on a "score" to base their decisions. But rather, they will consider a full-range of potential alternatives for intervention efforts. Bracken et al.'s (1994) admonition is particularly important in pulling together what may appear to be conflicting information (e.g., which kids really need help and what help do they actually need).

Another critical issue for schools is the time and cost of screening children. In this study, teachers completed the SSRS on all of their students which took about 10-15 minutes per child (13-20 hours), while the interview took about 30 minutes per identified child (6 hours each for teacher and psychologist). Commercially available instruments, like the SSRS, do cost the school in terms of purchase price, protocols, and computer software while the interview costs only in terms of the time of the participants (teacher, psychologist).

The problem-solving interview holds a prominent role in many professional assessment activities of school psychologists and special educators, particularly as it relates to the consultation process. This study adds to professional knowledge about some of the technical characteristics of the problem-solving interview. More research is needed to validate the interview in a similar fashion to other standardized assessment techniques because of its recommended usage in both the professional literature and federal special education guidelines (Macmann et al., 1996). Perhaps, the most important aspect of future research related to screening preschool children with social/behavioral concerns is the predictability of their behavior for long-term adjustment. Some research shows that young children's problems (acting out behaviors) often do not persist (Fischer, Rolf, Hasazi, & Cummings, 1984). This area of investigation is critical for early intervention efforts.

It appears that the reliability and validity of assessment outcomes and decisions might best be enhanced by an ongoing problem-solving process whereby assessment results and plans are evaluated over time. As long as classification decisions are based on cut scores or similar one-time-only guidelines, it is likely that procedures will continue to disagree on the kids who are identified and on targets for intervention. Until we know which procedures lead to the best long-term outcomes for children, our most cautious approach should be to utilize multiple approaches, informants, and contexts and not rely on any sole method of assessment.

References

Alessi, G. J. (1977). Problem-solving interview: Teacher or parent content agenda. In G. J. Alessi & J. H. Kaye (Eds.), *Behavioral assessment for school psychologists* (pp. 36-43). National Association of School Psychologists Professional Development Publication, *1*, (1).

Bagnato, S. J., & Neisworth, J. T. (1991). *Assessment for early intervention: Best practices for professionals.* New York: Guilford.

Bagnato, S. J., & Neisworth, J. T. (1994). A national study of the social and treatment "invalidity" of intelligence testing for early intervention. *School Psychology Quarterly, 9*, 81-102.

Barnett, D. W., & Carey, K. T. (1992). *Designing interventions for preschool learning and behavior problems.* San Francisco: Jossey-Bass.

Barnett, D. W., & Macmann, G. M. (1992). Decision reliability and validity: Contributions and limitations of alternative assessment strategies. *Journal of Special Education, 25*, 431-452.

Bergan, J. R., & Kratochwill, T. R. (1990). *Behavioral consultation and therapy.* New York: Plenum.

Bracken, B. A. (1987). Limitations of preschool assessment and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5*, 313-326.

Bracken, B. A. (1988). Ten psychometric reasons why similar tests produce dissimilar results. *Journal of School Psychology, 26*, 155-166.

Bracken, B. A., Bagnato, S. J., & Barnett, D. W. (1991). *Early childhood assessment* (NASP Position Statement): Washington, DC: National Association of School Psychologists.

Bracken, B. A., Keith, L. K., & Walker, K. C. (1994). Assessment of preschool behavior and social-emotional functioning: A review of thirteen third-party instruments. *Assessment in Rehabilitation and Exceptionality, 1*, 331-346.

Bracken, B. A., & Rhymer, P. (1991, November). *Technical adequacy of preschool language and*

articulation assessments. Paper presented at the annual conference of the American Speech, Hearing, and Language Association, Atlanta, Ga.

Carey, K. T. (1989). *The treatment utility potential of two methods of assessing stressful relationships in families: A study of practitioner utilization*. Unpublished doctoral dissertation, University of Cincinnati.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.

Cowen, E., Pederson, A., Babigian, H., Izzo, L., & Trost, M. (1973). Long-term follow-up of early detected vulnerable children. *Journal of Consulting and Clinical Psychology, 41*, 438-444.

Daly, E. J., Witt, J. C., Marten, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26*, 554-574.

Fischer, M., Rolf, J. E., Hasazi, J. E., & Cummings, L. (1984). Follow-up of a preschool epidemiological sample: Cross-age continuities and predictions of later adjustment with internalizing and externalizing dimensions of behavior. *Child Development, 55*, 137-150.

Flanagan, D. P., & Alfonso, V. C. (1995). A critical review of the technical characteristics of new and recently revised intelligence tests for preschool children. *Journal of Psychoeducational Assessment, 13*, 66-90.

Fleiss, J. L. (1981). Statistical methods for rates and proportions (2nd ed.). New York: John Wiley.

Gresham, F. M. (1991). What ever happened to functional analysis in behavioral consultation. *Journal of Educational and Psychological Consultation, 2*, 387-392.

Gresham, F. M. (1992). Social skills and learning disabilities: Causal, concomitant, or correlational? *School Psychology Review, 21*, 348-360.

Gresham, F. M., & Elliott, S. N. (1990). *Social Skills Rating System*. Circle Pines, MN: AGS.

Hall, J. D., & Barnett, D. W. (1991). Classification of risk status in preschool screening: A comparison of alternative measures. *Journal of Psychoeducational Assessment, 9*, 152-159.

Hall, J. D., Bramlett, R. K., Barnett, D. W., & Cox, F. (1994). Classification of risk status in kindergarten screening: A comparison of alternative measures. *Journal of Psychoeducational Assessment, 12*, 154-164.

Kratochwill, T. R. (Guest Editor). (1985). Miniseries on target behavior selection [Special Issue]. *Behavioral Assessment, 7* (1).

Landau, S., & Milich, R. (1990). Assessing children's social status and peer relations. In A. M. La Greca (Ed.), *Through the eyes of the child: Obtaining self-reports from children and adolescents* (pp. 259-291). Boston: Allyn and Bacon.

Macmann, G. M., Barnett, D. W., Allen, S. J., Bramlett, R. K., Hall, J. D., & Erhardt, K. E. (1996). Problem solving and intervention design: Guidelines for the evaluation of technical adequacy. *School Psychology Quarterly, 11*, 137-148.

McDermott, P. A. (1988). Agreement among diagnosticians or observers: Its importance and determination. *Professional School Psychology, 3*, 225-240.

Neisworth, J. T. (1990). Judgment-based assessment and social validity. (Special Issue). *Topics in Early Childhood Special Education, 10*.

Peterson, D. R. (1968). *The clinical study of social behavior*. New York: Appleton-Century-Crofts.

Vedder-Dubocq, S. A. (1990). *An investigation of the Parenting Stress Index for intervention decisions*. Unpublished doctoral dissertation, University of Cincinnati.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 1, 53-63

A Structural Model of Student Career Aspiration and Science Education

Jianjun Wang
California State University, Bakersfield

School-to-work transition is an issue of national significance in the 1990s. The gap between school and work places has been especially large for non-college bound students due to their limited training in sciences. While factors of educational productivity were extensively studied by many researchers, few investigations covered the school influence on student career orientation. A new feature of this study is to model the relationship between student career aspiration and educational productivity in science teaching. The structural equation model was cross-validated by the seventh grade data from the Longitudinal Study of American Youth (LSAY). The statistical findings were interpreted in terms of student aptitude-attributes, instructional characteristics, and contextual factors in the class, home, peer, and media environments.

Technological advancement in the 1990s demanded better-trained employees to operate smart machines and reconfigure them for effective applications. This market change imposed new challenges on science education, particularly for non-college bound students with limited academic preparations. Decker (1997) noted, "Increases in educational attainment were responsible for an estimated 11 to 20 percent of growth in worker productivity in the United States in recent decades" (p. 5). Thus, an important measure of educational productivity is the improvement of student achievement (e.g., Lavin, 1965; Walberg, 1981, 1982, 1983, 1984; Young, Reynolds, & Walberg, 1996). Under an assumption that factors of educational productivity may also affect student career orientation, the purpose of this study is to construct an empirical model articulating student career aspiration and factors of educational productivity. Since a career choice can be ranked on a scale of social economic index (Stevens & Cho, 1985; Stevens & Featherman, 1981), the model construction may not only facilitate assessment of student career orientation but also reveal critical factors of science education for maintaining the best job market in the United States.

Jianjun Wang is Associate Professor of Educational Statistics and Research, Department of Teacher Education, at California State University, Bakersfield. This research was supported by a grant from the American Educational Research Association which receives funds for its "AERA Grants Program" from the National Science Foundation and the National Center for Education Statistics (U.S. Department of Education) under NSF Grant #RED-9452861. Opinions reflect those of the author and do not necessarily reflect those of the granting agencies. Correspondence regarding this article should be sent to Jianjun Wang, Educational Statistics and Research, Department of Teacher Education, California State University, Bakersfield, CA 93311-1099 or by e-mail to jwang@academic.csubak.edu.

Literature Review

Science education played an important role in school-to-work transition. The National Research Council (1996) reported that "the economic productivity of our society is tightly linked to the scientific and technological skills of our work force" (p. ix). Non-college bound students are a component of the work force with the least training in school. Thus, their career orientation has been delimited by the scope of education, and an investigation of the empirical relations may help raise student career aspirations toward future employment.

Indicators of Career Aspiration

The anticipated job for adulthood is an indicator of student career aspiration (Miller, Hoffer, Suchner, Brown, & Nelson, 1992). Early employment during the school years may not necessarily signify smooth school-to-work transitions for non-college bound students (Medrich, 1996). Halperin, Melaville, and Taylor (1988) cautioned, "For most non-college-bound youth, the problem is not unemployment but extended underemployment in low-wage, dead-end jobs" (p. 25). Therefore, the future career aspiration is one of the critical factors behind the school-to-work transition. Raelin (1980) analyzed a national longitudinal data base and concurred that "by far the most important attitude uncovered by this study in terms of its contribution to later work experience is career aspiration" (p. 132).

Nonetheless, the importance of career aspiration has not caught the attention of the American public (Bishop, 1996). Consequently, Wilson (1996) noted that "The unemployment rates among both low-skilled men and women are five times that among their college-educated counterparts" (p. 28). At the non-college bound level, whether students have been well-prepared for education-based, promising jobs is an ultimate measure directly relying on the factors of educational productivity (Raelin,

1980; Wilson, 1996). On the basis of existing research literature, career aspiration has been included in this study to bridge the transition between school and workplace.

A Theory of Educational Productivity

Levels of educational productivity were typically measured by student academic achievement (Young, Reynolds, & Walberg, 1996). Fraser, Walberg, Welch, and Hattie (1987) noted that "the public ranks research on educational productivity higher in priority than most other fields of scientific investigation in the other natural and social sciences" (p.149). Among the existing research on educational productivity, Walberg (1981, 1986) proposed a nine-factor theory through synthesizing several thousand investigations. Reynolds and Walberg (1992a) elaborated:

The nine productivity factors are divided into three sets.

A. Aptitude includes --
   1. Student ability or prior achievement
   2. Motivation
   3. Age or developmental level

B. Instruction is indexed by --
   4. Quantity or amount of time
   5. Quality or appropriateness for the student

C. Psychological environment includes --
   6. Class climate
   7. Home environment
   8. Peer group
   9. Exposure to mass media outside of school including television. (p. 307)

Guided by the productivity theory, more than 120 studies have been conducted in the research community (Young, Reynolds, & Walberg, 1996). Ralph W. Tyler (1984), a predecessor of educational measurement, appraised Walberg's theory highly, and wrote:

> Herbert Walberg has done a superior interpretation of one of the most massive collections of data on school learning. He avoids the common weaknesses of many statistical reports; he recognizes the complexity of much human learning and does not try to reduce it to a simplistic model; he discusses the meaning of the data as well as indicating the quantitative results; he does not confuse statistical significance with substantive or social significance; he seeks to explain interactions among variables in commonsense terms; and he examines and reports both macro studies and micro studies. (p. 27)

In the 1990s, Walberg and his colleagues reconfirmed the model of educational productivity through analyses of a national data base from the *Longitudinal Study of American Youth* (LSAY), a project funded by the National Science Foundation (NSF) (Reynolds & Walberg, 1991, 1992b; Young, Reynolds, & Walberg, 1996). Nonetheless, few researchers extended Walberg's theory to other measures of educational productivity, such as preparing students for education-based, promising jobs. To fill in the void, a unique feature of this study is to examine student career aspirations along with various factors of educational productivity.

Data Selection

Most national data covering issues of future planning were collected from longitudinal studies. Thus far, the information of non-college bound students has been collected in four projects, the *National Longitudinal Study of the High School Class of 1972* (NLS-72), *High School and Beyond* (HS&B), the *National Education Longitudinal Study of 1988* (NELS:88) and the *Longitudinal Study of American Youth* (LSAY). The first three projects were conducted by the *National Center for Education Statistics* (NCES), while LSAY was supported by the *National Science Foundation* (NSF). Davis and Sonnenberg (1995) analyzed the connection among the NCES data bases, and noted that NELS:88 was built on the experience of NLS-72 and HS&B.

Similarly, for the NSF project, Suter (1992) observed that "efforts were made by the LSAY research team to include items from both NLS-72 and HS&B" (p. 131). Hoffer (1988) further delineated differences between NELS:88 and LSAY:

> The NELS88 cognitive tests, for example, included only about half the number of items in the LSAY. And the LSAY attitudinal batteries included at least two and usually three items for each dimension, while the NELS88 batteries have only one item for each dimension. (pp. 11-12)

Although students' future career choices were assessed in NELS:88, the LSAY project had a more informative item in this area. Specifically, the LSAY researchers further converted the career responses into the U.S. Bureau of Census codes with a *socioeconomic index* score assigned to each code (Miller, Hoffer, Suchner, Brown, & Nelson, 1992). The scale transformation facilitated the use of parametric statistical methods to model career aspirations. To date, career aspirations of the seventh graders have not been investigated by the NCES longitudinal projects, including NLS-72, HS&B, and NELS:88 (Davis & Sonnenberg, 1995). Therefore, the LSAY data represented a unique opportunity to examine career orientations at the initial stage of adolescence.

Some researchers already noted the potential of using the LSAY data to study student career aspiration. Miller and Brown (1992), for instance, reported that "Results of the LSAY data analyses permit an understanding of the relative importance of the factors constituting a model to predict the development of career expectations" (p. 221). On the other hand, Walberg and his colleagues have employed the LSAY data to reconfirm their theory of educational productivity (Reynolds & Walberg, 1991, 1992a, b; Young, Reynolds, & Walberg, 1996). Despite the existence of relevant information on both sides, no researcher has employed Walberg's theory to investigate relations between educational productivity and student career aspiration. Hence, the further investigation presented in this article may produce more empirical findings to enrich the existing knowledge base.

In 1987, approximately 3000 seventh grade students were selected randomly in the LSAY project and followed twice yearly for more than five years (Chicago Academy of Sciences, 1995). Meanwhile, parents, teachers, and principals also filled out questionnaires and received telephone interviews. The national probability sample was stratified by geographic region (east, west, north, and south) and community (urban, suburban, and rural). To maintain the data representation in the model reconfirmation, the LSAY sample was split into two parts by selecting every other case. Cases in the odd-numbered half were employed to develop a structural equation model, and the even-numbered half was adopted to cross-validate the findings between career aspiration and educational productivity.

Despite various changes since 1987, little variation was found on the basic condition of American mathematics and science education. Baker (1997) reviewed findings of the Third International Mathematics and Science Study (TIMSS), and noted that "American eighth-graders in 1996 placed about where they did in the early 1980s in the Second International Math and Science Studies (SIMS and SISS respectively)" (p. 295). Thus, the LSAY data collected in the late 1980s remain relevant to the existing condition of science education.

In summary, the LSAY data have been chosen by Walberg and other researchers to develop structural equation models on student science achievement (Reynolds & Walberg, 1991, 1992b; Young, Reynolds, & Walberg, 1996). However, few researchers extended these studies to issues of student career orientation. While science achievement is an indicator of school success, a final goal of school is to prepare students for school-to-work transition (Murnane & Levy, 1996). Under an assumption that factors of educational productivity may also affect issues of school-to-work transition, Walberg's theory has been adopted in this study to construct a structural equation model on career aspiration. The data selection was grounded on an extensive review of the existing public data bases. Hence, empirical relations identified in this study can be verified by other researchers through secondary analyses.

Research Questions

While knowledge acquisition is a major task of school education, Jenkins (1998) suggested that "knowledge development is a critical ingredient in the career aspiration formula" (p. 1). Accordingly, research questions that guide the analyses of LSAY data are:

1. What are the relations between factors of educational productivity and student career aspirations?
2. Is the career aspiration model strongly supported by the LSAY data base?
3. How can the empirical findings be interpreted in a broad education context?

Methods

To avoid underemployment among non-college bound students, it is important to prepare them for education-based, promising jobs in the future. In the LSAY project, students' choices of future occupation were scaled on a social economic index (Stevens & Cho, 1985; Stevens & Featherman, 1981), and released in the LSAY public data base (Miller et al., 1992). Meanwhile, the career utility of science classes was assessed by a Likert scale item in LSAY. The variable codes were reversed positively with 1 representing "no use" and 5 representing "very useful." These two variables, the *expected occupation* and the *career utility of science*, have been chosen in this study to indicate student career aspiration.

The original Walberg theory contained nine productivity factors (Walberg, 1981, 1986). In analyzing the LSAY data base, Reynolds and Walberg (1992b) reported that: "Because the students in the sample were all from the same grade level, age was relatively constant and therefore omitted" (p. 373). The observation was confirmed in this empirical investigation. Consequently, construction of the structural model was based on the remaining eight factors of educational productivity (Table 1).

Structural equation modeling is a comprehensive statistical approach to constructing relations among indicator variables and latent factors (Hoyle, 1995). The LISREL8 software was employed to estimate structural parameters among the standardized factors and indicators. Researchers demonstrated that even for models of the same structure, indicators of a latent factor can still be chosen differently, resulting in different estimates of structural relations (e.g., Reynolds & Walberg, 1991, 1992b). Since

the entire LSAY data contain more than 8,000 variables, individual researchers may be perplexed by the overwhelming amount of information. Fortunately, the LSAY researchers constructed a set of composite variables to enhance the information coverage on student aptitude-attributes, instructional characteristics, and psychological environment (Miller et al., 1992). The composite variables have been included in an LSAY CD-ROM disk (Chicago Academy of Sciences, 1995). Accordingly, indicators of the latent factors were selected from composite variables developed by the LSAY researchers (Miller et al., 1992). Each composite variable was based on multiple items in the LSAY survey instrument. For instance, the *family socioeconomic status* variable was composed by three sources of information, *parental education, parental occupational status*, and *a household possession index* (Miller et al., 1992). The use of multiple sources of information was recommended by many researchers to reduce potential measurement errors (e.g., Bentler, 1980; Hayduk, 1987; Reynolds & Walberg, 1991). To facilitate the model verification, correspondence between the composite indicators and the latent factors is presented in Table 1. The means and standard deviations of the indicators are included in Table 2. After pairwise deletions of the missing value, the minimum number of observations employed for the correlation calculation was 2482, above 80% of the sample size.

Table 1
Description of Latent Factors and the Corresponding LSAY Indicators

Latent Factor	LSAY Indicator	Description
Career Inspiration	AA22AS	SEI of the expected occupation
	AASCIF	career utility of science class
Education outcome	ASCIIRT	student science achievement
	SSCAT1	student attitude toward science
Motivation	SIPPS1	student independence & persistence
	SSFES1	student self-esteem
Instructional quantity	AA27A	student report of homework hours
	SSHWE1	student school work ethic
Instructional quality	SSCTCPH1	science teacher academic push
Home environment	PSCPH1	parental science push
	PACPH1	parental academic push
	PCOPH1	parental college push
	SES3	family socioeconomic status
Class environment	SFESX1	student's fear of success in class
Peer environment	KSCPH1	peer science push
	KACPH1	peer academic push
Mass media	SNWAQ1	student news acquisition

Table 2
Means and Standard Deviations of the Indicator Variables

Indicators	Description	N	Mean	SD
AA22AS	SEI of the expected occupation	2923	58.68	23.10
AASCIF	career utility of science class	2750	3.78	1.42
ASCIIRT	student science achievement	3065	50.35	10.15
SSCAT1	student attitude toward science	3051	14.27	2.81
SIPPS1	student independence & persistence	3051	10.29	2.93
SSFES1	student self-esteem	3062	16.38	3.69
AA27A	student report of homework hours	3064	6.00	6.55
SSHWE1	student school work ethic	3039	12.32	3.67
SSCTCPH1	science teacher academic push	2700	3.31	1.30
PSCPH1	parental science push	3107	1.63	1.20
PACPH1	parental academic push	3107	7.53	2.09
PCOPH1	parental college push	2482	1.94	1.10
SES3	family socioeconomic status	3096	-0.02	0.76
SFESX1	student's fear of success in class	2852	3.68	2.03
KSCPH1	peer science push	3107	1.16	1.20
KACPH1	peer academic push	3107	2.04	1.00
SNWAQ1	student news acquisition	3102	1.98	1.73

When Walberg and his colleagues developed the model of educational productivity, they first examined a simple productivity model (Parkerson, Lomax, Schiler, & Walberg, 1984). Only when the model did not fit a data base, had they investigated more complex models of variable relations. Similarly, to facilitate development of a new model of career aspiration, a simple model is chosen in this study to articulate relations between student career aspiration and factors of educational productivity (Figure 1). The structure of the simple model is similar to the one investigated by Parkerson, Lomax, Schiler, and Walberg (1984) except for the inclusion of variables on student career aspiration. The model structure is also in line with the description that "The Walberg productivity model posits direct, simultaneous influences of the nine factors on outcomes" (Reynolds, & Walberg, 1991, p. 98).

Besides the direct structural relations in Figure 1, correlations among the productivity factors were concurrently accommodated in this model. To add these relations, a total of 28 lines would have been included to connect the eight productivity factors. Moreover, additional graphics could be attached in Figure 1 to represent measurement errors of the multiple indicators. Due to mixture of these complex relations, the path diagram needs to be simplified to improve readability. In this study, only the direct path lines were drawn in Figure 1 to highlight the articulation between career aspiration and factors of educational productivity. Correlation coefficients and measurement errors are separately presented by tables in the next section (Tables 3 & 4).

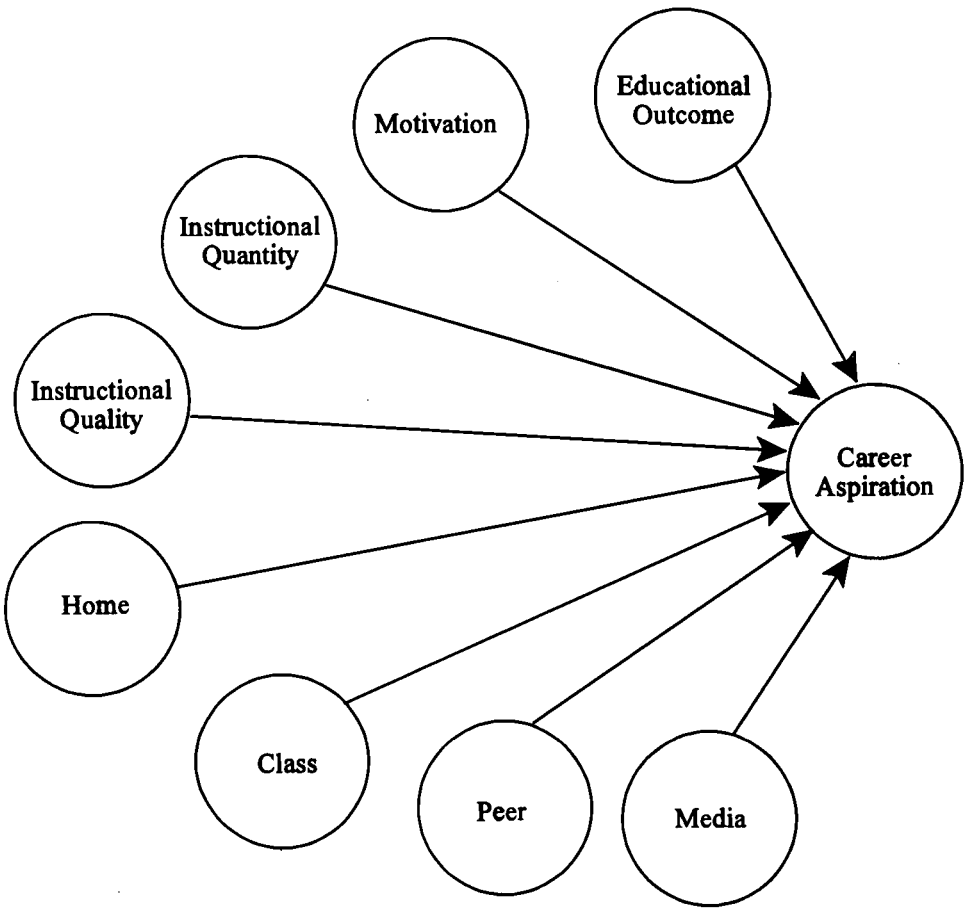

Figure 1. A Structural Model of Student Education and Career Aspiration

Results

Parameters describing the direct link between educational productivity and career aspiration were estimated by the maximum likelihood method and are presented in Figure 2. Since the whole LSAY sample was split into two halves, the results based on the even numbered half were enclosed within boxes (Figure 2). Data in the odd-numbered half converged to the results after 311 iterations, while the even-number half completed the convergence in 252 iterations. Correlations among the productivity factors are listed in Table 3.

For productivity factors with multiple indicators, factor loadings were calculated to reflect the weight of the LSAY composite variables on the latent factor construction (Table 4). Measurement errors were also estimated for the multiple indicators (Table 4).

The standardized root mean square residual (RMR) for the model was 0.05, and the goodness-of-fit index (GFI) was above 0.93 (Table 5). Both the low RMR and high GFI values suggested a good fit of the LSAY data for the structural equation model (Joreskog & Sorbom, 1993). In addition, the small RMSEA and large CFI indexes in Table 5 also supported the empirical model of career aspiration. Thus, the structural equation model was cross-validated by both halves of the LSAY sample.

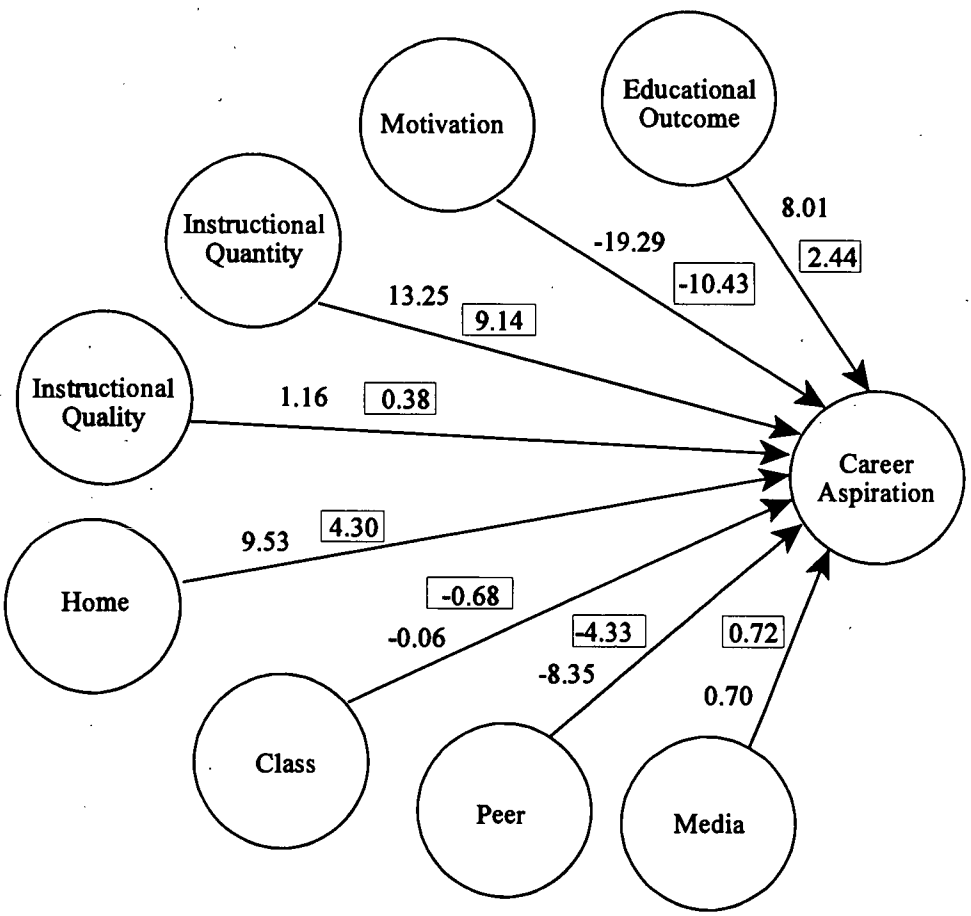

Figure 2. Parameters Describing the Direct Effects of Educational Productivity on Career Aspirations.

Table 3
Correlations of the Productivity Factors in the Structural Equation Model

	ability	motivation	quantity	quality	home	class	peer	media
Odd-Number Half								
ability	1.00							
motivation	0.89	1.00						
quantity	0.63	0.83	1.00					
quality	0.26	0.32	0.11	1.00				
home	0.61	0.64	0.37	0.22	1.00			
class	-0.03	-0.06	-0.04	0.02	0.03	1.00		
peer	0.50	0.57	0.63	0.05	0.78	0.08	1.00	
media	0.21	0.27	0.24	0.06	0.27	0.05	0.32	1.00
Even-Numbered Half								
ability	1.00							
motivation	0.70	1.00						
quantity	0.56	0.90	1.00					
quality	0.25	0.27	0.18	1.00				
home	0.53	0.58	0.45	0.32	1.00			
class	0.02	-0.09	0.01	-0.03	-0.00	1.00		
peer	0.42	0.45	0.56	0.24	0.75	0.10	1.00	
media	0.16	0.23	0.19	0.11	0.29	0.08	0.35	1.00

Table 4
Factor Loadings and Measurement Errors for Factors with Multiple Indicators

Productivity Factor	LSAY Variable	Factor Loading (w)	Measurement Error
Odd-Numbered Half			
Career Inspiration	AA22AS	0.28	0.04
	AASCIF	0.37	0.05
Education outcome	ASCIIRT	0.35	0.04
	SSCAT1	0.75	0.06
Motivation	SIPPS1	0.50	0.04
	SSFES1	0.47	0.04
Instructional quantity	AA27A	0.13	0.04
	SSHWE1	0.64	0.14
Home environment	PSCPH1	0.58	0.04
	PACPH1	0.52	0.04
	PCOPH1	0.33	0.04
	SES3	0.36	0.04
Peer environment	KSCPH1	0.49	0.04
	KACPH1	0.66	0.04
Even-Numbered Half			
Career Inspiration	AA22AS	0.23	0.04
	AASCIF	0.45	0.06
Education outcome	ASCIIRT	0.36	0.04
	SSCAT1	0.87	0.08
Motivation	SIPPS1	0.48	0.04
	SSFES1	0.53	0.04
Instructional quantity	AA27A	0.13	0.04
	SSHWE1	0.58	0.12
Home environment	PSCPH1	0.52	0.04
	PACPH1	0.50	0.04
	PCOPH1	0.29	0.04
	SES3	0.33	0.04
Peer environment	KSCPH1	0.55	0.04
	KACPH1	0.62	0.04

Table 5
The Model Fitness to the Split LSAY Data Bases

Indexes	Odd-Numbered Half	Even-Numbered Half
Comparative Fit Index (CFI)	0.78	0.80
Goodness of Fit Index (GFI)	0.93	0.94
Root MSE of Approximation (RMSEA)	0.07	0.07
Standardized Root MS Residual (STD RMR)	0.05	0.05

Discussion

While school outcomes can be measured in a variety of ways, student career aspiration and educational productivity are important indexes reflecting the condition of American welfare. Decker (1997) observed:

> Education appears to play an important role in worker productivity in all industrialized countries. The industrialized countries with the highest productivity levels tend to have highly educated work forces, and the convergence in productivity among these countries generally parallels that in educational attainment. (p. 5)

In this study, a positive link was found between career aspiration and education outcome (Figure 2). Students' *achievement in* and *attitude toward* science were indicators of *educational outcome* (Table 1). Some

researchers attached different importance to the *achievement* and *attitude* indicators. Berliner and Biddle (1995), for example, pointed out, "If schools are truly to serve the needs of business, it appears they should concentrate less on skill training and more on the values that students will need when they enter the workplace" (p. 89). Table 4 showed that the factor loading for student attitude was twice as high as that of achievement. Thus, the results seemed to reconfirm the importance of attitude variables on student career preparation.

In addition, according to the productivity theory, instructional factors were classified as *quality* and *quantity* of instruction (Reynolds & Walberg, 1991). The *quantity* was indicated by student weekly *homework hours* and *school work ethic* (Table 1). Miller et al. (1992) elaborated the scale of *school work ethic*:

> Student School Work Ethic - This scale measures the students attitudes and behaviors toward school work. The questions included are: How do you feel about the statement: I don't like to do any more school work than I have to; How often do you do each of the following: goof off in school; put off studying as long as I can; try hard to do my best in school; do the things I want to do first and leave my studying and other work to the last minute. (p. C2A-11)

Apparently, the composed *ethic* scale represented intrinsic efforts behind student responses to teachers' academic push and homework assignments. The teachers' academic push, on the other hand, indicated the *quality* of instruction (Table 1). Figure 2 showed that the quality factor had a ten-times smaller path coefficient than that of instructional quantity. Thus, this result and the factor loadings in Table 4 illustrated that the intrinsic effort played a more important role than the external teacher push or school assignment.

Another productivity factor with positive influences on career aspiration is student home environment (Figure 2). Trice, Hughes, Odom, Woods, and McClellan (1995) pointed out, "It has long been established that children aspire to the careers of their parents at rates significantly above chance" (p. 309). In the LSAY project, the parental career was a variable included in the family socioeconomic status (SES). According to Miller et al. (1992), SES was an "Equally weighted average of the non-missing values on (a) composite parental education, (b) composite parental occupational status, and (c) a household possessions index" (p. 48). The factor loadings suggested that SES had less weight on the home environment than the parental academic or science push (Table 4). Although SES may have indicated family status, it was the parental commitment that channeled the family resources to support kids' education. Thus far, many researchers have investigated the link between family SES and career aspiration (Grandy & Stahmann, 1974; Trice & McClellan, 1993; Trice, McClellan, & Hughes, 1992). In light of the factor loadings in Table 4, these former studies can be expanded to disentangle more profound variables, such as the contribution of parental education commitment.

Figure 2 also presented a weak impact from the mass media. In this study, the *media* environment was identified by a composite variable of *student news acquisition*. This variable reflected student utilization of news magazines, newspapers, and TV news (Miller et al., 1992). Table 3 indicated that *home* and *instructional quality* had weak correlations with the *media* environment. These results seemed to suggest that the news media were not strongly linked with the home and school settings. While it is against the First Amendment to control the news outlet, parents and teachers should make mindful discretion on the media exposure at home and in school. Perhaps, the enhanced discretion will strengthen the correlation between *news media* and *instructional quality*, and therefore, help raise career aspirations for students' futures.

On the other hand, negative path coefficients deserve special discussion. The most notable one was the negative coefficient from the *motivation* factor (Figure 2). In this case, the contribution from *motivation* was in sharp contrast with the positive link from *educational outcome* (Figure 2). The factor loading showed that the *motivation* factor was almost equally represented by indicators of student self-esteem and independence (Table 4). Since most seventh grade students were at an initial stage of establishing self-identity (Meeus, Geode, Kox, & Hurrelmann, 1992), it was possible that some students with less independence and unclear self-concepts might project their future jobs based on less realistic dreams. Therefore, the disagreement between student perceptions and performance was not entirely unusual. In the Third International Mathematics and Science Study (TIMSS), for instance, Beaton, Martin, Mullis, Gonzalez, Smith, and Kelly (1996) observed:

> In all except three countries, the majority of students agreed or strongly agreed that they did well in science or science subject areas - a perception that did not always coincide with the comparisons in achievement across countries on the TIMSS test. (p. 4)

Thus, the discrepancy, seemingly unlikely for adults, was quite plausible for the seventh grade students.

Another negative factor, *peer environment*, was indicated by two composite scales, *peer academic push* and *peer science push* (Table 1). Again, the negative feature

can be tracked back in the research literature. Bishop (1996) pointed out,

> In the United States, the peer group actively discourages academic effort. No adolescent wants to be considered a "nerd," "brain geek," or "grade grubber" or to be viewed as "acting white," yet that is what happens to students who study hard. (p. 83)

Similarly, the *class* environment was represented by a composite variable of classmates' negative pressure on school success. According to the LSAY codebook, this measure was based on students' agreement/disagreement with three survey items:

> In this school you were not very popular if you got better grades than other students; I worried that my success may make other students dislike me; I sometimes avoided discussing my accomplishments because other students might get jealous. (Miller et al., 1992, p. C2A-12)

Because these indicators reflected students' fear of academic success, it was no surprise to observe a negative correlation coefficient between the *motivation* and *class* environment (Table 3).

It should be noted that the *class* environment had weak links with other factors of educational productivity ($|r| \leq 0.10$), and the *peer* environment also had a weak correlation with *instructional quality* (Table 3). Perhaps, proper guidance from parents and teachers may help break the isolation of the *peer* and *class* environments, and facilitate positive interactions among peers at home and classmates in school.

In the LSAY project, student ability was measured by a test instrument from the *National Assessment of Educational Progress* (NAEP). The test performance was highly correlated with student motivation, instructional quantity, as well as the home and peer environments (Table 3). Reynolds and Walberg (1991) noted that the quantity of instruction "appears to be a particularly important variable, in that it mediates motivation, class environment, peer environment, and mass media" (p. 105). Although Reynolds and Walberg's (1991) analyses of the LSAY data did not include factors of career aspiration, similar links of *instructional quantity* were reconfirmed in this study with large correlation coefficients (Table 3).

In contrast, the *instructional quality* factor had much weaker correlations with other productivity factors (Table 3). Reynolds and Walberg (1992b) acknowledged that "detailed measures of instructional quality through observer ratings or surveys of classroom context (e.g., task orientation and organization of content) appear warranted. Unfortunately, such assessments are rarely possible in large-scale surveys with many other priorities, such as the present survey [LSAY]" (p. 381). Accordingly, the difference partially resulted from limitations of large-scale surveys, which failed to develop an in-depth indicator to penetrate the factor of instructional quality.

Although many researchers conducted studies on educational productivity, few investigations covered the factor of career aspiration. Trice and McClellan (1993) reported that half of mid-career adults had their current occupations closely associated with the early aspirations during the period of adolescence. However, at the seventh grade level, the factor loadings of career aspiration were not as high ($|w| \leq .45$) (Table 4). It can be suspected that education at upper grades may have strengthened the long-term influence on school-to-work transition.

Finally, it should be highlighted that all statistical findings were cross-validated by the split-sample analysis. Measurement errors were represented by the small *standard deviation for errors* in Table 4. Relative deviations among the model-fitting indexes were less than 2.56% (Table 5). These general agreements represented consistent confirmation of the structural equation model between the even- and odd-numbered subsamples.

In summary, career aspiration is one of the key factors behind the school-to-work transition (Raelin, 1980; Wilson, 1996). Under an assumption that factors of educational productivity also influence student school-to-work transition, Walberg's model of educational productivity was extended in this study to include the factor of career aspiration. The model was cross-validated by the split-half data from the LSAY seventh grade investigation. The statistical findings were interpreted in terms of a broad education context, including student aptitude-attributes, instructional characteristics, and psychological environment. Since no other national longitudinal studies besides LSAY have covered the seventh grade population, the model is subject to more data reconfirmations in the future.

References

Baker, D. P. (1997). Surviving TIMSS. *Phi Delta Kappan, 79*, 295-300.

Beaton, A., Martin, M., Mullis, I., Gonzalez, E., Smith, T. & Kelly, D. (1996). *Science achievement in the middle school years*. Chestnut Hill, MA: Boston College.

Bentler, P. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31*, 419-456.

Berliner, D. & Biddle, B. (1995). *The manufactured crisis.* New York, NY: Addison-Wesley.

Bishop, J. (1996). Singling the competencies of high school students to employers. In L. Resnick and J. Wirt (Eds.), *Linking School and Work.* San Francisco, CA: Jossey-Bass.

Chicago Academy of sciences (1995). *Longitudinal Study of American Youth: Base year through year 5.* Chicago, IL: The International Center for the Advancement of Scientific Literacy and The Chicago Academy of Sciences.

Davis, C. & Sonnenberg, B. (1995). *Programs and plans of the National Center for Education Statistics.* Washington, DC: U.S. Department of Education.

Decker, P. (1997). *Findings from education and the economy: An indicators report.* Washington, DC: U.S. Department of Education.

Fraser, B., Walberg, H., Welch, W., & Hattie, J. (1987). The need for educational productivity research. *International Journal of Educational Research, 11*(2), 147-154.

Grandy, T. G., & Stahmann, R. F. (1974). Family influence on college students' vocational choice: Predicting Holland's personality types. *Journal of College Student Personnel, 15*, 404-409.

Halperin, S., Melaville, A., & Taylor, A. (1988). *The forgotten half: Non-College youth in America.* Washington, DC: The William I Grant Foundation Commission on Work, Family, and Citizenship.

Hayduk, L. (1987). *Structural equation modeling with LISREL: Essentials and advances.* Baltimore, MD: Johns Hopkins University Press.

Hoffer, T. B. (1988, April). *The Longitudinal Study of American Youth and the National Education Longitudinal Study of 1988: A comparison.* Paper presented at the annual meeting of American Educational Research Association, New Orleans, LA.

Hoyle, R. (1995). *Structural equation modeling.* Thousand Oaks, CA: Sage.

Jenkins, J. T. (1998). *Career aspirations* [Online]. Available: http://www.peptalk.com/2career.htm [1998, May 12].

Joreskog, K., & Sorbom, D. (1993). *LISREL8 user's reference guide.* Chicago, IL: Scientific Software International.

Lavin, D. (1965). *The prediction of academic performance.* New York, NY: Russell Sage Foundation.

Medrich, E. (1996). *Preparation for work.* Washington, DC: U.S. Government Printing Office.

Meeus, W., Geode, M., Kox, W., & Hurrelmann, K. (1992). *Adolescence, careers, and cultures.* New York, NY: Walter de Gruyter.

Miller, J., & Brown, K. (1992). The development of career expectation by American youth. In W. Meeus, M. Goede, W. Kox, & K. Hurrelmann (Eds.), *Adolescence, careers, and cultures.* New York, NY: Walter de Gruyter.

Miller, J., Hoffer, T., Suchner, R., Brown, K., & Nelson, C. (1992). *LSAY codebook.* DeKalb, IL: Northern Illinois University.

Murnane, R. & Levy, F. (1996). *Teaching the new basics: Principles for educating children to thrive in a changing economy.* New York, NY: Free Press.

National Research Council (1996). *National science education standards.* Washington, DC: National Academy.

Parkerson, J. A., Lomax, R. L., Schiller, D. P., & Walberg, H. J. (1984). Exploring causal models of educational achievement. *Journal of Educational Psychology, 76* (4), 638-646.

Raelin, J. (1980). *Building a career.* Kalamazoo, MI: W. E. Upjohn Institute for Employment Research.

Reynolds, A. J., & Walberg, H. J. (1991). A structural model of science achievement. *Journal of Educational Psychology, 83*(1), 97-107.

Reynolds, A. J., & Walberg, H. J. (1992a). A process model of mathematics achievement and attitude. *Journal for Research in Mathematics Education, 23*(4), 306-328.

Reynolds, A. J., & Walberg, H. J. (1992b). A structural model of science achievement and attitude: An extension to high school. *Journal of Educational Psychology, 84*(3), 371-382.

Stevens, G., & Cho, J. (1985). Socioeconomic indexes and the new 1980 census occupational classification scheme. *Social Science Research, 14*, 142-168.

Stevens, G., & Featherman, D. (1981). A revised socioeconomic index of occupational status. *Social Science Research, 10*, 364-395.

Suter, L. E. (1992). *Indicators of science and mathematics education 1992* (NSF93-95). Washington, DC: National Science Foundation.

Trice, A. D., & McClellan, N. (1993). Do children's career aspirations predict adult occupations? *Psychological Reports, 72*, 367-370.

Trice, A. D., McClellan, N., & Hughes, M. A. (1992). Origins of children's career aspirations: II. Direct suggestions as a method of transmitting occupational preferences. *Psychological Reports, 71*, 253-254.

Trice, A. D., Hughes, M. A., Odom, C., Woods, K., & McClellan, N. (1995). The origins of children's career aspirations: IV. Testing hypotheses from four theories. *The Career Development Quarterly, 43*, 307-322.

Tyler, R. W. (1984). A guide to educational troubleshooting. *Educational Leadership, 41*(8), 27-30.

Walberg, H. (1981). A psychological theory of educational productivity. In F. H. Farley and N. Gorden (Eds.), *Psychology and education*. Berkerley, CA: McCutchan.

Walberg, H. (1982). What makes schooling effective? A synthesis and critique of three national studies. *Contemporary Education Review, 1*, 23-34.

Walberg, H. (1983). Scientific literacy and economic productivity in international perspective. *Daedalus, 112*, 1-28.

Walberg, H. (1984). Improving the productivity of America's schools. *Educational Leadership, 41*, 19-30.

Walberg, H. (1986). Synthesis of research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). Washington, DC: American Educational Research Association.

Wilson, J. W. (1996). *When work disappears*. New York, NY: Alfred A. Knopf.

Young, D., Reynolds, A., & Walberg, H. (1996). Science achievement and educational productivity: A hierarchical linear model. *Journal of Educational Research, 89*(5), 272-287.

JOURNAL SUBSCRIPTION FORM

This form can be used to subscribe to RESEARCH IN THE SCHOOLS without becoming a member of the Mid-South Educational Research Association. It can be used by individuals and institutions.

◇◇

Please enter a subscription to RESEARCH IN THE SCHOOLS for:

Name: ______________________________

Institution: ______________________________

Address: ______________________________

____

			COST
Individual Subscription ($25 per year)	Number of years	______	______
Institutional Subscription ($30 per year)	Number of years	______	______
Foreign Surcharge ($25 per year, applies to both individual and institutional subscriptions)	Number of years	______	______
Back issues for Volumes 1, 2, 3, and 4 ($30 per Volume)	Number of Volumes	______	______
TOTAL COST:			______

MAKE CHECKS PAYABLE TO MSERA
SEND FORM AND CHECK TO:

Dr. James E. McLean, Co-Editor
RESEARCH IN THE SCHOOLS
University of Alabama at Birmingham
School of Education, 233 Educ. Bldg.
901 13th Street, South
Birmingham, AL 35294-1250

MID-SOUTH EDUCATIONAL RESEARCH ASSOCIATION

The Mid-South Educational Research Association (MSERA) was founded in order to encourage quality educational research in the Mid-South and to promote the application of quality educational research in schools. Members of MSERA share interests in educational research, development, and evaluation. While most members are from institutions of higher education, many others represent state departments of education, public and private agencies, and public school systems. Graduate students comprise a significant portion of the membership. A majority of MSERA members are from the six states represented by the organization, but others are from many other states and several foreign countries. The MSERA is the largest regional educational research organization in the country.

The organization provides several services for its members. The annual meeting, held every November, offers many formal and informal opportunities for professional development through special training courses, sharing of research findings, and programmatic interests with colleagues. Members receive a subscription to *RESEARCH IN THE SCHOOLS* and the *Mid-South Educational Researcher*. The MSERA also provides recognition and cash rewards for its outstanding paper, an outstanding dissertation, and professional service.

MSERA Membership/Renewal Form

(Please print or type)

Name _____

Organization _____

Address _____

_____

Telephone Work: _____

Home: _____

Fax: _____

e-mail: _____

Amount Enclosed:

MSERA 1999 Membership ($25 professional, $15 student)	$_____
MSER Foundation Contribution	$_____
TOTAL	$_____

Make check out to MSERA and mail to:

Dr. Robert Calvery
Southside School District
70 Scott Drive
Batesville, Arkansas 72501

RESEARCH IN THE SCHOOLS
id-South Educational Research Association
d the University of Alabama at Birmingham
901 South 13th Street, Room 233
Birmingham, AL 35294-1250

RESEARCH IN THE SCHOOLS

A nationally refereed journal sponsored by the
Mid-South Educational Research Association
and the University of Alabama at Birmingham.

Volume 6, Number 2 **Fall 1999**

James E. McLean and Alan S. Kaufman, Editors

RESEARCH IN THE SCHOOLS

Information for Authors

Statement of Purpose

RESEARCH IN THE SCHOOLS (ISSN 1085-5300) publishes original contributions in the following areas: 1) *Research in Practice*--empirical studies focusing on the results of applied educational research including cross-cultural studies, 2) *Topical Articles*--scholarly reviews of research, perspectives on the use of research findings, theoretical articles, and related articles, 3) *Methods and Techniques*--descriptions of technology applications in the classroom, descriptions of innovative teaching strategies in research/measurement/statistics, evaluations of teaching methods, and similar articles of interest to instructors of research-oriented courses, 4) *Assessment*--empirical studies of norm-referenced, criterion-referenced, and informal tests in the areas of cognitive ability, academic achievement, personality, vocational interests, neuropsychological functioning, and the like, and 5) *Other* topics of interest to educational researchers. *RESEARCH IN THE SCHOOLS* is devoted to research conducted in *any* educational setting from a conventional elementary school or high school to a training program conducted within an industry. Likewise, there are no age restrictions on the sample, since the educational settings may include preschools, continuing education classes for adults, or adaptive skills courses in nursing homes. Studies conducted in settings such as clinics, hospitals, or prisons are ordinarily inappropriate for *RESEARCH IN THE SCHOOLS* unless they involve an educational program within such a setting. One goal of *RESEARCH IN THE SCHOOLS* is to provide a training ground for graduate students to learn effective reviewing techniques. Consequently, the journal utilizes a Graduate Student Editorial Board composed mostly of students in educational psychology and educational research. Members of this Editorial Board, each sponsored by a professor, provide supplementary reviews for a selection of submitted articles, and receive both direct and indirect feedback of the quality of these reviews.

Preparing Manuscripts

Authors should prepare manuscripts in accordance with the stylistic rules and guidelines delineated in the *Publications Manual of the American Psychological Association* (4th ed., 1994), which is available from: Order Department, American Psychological Association, PO Box 2710, Hyattsville, MD 20784. Number the pages consecutively. All manuscripts will be subject to editing for sexist language.

Author Identification

Authors should put the complete title of the article on the first text page, but they should exclude their names. Subsequent pages should include only a running head. They should prepare a separate sheet with the complete title of the article and their names and affiliations; this procedure will ensure anonymity in the review process. Authors should supply addresses and phone numbers, and electronic mail addresses and fax numbers (if available), for potential use by the editorial staff and, later, by the production staff. Unless otherwise stated, the first-named author will be sent correspondence, galley proofs, copyright forms, and so forth.

Submission of Manuscripts

Submit manuscripts in triplicate to **James E. McLean, Co-Editor, RESEARCH IN THE SCHOOLS, School of Education, 233 Educ. Bldg., The University of Alabama at Birmingham, 901 13th Street, South, Birmingham, AL 35294-1250. Please direct questions to jmclean@uab.edu.** All copies should be clear and readable; dot matrix is acceptable only if it meets these qualities of legibility. Length of the manuscripts, including references and tables, should ordinarily range from about 10 to 40 typed, double-spaced, 8-1/2 X 11-inch pages, using 11-12 point type. Abstracts are limited to 125 words. Brief reports of research are not encouraged. Authors are encouraged to keep a hard copy of the manuscript to guard against loss. It is assumed that all manuscripts submitted for publication are original material and have not been simultaneously submitted for publication elsewhere. When manuscripts are accepted for publication, authors are encouraged to submit the final version on a computer disk along with the hard copy.

Copyright and Permissions

Volume 6, Number 2 **Fall 1999**

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 2, 1-8

Examining Professional Development of Principals As Adult Learners

Diane Ricciardi
University of Louisville

The study explored professional development needs related to demographic characteristics of principals and examined the relationship of professional development needs and self-reported job performance ratings. Principals (N = 140) gave recommendations for improving professional development and rated the usefulness of professional development content and the effectiveness of training delivery methods. A positive relationship was found between professional development needs and self-reported job performance ratings. Professional development content was not balanced among the 21 performance domains identified by NPBEA. Principals reported lecture to be the primary delivery method used for training; yet, lecture received mixed ratings of effectiveness. It is suggested that applying adult learning theory could enrich professional development opportunities for principals. Study results supported the work of the NPBEA and numerous adult learning theorists.

Many issues in the 1990s confront principals with an increased sense of urgency and serve to change their role as building leaders. Social conditions alter the way schools provide services to children. The emerging policy environment calls for higher achievement for a broader range of students. Site governance increases the scope of school leadership (Alexander, 1986). These issues and others continue to reshape the job of the principal and to underscore the importance of well trained school leaders.

Attention to the professional growth of principals is timely. Yet most training activities focus on preservice education and on induction needs of new principals (Erlandson, 1994). Training for experienced principals appears to be the most neglected, least developed component of training. Barth (1993), an advocate of professional development for principals, stressed, ". . . I am convinced that being a learner, a lifelong adult learner, is the most important characteristic of a school leader and of a professional" (p. 219).

Even though the worth of professional development is seldom argued, one questions whether the institution of the principalship assists or inhibits an individual's professional growth, and whether principals are equipped to handle the rapid changes occurring in schools. As universities, professional organizations, and school districts provide training, each must consider principals' unique learning needs. To address the professional growth of principals, training needs and training effectiveness must be analyzed.

This article summarizes findings of a recent study that examines the professional development for practicing principals. Using the andragogical model of adult learning theory (Knowles, 1984), the author analyzes training topics, delivery methods, and recommendations for improving training. Supported by the framework of adult learning theory and study findings, a model for improving principals' professional development is presented. The study addressed several questions:

1. What are the perceived professional development needs of principals?
2. How do professional development needs vary according to demographic characteristics of principals and schools: (a) years working as a principal, (b) years working in education, (c) immediate prior position in education, (d) level of formal training, (e) gender, (f) school level, (g) school size, (h) school location, and (I) school socioeconomic level?
3. How do professional development needs relate to the self-reported job performance ratings that principals receive from supervisors?
4. What is the content of professional development for principals, and how do principals rate the usefulness of the content?

Diane Ricciardi, Assistant Professor, Department of Administration and Higher Education at the University of Louisville, focuses on training of school principals. She has served as a middle school principal, high school guidance director, counselor, and teacher in the South Carolina public schools. Correspondence regarding this article should be directed to Dianne Ricciardi, Department of Administration and Higher Education, University of Louisville, #329 School of Education, Louisville, KY 40292 or by e-mail at pdricc01@athena.louisville.edu.

5. What methods are used to deliver professional development, and how do principals rate the effectiveness of delivery methods?
6. What recommendations do principals make to improve professional development?
7. How do study findings support adult learning theory?

First, the author examined the conceptual framework of adult learning theory and the current literature supporting this study.

Conceptual Framework

Historically, training programs for educators have not been grounded in adult learning theory. Schools, designed around the needs of youth, seldom serve the professional growth needs of the adult workers (Levine, 1982). Noting that children and adults learn differently, Knowles (1988) found that adults need experiences that offer strong connections to the workplace, ample opportunities for self-directed learning, and a wide-range of instructional approaches. Louis (1993) noted that "there are no strategies for development that adequately take into account the needs of educational leaders at different stages of their careers," nor are there systematic efforts "to engage administrators in regularly updating their skills and knowledge to deal with critical changes in their settings" (p. 15). Little research has examined the alignment of professional development for principals and adult learning needs. This lack of theoretical base is problematic, and the needs of adult learners have not been operationalized fully into training practices for principals (Pitner, 1982).

Noting lack of field-based experiences as a major concern, leaders of the National Association of Secondary School Principals (NASSP) called for reform of principal preparation programs. In its monograph, *Developing School Leaders: A Call for Collaboration* (NASSP, 1992), the organization framed professional development for principals as a seamless garment that should extend through preservice, induction, career changes, and retirement. The national call for reform countered past assumptions that principal training and certification programs both qualified an individual for the job and provided the individual with all the skills needed throughout the career.

Responding to the call for reform, the National Policy Board for Educational Administration (NPBEA) (1990) worked to redesign preparation programs for principals and to examine the need for a national certification process. The group studied the changing nature of the work of principals and determined that a mismatch between training programs and job requirements existed. The NPBEA developed a framework of 21 job performance domains (see Table 1) and recommended delivery systems for training principals.

Recently, the Interstate School Leaders Licensure Consortium (ISLLC), sponsored by the Council of Chief State School Officers (1997), released standards and indicators describing the knowledge, performances, and dispositions needed by school administrators. The consortium's new assessment for administrators is designed to test application of administrative knowledge in the workplace. The new ISLLC standards and assessment likely will impact principal training and certification during the coming decade.

Method

Participants

The population for this study consisted of 883 experienced public school principals listed in the South Carolina Basic Educational Data System. A stratified sample by gender and school levels ($N = 275$) was drawn using a table of random numbers. Of the sample, 37% were female, 58% served in elementary schools, 19% worked in middle schools, and 23% were in high schools. A sample of this size was expected to produce less than a 10% degree of error (Minium, 1978). An initial mailing and one follow-up mailing yielded a 51% response rate. The response rate met the minimum rate of 50% considered adequate for analysis and reporting (Babbie, 1990). Of the 140 principals returning the questionnaire, 37% were female, 46% represented elementary schools, 31% were from middle schools, and 23% were from high schools. This response was fairly representative of principals statewide by gender and school level, with a slight overrepresentation of middle school principals and a slight underrepresentation of elementary school principals. Respondents averaged 22.3 years working in public education and 10 years serving as a school principal. Both averages were similar to national averages (Doud, 1988).

Questionnaire

The questionnaire, *Survey of Professional Development Needs of Experienced Principals,* was designed with the assistance of expert reviewers and was piloted and tested for content validity and reliability by a panel of 11 practicing principals. The questionnaire incorporated the 21 job performance domains from the National Policy Board for Educational Administration (NPBEA). These domains, developed over a number of years, represented current, well-respected research findings of national organizations dedicated to the improvement of educational leadership. The NPBEA grouped the 21 performance domains into four content domain categories. The

functional domains deal with administrative processes and techniques that allow the school to function. The *programmatic* domains relate to the core technologies of curriculum and instruction and the school services that support student learning. The *interpersonal* domains deal with the value of human relationships and communication within schools. The *contextual* domains relate to the cultural, economic, and political influences surrounding schools. In this study, internal consistency reliability was estimated by computing alpha coefficients for each of the four content domain categories used to rate professional development need and for the total group of 21 performance domains. These coefficients ranged from .90 to .97, indicating the survey instrument had an acceptable level of internal consistency above the minimum (.70) recommended by Nunnally (1978) for use of composite scores in statistical analysis.

The first part of the questionnaire collected information about the independent variables which included demographic data about principals and schools. The second part of the questionnaire consisted of 21 items, each asking two questions. In the first question, principals rated their perceived level of training need in each of the 21 job performance domains. Ratings were given on four-point Likert-type scales (1 = *Not a Need*; 2 = *Mildly Important Need*; 3 = *Important Need*; 4 = *Extremely Important Need*). A higher rating indicated a greater perceived level of professional development need in each of the 21 performance domains. In the second question, principals reported their most recent job performance evaluation ratings given by supervisors in each of the separate 21 performance domains. These self-reported ratings were given on four-point Likert-type scales (1 = *Exemplary*; 2 = *More Than Satisfactory*; 3 = *Satisfactory*; 4 = *Needs Improvement*). A higher rating indicated a greater need for job performance improvement in each of the separate 21 performance domains.

On the third part of the questionnaire, principals listed the professional training activities in which they had participated during the past two years, and rated the usefulness of each training activity on four-point Likert-type scales (1 = *Not Useful*; 2 = *Mildly Useful*; 3 = *Useful*; 4 = *Extremely Useful*). Next, principals indicated the delivery method used for each activity (1 = *Lecture Workshop*; 2 = *Peer Coaching*; 3 = *Small Group Instruction*; 4 = *Graduate Course Format*; 5 = *Hands-On Simulation*; 6 = *Other*) and rated the effectiveness of each delivery method on four-point Likert-type scales (1 = *Not Effective*; 2 = *Mildly Effective*; 3 = *Effective*; 4 = *Extremely Effective*). In the last part of the questionnaire, principals answered three open-ended questions that elicited participant recommendations for improving the usefulness and accessibility of, and support for professional development.

Procedure

The questionnaire, a self-addressed, stamped envelope, and a cover letter in which anonymity was assured, were mailed to the principals in the fall semester. Each survey was coded to maintain confidentiality of respondents and to allow a follow-up mailing to non-respondents. Three weeks after the first mailing, subjects who had not returned the questionnaire received a second set of materials.

Once received, quantitative data were processed using SAS statistical software. Frequencies and percentages were compiled for demographic information and to answer questions about training needs, participation, ratings of usefulness, and effectiveness of delivery methods. Stepwise Regression Analyses were performed to identify the relationship between demographic factors and principals' professional development training needs in the four content domain categories. Pearson Correlation Coefficients were computed to examine the relationship between perceived professional development needs and self-reported job performance improvement ratings. A content analysis served to categorize principals' recommendations for improving professional training and to relate these findings to the framework of adult learning theory. A coding sheet was used to tally recommendations into categories. Two independent researchers replicated the content analysis to verify accuracy. Adjustments to increase the clarity of category descriptions were made based upon recommendations from the additional reviewers.

Findings

Professional Development Needs

Principals identified their most important professional training needs in the specific domains of (a) Curriculum Design and (b) Instruction and the Learning Environment. (See Table 1.) The National Policy Board for Educational Administration found these two performance domains to be the most important for improving K-12 student achievement. Principals also gave priority to training needs which focused on the daily chores involved in running schools. The NPBEA contended that training in the domains should begin early in preparation programs and should be offered to principals throughout their careers. Participant ratings of the 21 professional development need domains appear in Table 1 in rank order according to participant mean score.

Table 1
Rank Order by Mean Score for Importance of Professional Development Needs

Performance Domain Areas	*M*	*SD*
Curriculum Design	3.25	.92
Instruction and the Learning Environment	3.23	1.16
Staff Development	3.13	.98
Legal and Regulatory Applications	3.08	.91
Problem Analysis	3.07	1.02
Public and Media Relationships	3.05	.92
Motivating Others	3.04	1.01
Measurement and Evaluation	3.02	.82
Sensitivity	2.99	1.02
Judgment	2.94	1.14
Implementation	2.92	1.04
Written Expression	2.89	1.08
Leadership	2.88	1.06
Delegation	2.88	.99
Oral and Nonverbal Expression	2.87	1.03
Student Guidance and Development	2.87	.94
Organizational Oversight	2.86	1.04
Policy and Political Influences	2.81	.93
Resource Allocation	2.81	1.05
Information Collection	2.77	.92
Philosophical and Cultural Values	2.67	.94

Relationship of Needs and Demographic Characteristics

The independent variables comprising demographic characteristics of the participants and their schools (i.e., gender, prior position, years in education, years in the principalship, school size, school location, school level) were examined as possible predictors of professional development needs. The 21 professional development need domains comprised four content domain categories: (a) functional, (b) programmatic, (c) interpersonal, and (d) contextual. Additive composite scores were formed from the items within the four content domain categories and served as dependent variables. Stepwise Multiple Regression Analyses were used to regress the independent variables on each of the four composite scores.

Results of the above analyses rendered two significant findings, both related to the participant school location. Status as a rural principal was a significant predictor of the composite rating of need for interpersonal professional development ($R^2 = .03$, $p < .05$). The relationship between rural status and the dependent variable was negative ($b = -1.29$), indicating that rural principals perceived significantly less need for interpersonal professional development than did non-rural principals. Also, status as a suburban principal was a significant predictor of the composite rating of need for contextual professional development ($R^2 = .03$, $p < .05$). The relationship between suburban status and the dependent variable was positive ($b = 1.17$), indicating that suburban principals perceived significantly more need for contextual professional development than did non-suburban principals. No other relationships between the demographic variables and the dependent variables of professional development need met the .05 level of significance.

Relationship of Needs to Self-Reported Job Performance Improvement Ratings

For this correlational analysis, principals self-reported the most recent job performance improvement ratings given to them by supervisors during formal performance evaluations. Job performance improvement ratings were analyzed in each of the 21 job performance domains. (See Table 2). These self-reported ratings were correlated to principals' professional development needs for each of the 21 performance domains. A positive correlation in each of the 21 performance domains was found. All correlations were significant, although the degree of association was moderate or weak at best. In this analysis, the positive correlations were logical because they demonstrated that perceived levels of professional development needs increased as ratings of the need for job performance improvement increased.

The strongest correlation (® = .42, $p < .001$) fell in the specific performance domain area of Implementation. Other correlations of professional development needs with need for job performance improvement were found in the domains of (a) Written Expression, (b) Measurement and Evaluation, (c) Student Guidance and Development, (d) Organizational Oversight, and (e) Problem Analysis. All other correlations indicated only moderate relationships between the variables of perceived professional development needs and the self-reported job performance improvement ratings. For the highest areas of correlation, Table 2 provides means, standard deviations, and correlation coefficients which meet the .001 level of significance.

Table 2
Highest Correlation Between Perceived Professional Development Needs and Self-Reported Job performance Improvement Ratings Last Given by Supervisors

Performance Domain Areas	M	SD	r
Implementation			.42*
Professional Development Need	2.921	1.039	
Job Performance Improvement Rating	1.831	.857	
Written Expression			.35*
Professional Development Need	2.896	1.076	
Job Performance Improvement Rating	1.864	.836	
Measurement and Evaluation			.34*
Professional Development Need	3.022	.821	
Job Performance Improvement Rating	2.419	.822	
Student Guidance and Development			.33*
Professional Development Need	2.870	.939	
Job Performance Improvement Rating	2.235	.836	
Organizational Oversight			.31*
Professional Development Need	2.857	1.042	
Job Performance Improvement Rating	2.043	.873	
Problem Analysis			.30*
Professional Development Need	3.071	1.022	
Job Performance Improvement Rating	1.993	.895	

*$p < .001$.

Participation In and Usefulness Of Professional Development Content

The principals identified 415 professional development activities in which they had participated. The 415 activities are summarized according to the content categories shown in Table 3. The four right-hand columns in Table 3 report the percentages for participant ratings of usefulness for activities in each category. The highest participation categories were: (a) Instruction and the Learning Environment, (b) Curriculum Design, (c) Legal and Regulatory Applications, and (d) Leadership. When comparing participation to perceived professional development needs, principals reported that professional development in (a) Instruction and the Learning Environment, (b) Curriculum Design, and (c) Legal and Regulatory Applications represented both their top professional development needs and their highest levels of participation. Principals also reported top professional development needs in (a) Problem Analysis and (b) Public and Media Relationships; however, they reported extremely low participation in these training areas. Only 7 (2%) of the 415 professional development activities attended by principals dealt with Public and Media Relationships, and only 1 (.1%) of the 415 sessions was about Problem Analysis. The low participation may indicate that training content targeting these topics was unavailable or was offered at inconvenient times or locations.

Overall, principals reported extremely low participation in professional development activities in 11 of the 21 performance domains. Principals attended fewer than 10 activities in each area: (a) Resource Allocation, (b) Public and Media Relationships, (c) Information Collection, (d) Problem Analysis, (e) Oral and Nonverbal Expression, (f) Written Expression, (g) Philosophical and Cultural Values, (h) Policy and Political Influences, (I) Judgment, (j) Implementation, and (k) Delegation. This low participation showed that principals were not exposed to professional development content that was balanced across the twenty-one performance areas identified as being critical for job performance (NPBEA, 1990).

Principals rated 87% of all professional development activities in which they participated as useful, rating (a) Instruction and the Learning Environment, (b) Curriculum Design, and (c) Legal and Regulatory Application as the content most useful to their jobs. These findings confirmed NPBEA findings (1990) that (a) Instruction and the Learning Environment and (b) Curriculum Design provide direction for principals' performance in all other areas and serve as the core knowledge and skills that principals need to improve student achievement in their schools. Table 3 presents the frequencies of professional development activity participation and ratings of activity usefulness.

Table 3
Percentages for Professional Development Activity Categories and Levels of Activity Usefulness

Professional Development Content Areas	Total Activities %	Not Useful %	Mildly Useful %	Useful %	Extremely Useful %
Instruction and the Learning Environment	22%	1%	8%	46%	46%
Leadership	13%	0%	15%	31%	54%
Legal and Regulatory Applications	13%	0%	9%	51%	40%
Curriculum Design	13%	0%	9%	29%	62%
Staff Development	8%	0%	19%	58%	23%
Measurement and Evaluation	6%	8%	4%	64%	24%
Organization Oversight	6%	4%	8%	48%	30%
Student Guidance and Development	6%	4%	15%	31%	50%
Sensitivity	5%	0%	0%	42%	58%
Motivating Others	3%	0%	15%	31%	54%
Public and Media Relationships	3%	0%	43%	57%	0%
Resource Allocation	2%	0%	25%	63%	13%
Total	100%	1%	12%	43%	44%

Note. All training areas that correspond to percents that were <1 were omitted.

Delivery Methods for Professional Development Training

Principals provided information about delivery methods used for 412 professional development activities. (See Table 4.) Principals reported that the lecture workshop, used for 59% of all activities, was the primary delivery method for training; however, this method received mixed ratings of effectiveness. Approximately 32% of these training activities utilizing lecture were not rated as effective. Principals, therefore, did not believe that the lecture method was appropriate for all training activities. This finding supported work by Gemar (1979), who reported that short, how-to lecture workshops were not effective in helping principals build a broad base of knowledge or skills.

Other delivery methods, categorized as (a) hands-on simulation, (b) small group instruction, (c) graduate-course format, and (d) peer coaching, were used less frequently than the lecture method. Principals rated these other methods as very effective and gave many recommendations for using these methods more often. Although, principals rarely had opportunities to participate in alternative delivery methods, when they did participate, they rated these methods positively. Principals wanted delivery methods which increased their peer contacts, offered more intensive content knowledge, and expanded their active involvement in learning activities. Table 4 represents the frequencies of delivery method types used for training and the effectiveness ratings by delivery method type.

Table 4
Frequencies of Delivery Method Type Used for Professional Development Training and Effectiveness Ratings by Delivery Method Type

Delivery Method Type	Delivery Method Frequency (%)	Effectiveness Rating Scale of Delivery Method Type: Not Effective %	Mildly Effective %	Effective %	Extremely Effective %
Lecture Workshop	242 (59%)	8%	24%	45%	23%
Hands-On Simulation	68 (17%)	3%	10%	31%	56%
Small Group Instruction	48 (12%)	0%	10%	75%	15%
Graduate-Course Format	36 (9%)	0%	11%	61%	28%
Peer Coaching	18 (4%)	0%	5%	50%	45%
Total	412 (100%)	5%	18%	49%	28%

Recommendations for Improving Professional Development

Principals gave 204 written recommendations for improving professional development training. The researcher labeled the recommendations as they fell into five primary areas of concern and then calculated frequencies and percentages from each category. As indicated in Table 5, categories of concern included (a) the time when training was offered; (b) the content of training activities; (c) the location where training was held; (d) the networking support before, during, and after training; and (e) the availability of follow-up activities to support training. Principals indicated that the demands of the job often limited participation in professional development activities. Principals recommended that more convenient locations and times be selected for professional training, and that follow-up activities and networking with their peers would be useful. Principals recommended that training content be available on current issues impacting schools and be tailored to fit individual and situational needs. Table 5 presents the frequencies and percentage groupings of recommendations given by principals for improving professional development.

Discussion

Findings of this study support operationalizing the framework of adult learning theory into professional development experiences for principals. In the andragogical model (Knowles, 1984), adults learn knowledge and skills which allow them to perform social and professional roles. In this study, principals desired training which had a high degree of connection to their workplace. Principals prioritized learning needs in the areas of (a) Curriculum Design and (b) Instruction and the Learning Environment, both viewed by the NPBEA as the domains most important to the primary function of school leaders. Principals desired more training in the areas of legal issues, problem analysis, and various interpersonal skills. Findings of this study can be generalized to findings of the NPBEA which show that principals want training topics which are relevant and give them the knowledge and skills needed to deal with the day-to-day operations of schools.

Table 5
Rank Order Distribution of Frequencies and Percentages of Concerns Given As Recommendations to Improve Professional Development

Concern	Number of Recommendations	% of concerns
Time	48	24%
The time in which training is offered is not convenient for principals and could be changed to increase the usefulness of and accessibility to training.		
Training Content	45	22%
The content of training could address specific issues and individual needs of principals and schools to be more useful.		
Networking Support	41	20%
Contact with other principals before, during, and after training could increase the usefulness of training and could be helpful to principals.		
Location	36	18%
The location for training could be changed to improve the accessibility of training and the usefulness of training.		
Follow-up Training	34	16%
Follow-up training could be helpful to principals and could increase the usefulness of training.		
Total Recommendations	204	100%

Findings from the study support research of several learning theorists (Knowles, 1988; Levine, 1982; Rogers, 1980) who found adult learners to be both self-directed and problem-oriented. These findings support general recommendations for serving the needs of principals as adult learners to include: (a) developing systems to match the content and delivery methods of training to the professional needs of individual principals, (b) extending the range of learning opportunities provided in training, (c) encouraging principals to take a greater role in their own professional growth, and (d) supporting greater

collaboration for training efforts among professional organizations, universities, state departments of education, and local school districts.

A prescribed set of professional development needs cannot be identified for principals based upon individual demographic characteristics. Instead, many interwoven factors affect the knowledge and skills needed by principals. This conclusion supports Knowles's contention (1984) that adult learning experiences are connected to past experiences and are highly situational. Yet the dominant design of professional development for principals does not account for prior experiences. Past knowledge and skills that principals possess cause learning needs to be highly individualized and situational. Principals' training needs change over time, change from one school to the next, and differ according to past experiences.

Principals want training methods which vary from lecture workshops. Desiring interactive professional learning opportunities, principals gave the lecture method mixed ratings of effectiveness. Other methods of (a) hands-on simulation, (b) small group instruction, (c) graduate-course format, and (d) peer coaching, though used infrequently, were rated as highly effective. This finding supports the principle of adult learning theory which recognizes adults to be interactive and self-directed in their learning.

Recommendations and Conclusion

Five recommendations for improving professional development complete this article. Principals recommend that the timing for training be convenient and flexible to minimize time they spend away from their schools. Second, principals value what they can learn from peers and recommend that more principals serve as trainers and mentors. They desire training methods which engage them in hands-on learning activities working with peers. Third, principals want professional development to include more follow-up support and to be available on an on-going basis. Fourth, principals are concerned about the relevancy of training content and want content to focus on timely issues related to individual and school needs. Finally, principals express concern about the inconvenience associated with travel to training sessions which often are held in locations far from their schools making attendance difficult.

Results of this study should be interpreted with caution due to certain limitations. Findings are based on data from a response rate of 51%. Results may have varied if a greater number of participants had responded. Also, due to the self-reporting nature of the questionnaire, data concerning job performance improvement ratings and perceptions of professional development needs may have differed from data based on actual performance evaluations provided by principal supervisors. Because job evaluations are sensitive and confidential, for the purpose of this study, the author used participant self-reported data.

However, despite limitations, the present study does provide direction for improving professional development training for school principals. Findings suggest that training content and methods can be enriched with the application of adult learning theory. Staff developers can design more effective delivery methods for training that can address both the individual and situational needs of particular principals. This, in turn, can insure that future professional development activities, framed within the context of adult learning theory, guarantee greater job relevancy and provide principals with a high degree of involvement and self-direction in their own professional growth.

References

Alexander, L. (1986). Time for results: An overview. *Phi Delta Kappan*, 68, 202-204.

Babbie, E. (1990). *Survey research methods.* Belmont, CA: Wadsworth Publishing Company.

Barth, R. S. (1993). Reflections on a conversation. *Journal of Personnel Evaluation in Education*, 7, 217-221.

Council of Chief State School Officers (1997). *Using ISLLC standards to strengthen preparation programs in school administration.* Washington: Council of Chief State School Officers.

Doud, J. (1988). *The K-8 principal in 1988.* Alexandria, VA: National Association of Elementary School Principals.

Erlandson, D. A. (1994). *Building a career: Fulfilling the lifetime professional needs of principals.* Fairfax, VA: National Policy Board for Educational Administration.

Gemar, R. A. (1979, March). State of the art -- There are programs available. *Thrust for Educational Leadership*, 8, 15-17.

Knowles, M. (1984). Introduction: The art and science of helping adults learn. In M. Knowles (Ed.), *Andragogy in action: Applying modern principles of adult learning.* San Francisco: Jossey-Bass, Inc.

Knowles, M. (1988). *The modern practice of adult education: From pedagogy to andragogy.* Cambridge, MA: The Adult Education Co.

Levine, S. L. (1982). *Promoting adult growth in schools: The promise of professional development.* Boston: Allyn and Bacon.

Louis, K. S. (1993). *Educational leadership development in Minnesota: A non-system in need of transformation.* Unpublished manuscript.

Minium, E. W. (1978). *Statistical reasoning in psychology and education.* New York: John Wiley & Sons.

National Association of Secondary School Principals (1992). *Developing school leaders: A call for collaboration.* Reston, VA: National Association of Secondary School Principals.

National Policy Board for Educational Administration (1990). *Principals for our changing schools: Preparation and certification.* Fairfax, VA: Author.

Nunnally, J. C. (1978). *Psychometric theory.* New York: McGraw-Hill.

Pitner, N. J. (1982). *Training of the school administrator: State of the art.* Eugene, OR: Center for Educational Policy and Management, University of Oregon.

Rogers, C. R. (1980). *A way of being.* Boston: Houghton Mifflin.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 2, 9-24

"Drafted Into the Front Lines": Teacher Efficacy During School Desegregation in Columbus, Georgia, 1968-1975

Virginia E. Causey
Georgia State University

Despite federally coerced desegregation three decades ago, many schools today are more segregated than ever. Based on archival research and interviews with 40 participants, this study examined the historical roots and present significance of desegregation for those on the front line--teachers, administrators, and students. This case study of Columbus, Georgia, reveals that the demands of White constituencies remained the priority. Teachers of both races adopted a "color-blind" approach emphasizing the needs of White students. Teachers abdicated efficacy through focusing on survival and a "business as usual" approach. Black teachers lost leadership roles and endured second-class treatment. Despite pervasive conflict, many teachers experienced personal and professional growth and believed benefits came from desegregation. Their experiences resonate as educators grapple with issues of race and equity today.

A social revolution engulfed the South during the early 1970s. A hundred years of Jim Crow segregation in public schools was swept away almost overnight by federal force. "Drafted into the front lines" as the unlikely foot soldiers of this revolution (Crain, Mahard, & Narot, 1982, p. 97) were teachers, often unwilling and unprepared. The experiences of teachers in Columbus, a city of 150,000 people in west central Georgia, in the last years of freedom of choice desegregation and the first years of forced desegregation open a window into that turbulent time and illuminate implications for educators thirty years later.

Design of the Study

This study investigated the effects of desegregation on the efficacy of teachers in Columbus high schools. Columbus, part of the consolidated Muscogee County School District (MCSD), had five White high schools--two considered college preparatory and three working class. It had two Black high schools. During the years 1968-1975, the student population was about 70 percent White and 30 percent Black, and Black teachers made up about 25 percent of the teaching force.

The effort to uncover the significance of teachers' experiences involved three "cycles" of data collection. The first cycle focused on an exploration of court records, school district documents, school board minutes, newspaper accounts, and other contemporary and historic studies of the effects of desegregation in order to understand the context of desegregation in Columbus and the South. The second cycle produced the richest data: 40 interviews with administrators, teachers, and students from a variety of backgrounds who participated in school desegregation in Columbus. The author interviewed 17 Black and 17 White respondents, including five administrators, 19 teachers, and 10 students.[1] She also analyzed six additional oral history interviews from the Columbus State University Archives with three administrators, two teachers, and one student.[2] The final cycle narrowed in focus for more intense data collection on emergent themes. The author used the document base to triangulate with interview data in order to reduce bias and nostalgia and to gain as clear a picture as possible of the experience of teachers during desegregation.

Themes that arose from the study included the role race played in teacher assignments during freedom of choice and forced desegregation; teachers' reactions to forced desegregation; racial relations between teachers; issues of school leadership and ownership; crisis and growth experienced by teachers; the influence of desegregation on curriculum and pedagogy; how race affected perceptions about students and teachers; cultural conflict between teachers and students and its effect on discipline

Virginia E. Causey is Assistant Professor in the Department of Middle/Secondary Education and Instructional Technology at Georgia State University. Correspondence concerning this article should be directed to Virginia E. Causey, Dept. of Middle/Secondary Educ & Instructional Tech., Georgia State University, University Plaza, Atlanta, GA 30303 or by e-mail to vcausey@gsu.edu.

[1] Interviews were audiotaped. Transcripts and audiotapes were deposited in the archives at Columbus State University.

[2] See Table 1 for a description of the respondents according to position in the school district, gender, ethnicity, experience, and age. See Tables 2 and 3 for the interview protocols used by the author.

and instruction; and inservice opportunities for teachers. Emerging from the research was the pervasive role of White power and privilege in school desegregation. Whether overtly or subtly, meeting the needs of White constituencies remained the priority in Columbus schooling.

Table 1
Respondent Backgrounds

Name*	Sex	Race	Year of Birth	Years Teaching Experience in 1970	Occupation in 1997
Administrators:					
James Collier	M	Black	1923	15	Retired assistant principal
Samuel Fuller	M	Black	N/A**	N/A**	Retired principal
Rosa Goodman	F	Black	N/A**	N/A**	Retired principal
Marvin Grant	M	Black	1928	16	Retired principal
Andrew Hall	M	Black	1932	18	College professor
Fred Nix	M	White	N/A**	16	Retired principal
George Turner	M	White	N/A**	N/A**	Retired superintendent
Ronald Walker	M	Black	1927	18	Retired principal
Teachers:					
David Allen	M	Black	1942	1	Banker; left teaching in 1971
Angela Burton	F	Black	N/A	11	Teacher
Lynn Carson	F	Black	1936	9	Teacher
Sally Davidson	F	White	1948	0	Assistant principal
Margaret Finch	F	Black	N/A	9	Banker; left teaching in 1971
Delores Findlay	F	White	1941	7	Museum teacher
Deborah Gaston	F	Black	1941	6	Teacher
Irene Goff	F	Black	1934	12	Teacher
Beth Henderson	F	White	1931	0	Retired teacher
Connie Kent	F	White	1937	6	Guidance counselor
Ted Little	M	White	1940	1	Teacher
Harriet Lowe	F	Black	N/A**	N/A**	Retired teacher
Evelyn Majure	F	Black	N/A**	N/A**	Retired teacher
Susan O'Kelly	F	White	1947	0	Teacher
John Porter	M	Black	1940	8	Assistant superintendent
Charles Schenck	M	White	1944	0	Principal
Lila Smith	F	White	1942	6	Teacher
Jane Stoddard	F	White	1944	3	Retired Teacher
Karen Tate	F	White	1943	5	Teacher
Lawrence Thomas	M	Black	1931	14	Retired assistant principal
Valerie Trent	F	Black	1934	14	Retired assistant principal
Students:					
Leonard Anderson	M	White	1957	N/A	N/A
Julie Blassingame	F	Black	1955	N/A	N/A
Bob Hartford	M	Black	1957	N/A	N/A
Kevin Johnson	M	White	1957	N/A	N/A
Tim Martin	M	Black	1956	N/A	N/A
Katherine Paley	F	Black	1956	N/A	N/A
Meg Stevens	F	White	1956	N/A	N/A
Duke Tanner	M	White	1958	N/A	N/A
Lisa Tappan	F	White	1958	N/A	N/A
Steve Victor	M	Black	N/A**	N/A	N/A
Janice Yelverton	F	White	1957	N/A	N/A

* All names are pseudonyms.
** Archival interviews; personal data not available.

Table 2
Interview Protocol - Administrators and Teachers

1. Describe your first teaching position - where you taught, what you taught, what your school was like, such as school climate, quality of teaching, school facilities, materials available, etc.

2. What do you remember about the beginnings of desegregation in Muscogee County schools? What did teachers and students think about desegregation when it started? What was the reaction of parents and community members?

3. In the massive teacher and student reassignments beginning in 1971, what happened to you? What happened to your colleagues? How did you feel during this process?

4. When schools were desegregated on a 70% white/30% black basis in September 1971, how did people react? Do you remember conflicts between students? Between teachers? Between teachers and students? Between administrators and students or teachers? Do you remember special efforts at school or in the community at trying to get along with each other?

5. Do you remember any staff development workshops or meetings to help teachers deal with the problems of desegregation? Was there any effort made to help white teachers more effectively teach black children and black teachers more effectively teach white children? If so, who sponsored them? Were they helpful? Did teachers within the school work with each other to promote student learning? If so, how?

6. Were there any efforts within schools to promote interracial communication among students or among teachers? If so, were they effective?

7. In 1971, 1972, and 1973, every high school had a ratio of at least 60% white and 40% black. Yet by 1976, once again there were some virtually all-white schools and all-black schools. How did this happen?

8. Do you believe now, looking back over the past twenty-five years, that desegregation of schools has been successful in Columbus? Why or why not?

9. Do you believe the benefits of school desegregation have been worth the costs? Did we gain more than we lost by desegregating schools?

Table 3
Interview Protocol - Students

1. Describe the high school(s) you attended - what your school was like, such as the kinds of students who went there, school climate, quality of teaching, school facilities, materials available, extracurricular activities, etc.

2. What do you remember about the beginnings of desegregation in Muscogee County schools (freedom of choice)? What did teachers and students think about desegregation when it started? What was the reaction of parents and community members?

3. In the massive teacher and student reassignments beginning in 1971, what happened to you? What happened to your friends? How did you feel during this process?

4. When schools were desegregated on a 70% white/30% black basis in September 1971, how did people react? Do you remember conflicts between students? Between teachers? Between teachers and students? Between administrators and students or teachers? Were disciplinary matters handled effectively and fairly?

5. Do you remember any workshops or meetings or organizations to help students deal with the problems of desegregation? If so, who sponsored them? Were they helpful? Were there any efforts within schools to promote interracial communication among students or among students and teachers? If so, were they effective?

6. Did quality instruction and learning experiences take place in your classes? Were your classes effectively integrated? Were extracurricular activities open to all students on an equal basis?

7. In 1971, 1972, and 1973, every high school had a ratio of at least 60% white and 40% black. Yet by 1976, once again there were some virtually all-white schools and all-black schools. How do you think this happened?

8. Do you believe now, looking back over the past twenty-five years, that desegregation of schools has been successful in Columbus? Why or why not?

9. Do you believe the benefits of school desegregation have been worth the costs? Did we gain more than we lost by desegregating schools?

The interviews were crucial to understanding the process and outcomes of school desegregation because they provided information from the "inside" not otherwise available. Obviously, memories of a time nearly 30 years distant were selective, but that very selectivity can be instructive, particularly in comparing the interpretations of Black and White participants recalling those events that changed their lives. As "draftees" into newly desegregated settings, teachers had little control, and the concern for White superiority negated what efficacy many, especially Black teachers, might have retained in equitably organized schools and classrooms.

Freedom of Choice Desegregation

The School Board governing Columbus' Muscogee County School District had desegregated grades K-12 by 1967-1968 without judicial coercion. When the Fifth Circuit Court of Appeals ruled in *U.S. v. Jefferson County Board of Education* in 1967 that teaching staffs had to be desegregated as well, the MCSD Board responded with a plan to employ teachers where "race or color shall not be a factor" and to assign two or more teachers to schools where they would be in a racial minority (Muscogee County Board of Education, Minutes, 1968, p. 4546. Hereafter cited as MCSD Board Minutes). Fifty-nine of Columbus' 67 schools opened in September 1968 with one or more minority race teachers. The Board assigned 42 White teachers to formerly Black schools, and 82 Black teachers to formerly White schools (*Lockett*, 1968, August 8, pp. 6-8).

From the start, the needs of White schools, teachers, and students took precedence, as the following examples indicate. Both Black and White respondents in Columbus believed the first Black teachers to desegregate White schools were hand-picked veterans, a practice documented across the South (Buxton, Pritchard, Bingham, Jackson, & Talps, 1974, p. 20). Prior to the initial transfers, a White and a Black principal "shared information" about their teachers. Fred Nix[3], the White principal, recalled getting a list of Black teachers from which to choose: "I wouldn't say volunteers, but [Black teachers who] would do what was best for education in Muscogee County." What was "best for education" meant what most benefitted the White schools. "We had some first-class [Black] teachers," he remembered. A White teacher, Charles Schenck, similarly recollected two strong Black teachers came to his all-White school. He said the school system tried "to select your stronger, your more acceptable Black teachers to be assigned to [White schools]." Irene Goff, a veteran Black teacher, noted that the first Black transfers were Black principals' wives "used, say, as tokens to get the feel [of integration]." These women had advanced degrees, and, because of their husbands' positions within the system, were not likely to protest their appointments.

Black principals did not have the opportunity to choose from a list of White teachers. Usually, first-year White teachers were hired in the two Black high schools. Veteran White teachers were not so willing to do what was "best" for Black schools; few volunteered to go there. Some Black teachers resented the racial differences in employment patterns. Lynn Carson, an experienced Black teacher, noted, "Every time a vacancy came up in a Black school, they were replacing that Black teacher with a non-Black teacher." Most of the White high schools employed only the required two Black teachers before 1971, but by 1970, one Black school had 11 White teachers and the other had 14 (*Lockett*, Individual school reports," 1968-1970).

Issues of equity and trust immediately surfaced. Together for the first time as peers, White and Black teachers discovered that even their paychecks came in different colors. The perception among Black teachers was that Whites were probably being paid more, though the district supposedly equalized pay for Black and White teachers in 1946, and the average salary for Black teachers was slightly higher than average White pay due to seniority (*Lockett*, 1964).

Freedom of choice by 1970 had led to only 12.5% desegregation for Columbus students, with three all-Black and 19 all-White schools remaining out of 67 total (*Lockett*, 1970). Federal pressure had increased with Supreme Court rulings in *Green v. New Kent County* (1968) and *Alexander v. Holmes* (1969). Shortly after the latter, the federal Office for Civil Rights mailed letters to 112 Southern districts, including Muscogee County, telling them to desegregate by December 31, 1969, or face losing federal funds (Metcalf, 1983, p. 53), no small threat to Columbus schools which got almost $1.8 million that year (MCSD Board Minutes, 1969).

Superintendent William Henry Shaw and the School Board wanted to "keep the initiative." Two nearby counties recently had been placed under court-ordered plans, and the Board wanted "to forestall the possibility of having the federal courts take over operation of the schools." On January 5, 1970, it proposed adoption of a unitary system immediately by transferring teachers to create a racial ratio of 75 percent White and 25 percent Black on every faculty (Johnson, January 5, 1970; Johnson, January 6, 1970; MCSD Board Minutes, 1970, pp. 4990-4991). The switch would take effect at the end of the semester, February 2, when 110 White teachers and

[3] All names used are pseudonyms.

120 Black teachers, about 17 percent of the teaching force, would change schools. Transfers would be based on seniority, and teachers were asked to give two choices for their school assignment (MCSD Board Minutes, 1970, p. 4996).

The community erupted. Opponents of the plan jammed the January School Board meeting (MCSD Board Minutes, 1970, p. 5005). Two hundred and fifty students staged a walk-out at a working-class White high school on January 16. They were suspended, but reinstated after Georgia's segregationist governor Lester Maddox supported student protests in Columbus (Battle, 1970; "Appeal Due in Bootle's Ruling," 1970). On January 28, two White parents presented a Muscogee Superior Court judge with 170 signatures on a petition in a successful attempt to enjoin the Board and stop teacher transfers (Johnson, January 28, 1970). Governor Maddox inflamed passions further on February 1 with a speech to 400 people from the steps of the Muscogee County courthouse. He declared "Freedom of Choice Day" in Georgia and urged parents to "put children above [federal] dollars" and to resist any forced transfers (Johnson, February 1, 1970). The Georgia General Assembly then passed a bill making pupil and teacher transfers based on race illegal. When the Superior Court judge dismissed the White parents' suit as moot, the School Board stated that it would comply with state law and would not transfer teachers "unless it is hereafter ordered to do so by a court of competent jurisdiction" (Johnson, February 25, 1970; MCSD Board Minutes, 1970, p. 5049).

Forced Transfers of Teachers

The 1970-71 school year was tense. A court order appeared imminent, especially after *Swan v. Charlotte-Mecklenberg Board of Education* in April 1971 upheld forced busing and using racial ratios to achieve desegregation (Metcalf, 1983, pp. 101-104). The Fifth Circuit Court on May 29, 1971, ordered Muscogee County to dismantle its dual system and comply with *Swan* (P. Smith, May 29, 1971). In June, the Board approved a pupil assignment plan with 70 percent White and 30 percent Black students in each school. Busing would be implemented as needed. The faculty ratio would be 75 percent White/25 percent Black in every school, and the Board promised that "the oldest teachers in point of service in each school shall be given their choice of remaining in such school or being assigned to a different school" (MCSD Board Minutes, 1971, p. 5464; P. Smith, June 24, 1971).

Teachers were anxious. Though officially notified by a letter in their July paychecks (Laney, 1971), teachers usually met with principals before the school year ended to learn their assignments. One White school simply posted a list by lowest seniority and made assignments from the top. Ted Little, a second-year teacher, remembered, "I was real scared. I didn't want to go to [the Black school]." Just before his name was reached on the list, "they fulfilled their quota. . . . I didn't have to go." Another young White teacher, Charles Schenck, joked with his principal he was being "exiled" to a Black school. One young White female teacher, Jane Stoddard, resented men counting military service toward their seniority, but she was afraid to protest. She said, "You didn't want to bring attention to yourself--anything that might be thought to be negative. . . . [Might] move your name to the top of the list." She said that if she had been transferred to a Black school, she would have resigned. Another White female, Delores Findlay, recalled the helplessness she felt, wishing teachers had more input into the process: ". . . [A]ll that was done by principals and administrators, and the teachers just had to sit and wait and see [what happened]." She felt teachers would have helped because it was "fairly chaotic as far as how they did it. . . ." Angela Burton, a veteran Black teacher, echoed that frustration. She selected her two choices of White schools, but her principal told her that she had been assigned to another school. "I started screaming, . . . 'That's not my choice.' . . . and I went on and on and on . . . . We [teachers] stood out in the hall and cried."

Black teachers and many Whites were convinced the Black teachers' transfers were not based on seniority. They thought the White schools simply took the most talented and experienced Black teachers. Some had been department chairs with 15 or 20 years' experience, yet were transferred to White schools. Despite these perceptions, most transfers apparently *were* based on seniority. The tradition of long service among Black teachers made careers of 40 years the norm. Lynn Carson called a colleague ending her career at 30 years "quite early retired." Since the ratio for Black teachers in all schools was 30 percent of the total faculty, the Black high schools had to give up 70 percent of their teachers. Each retained only 17 teachers from faculties of more than 50 (*Lockett*, Individual school reports, 1971). Teachers with fewer than 20 years' experience got transferred. There is evidence, however, that a couple of elite White schools picked the leaders among those Black teachers transferred. This pattern prevailed across the South as the White schools got the most competent Black teachers (Anderson, 1971; Foster, 1990). At one Black Columbus high school, every faculty member with an advanced degree was sent to a White school (MCSD Board Minutes, 1971, pp. 5483-5501A; MCSD Board Minutes,

1972, following p. 5865). This left the most elderly and those with the least training at the Black school.

White teacher transfers conformed to the seniority standard, with the youngest in service going to the traditionally Black schools (MCSD Board Minutes, 1971, pp. 5483-5501A; MCSD Board Minutes, 1972, following p. 5865). Some teachers believed that placing young White teachers in Black schools was as intentional as assigning Black veterans to White schools. Young teachers might have been more flexible and perhaps would have worked together more. Control may also have been a motive. Sally Davidson, a young White teacher, noted, "[We] were desperate for jobs, so [we] were much better at just shutting up [and] doing what [we] were supposed to do." A Black teacher, Lawrence Thomas, agreed: "[As a first year teacher,] you don't have any leverage and you're a neophyte. . . . "

Some Black teachers were anxious and bitter at their transfers. Angela Burton said she "felt totally violated." But most Black teachers faced their transfers more confidently than Whites. Several of the men had experience in a desegregated military, and many of the women had worked for or with Whites. Some may also have been more committed to achieving desegregation. As Lynn Carson put it, "If we have integration, we have to suffer the consequences, whatever they are." She had participated in civil rights demonstrations as a student and believed the change would come "on the backs of all of us. So I might as well pay my dues as anybody else."

The culture of segregation intensified teachers' anxiety. A White teacher, Lila Smith, grew up within three blocks of her Black school assignment, yet never knew it was there. Lynn Carson had the same experience; she had never even driven by her assigned White school only a mile from her house. Angela Burton, bolder than many of her colleagues, decided to visit her assigned working-class White school beforehand. It had a tough reputation. She recalled, "We had heard all kinds of things like . . . they're rednecks, they're hateful, they're mean, they don't like Blacks. . . ." When she arrived, she discovered night school was in session. "So when I went in, of course, I met the [night school principal] that was not so pleased with my coming in. And he made that known, that I should not be there." She left, dreading her assignment.

Teachers' Reactions to Forced Desegregation: Survival

Several studies during the desegregation era identified positive teacher attitudes as the key to "success" in desegregation. Kritek (1979, p. 19) found that "teachers . . . bear the primary responsibility for making sure that integration 'works.'" Their behavior, according to a 1974 Rand Corporation report, was the most critical factor in a desegregated school (Dickinson & Freeland, 1981, p. 40). Racial prejudice, failure to serve the curricular needs of diverse students, and disregard of the insecurities arising from desegregation contributed to a climate that wasn't conducive to learning (Robinson, Robinson, & Bickel, 1980). Where teacher attitudes consistently supported desegregation, student racial attitudes were more positive, Black students felt more accepted, teacher expectations were higher for students in the minority, and Black academic achievement was higher (Crain, Mahard, & Narot, 1982). In Columbus, however, teachers weren't trained to deal effectively with the social revolution taking place in their midst, though the literature of the period clearly advocated preparing teachers to work across racial lines, especially training them in human relations, multicultural curricular approaches, democratic problem-solving approaches, and pedagogical approaches (Crain, Mahard, & Narot, 1982; Johnson & Hall, 1968; National Education Association, 1980; Wey, 1965). Districts where leaders used available federal funds for inservice training found less resistance to desegregation (Dickinson & Freeland, 1981; Rodgers & Bulloch, 1974).

But in Muscogee County, it was "business as usual." Rosa Goodman, a Black principal, remembered that Superintendent Shaw told principals that if they couldn't communicate effectively with persons of another race, then something was wrong with their training. No respondent remembered workshops or inservice activities the summer before forced teacher and student transfers nor during the planning week just before school opened. Delores Findlay noted, ". . . [T]he teachers knew that we needed help, we needed inservice, but there wasn't time. It's sink or swim. . . . " Another White teacher, Karen Tate, agreed: the expectation was "do the best you can." John Porter, a Black teacher and coach, remembered ". . . [It was] going to be school as usual. That we should not overreact to the fact that this is the first day of totally integrated school." No administrators made special efforts to prepare teachers: ". . . [W]e were just left to handle it as mature individuals." One Black teacher, Valerie Trent, was more cynical: she perceived the attitude of the leadership of the district as "Let's get it over with before they cut off our federal funds. . . ." She stated that "the Black teachers and the White teachers needed training on the learning styles of children because the methods and the things that they were using sometimes did not work. . . ." But the attitude of school district leaders was to maintain "business as usual" in a setting as *un*usual as any the teachers had experienced.

Across the South many teachers in the first year of desegregation suffered "culture shock" at being thrust into a school culture different from their own. Symptoms included depression, self-doubt, hostility, anger, seclusion, helplessness, and a longing for the company of

one's own culture (Fuchs, 1969, p. 21). More White teachers seemed to experience culture shock, perhaps because Black teachers historically had to cope in a predominantly White world and make accommodations (Buxton et al., 1974; Kron & Faber, 1973). Teacher performance was adversely affected. Students suffered if their teacher was anxious, frustrated, or afraid. Jane Stoddard, a young White teacher, described her culture shock when the student composition at her college prep school changed: "It was almost like joining the Peace Corps. Going to a place you've never been, living with people you've never dealt with, not understanding their culture, where they were coming from. It was just [that] the Peace Corps trains you for it." Several White teachers shared the goal of simply "surviving," enduring each school day and the year. Ted Little, who remained at a White school that experienced significant conflict, recalled doing nothing but staying in his classroom and going home as soon as he could. He wanted nothing to hold him at school longer than he had to stay. He felt "like a turtle closed up in his shell. Just an old turtle barely poking his head out." After the first year of desegregation, Sally Davidson remembered that many White teachers at a formerly all-Black high school were overjoyed that they could transfer out.

Racial Relations and the Treatment of Black Teachers in Desegregated Schools

Relationships between Black and White teachers were often distant or hostile, and that climate negatively affected the efficacy of both. The White teachers were "not over friendly," remembered Deborah Gaston, an experienced Black teacher. "They would speak to you" but made no move to ease the Black teachers' adjustment to their school. Black teachers thought that some White teachers dressed unprofessionally. David Allen and Lawrence Thomas pointed out that men didn't wear ties; the women wore pants--unlike the dress at the former Black schools. Valerie Trent remembered White teachers' insensitive remarks: ". . . There was one math teacher, we finally taught her not to tell us about her maid every day." Black and White teachers essentially remained segregated within many schools in the first years of desegregation. Cafeterias and faculty lounges rarely desegregated. As Valerie Trent put it, ". . . [W]e didn't trust them, and they didn't trust us. . . ." Irene Goff felt that members of her department resented her, and she recalled with bitterness that if she met White colleagues in public, "they wouldn't speak to you. . . . [T]hey act like you are poison. Didn't say one word. . . . Turn their head, if anything. Didn't want anybody to know they're teaching in the same place that you are." Angela Burton found little closeness among the teachers at her White school, but "we changed the attitudes of the faculty there. . . . We became friends . . ., so I figure we [Black teachers] improved the whole element."

Teachers had more positive memories of faculty relationships within the former Black schools. Sally Davidson found mentors among the Black teachers within her department. Two White teachers, Lila Smith and Susan O'Kelly, saw the few remaining veteran Black teachers at their school as a stabilizing influence, often providing information and support for the inexperienced White faculty. John Porter recalled White teachers and Black coaches working together to motivate student behavior and achievement.

Many veteran Black teachers received inequitable treatment in White schools. They were treated as if they were first-year teachers with no status, typical of newly desegregated schools in the South (Deever, 1992; Foster, 1990; Gaston, 1971). Often they had no permanent room and "floated" among White teachers' classrooms. Some of the White teachers obviously resented this arrangement. Deborah Gaston said White teachers got angry if she erased anything on the chalkboard. She was afraid to sit at the White teachers' desks and had nowhere to store her materials. One White teacher refused even to leave the classroom during her lessons. Gaston suspected this was a way to monitor her teaching ability.

Black teachers believed that White teachers did not value their ideas and contributions. Lynn Carson, a veteran Black teacher, noted, "All of us from the total Black schools, it was not like we brought any ideas to the table. It was like, 'We want to show you how to do it. . . .' And you came in like, 'I've been doing this for ten years. What are you going to show me how to do? . . . Let *me* show *you* how to do it.'" Black teachers felt that when they did share their ideas, White teachers appropriated them. Irene Goff remembered, "They're still your ideas, [but White teachers] capitalize on them and pretend they started it. . . . I have to keep my mouth closed then, see."

Some Black teachers found course load assignments inequitable. Goff, an English teacher, said her department chair assigned her the "worst classes," but backed down when the Goff told her, "Just because I'm new, you don't have to give me the worst you have." Assertive Angela Burton had less success. The White principal assigned every new course to her. She finally asked him why her colleagues couldn't share the load of course development, since "we're all supposed to teach everything that's offered in [our] department. . . . He looked at me and he said, 'I think you can handle it.' I said, 'Don't do me no

small favors.'. . . So I left, knowing I had to take that class because he wasn't changing."

Even in the former all-Black schools, Black teachers sometimes lost authority. Lawrence Thomas was the head football coach and athletic director, but he remembered that after desegregation he no longer could choose his coaching staff. A "lackadaisical" and "uncooperative" White coach was transferred into his program. Tension grew until the White coach ordered track shoes in his old school's colors. Mr. Thomas sent the shoes back and took over the White coach's budget and team. The White coach bypassed the school's Black principal and appealed directly to the Superintendent, but Dr. Shaw ultimately backed the athletic director, who fired the White coach.

Despite in-school segregation and conflict, some teachers forged interracial collaborations. Margaret Finch, a science teacher, said she and a White colleague teamed up "to see how we could be competitive, how we could reach the students, and how we could do things together." She said the male faculty on her hall also were very supportive and protective of her. Two White teachers, Sally Davidson and Karen Tate, also recalled collaborating with Black colleagues. Angela Burton developed personal friendships in her department across racial lines: ". . . I would go and visit . . . their homes. They would come visit me when I got sick. . . ." Another Black teacher, David Allen, described his cordial relationship with two White teachers, though it had limits: ". . . [W]e'd chitter-chat about things and talk about folks and stuff like that but we never really talked about the race issue."

The climate in most Columbus schools improved considerably the second year. Valerie Trent, a Black veteran teacher, said Whites began to see "that every time you saw somebody Black, he wasn't going to hit you or steal your pocketbook. . . ." She thought both sides had been taught Whites were superior and Blacks submissive, and it took time to get beyond that. Black respondents often cited dispelling stereotypes as one benefit of desegregation, as have teachers and researchers in other desegregated settings (Metz, 1994; Nicholson, 1976). ". . . [Whites] were always given the lower concept about Blacks," said Irene Goff. "It was always told that Blacks would steal and fight and [were] greedy, dirty. [They learned] all these things were lies." A White teacher, Karen Tate, saw the growth of broader professional understanding: "[Now] you don't fear the child that comes into your room." Students and teachers came to see each other as "people" and not mere representations of race. Black respondents especially thought it important that Whites and Blacks learn about diverse people in a peer setting in order to succeed in an increasingly multicultural world.

Loss of Leadership and School Ownership

Despite the improving climate, many White and Black teachers resigned or took early retirement between 1971 and 1975. The Muscogee County School District's rate of teacher resignations and retirements increased about 25 percent for the period 1970-1975, compared to the previous four years, with an average loss of about 260 teachers per year (MCSD Board Minutes, 1966-1975). This came at a time when stability and experience would have been beneficial. Black teachers insisted the number and percentage of Black teachers in the school district declined, a trend evident across the South in the 1970s (Bosma, 1977; Busby and Barrett, 1988; Commission on Civil Rights, 1972; Deever, 1992; Hudson & Holmes, 1994; Metcalf, 1983). But Muscogee County employment records indicated almost 27 percent Black teachers in Columbus high schools in 1968 and just under 29 percent in 1975 (*Lockett*, 1970; *Lockett*, Individual school reports," 1968, 1971-1975).

Overall percentages may have held firm, but Black schools lost veteran teachers' leadership and the strong educational traditions they had fostered, as happened in former Black schools everywhere during desegregation (Bell, 1983; Nicholson, 1976). Black respondents believed that the best Black teachers were parceled out to the White schools where they were underutilized and unappreciated. In return, Black schools got inexperienced White teachers unresponsive to Black children. David Allen pointed out that a White teacher might be fearful of Black children, "she didn't want to touch them because she thought they were dirty. . . ." Black children lost that "loving, nurturing touch" provided by Black teachers. Similarly, Foster (1990, p. 136) found that desegregation removed the Black teacher's ability to be an "admonisher, urger, and meddler" for Black students. In desegregated settings, Black teachers lamented their diminishing ability to positively influence the Black students.

Blacks and Whites both recognized that school desegregation equalized resources and opportunities. For the first time, former Black schools gained new textbooks and athletic and instructional equipment and not the usual castoffs from the White schools. Black teachers marveled at the difference in resources. Lynn Carson remembered, ". . . I was just standing there with my eyes open, my mouth open, and couldn't speak. . . . Listen, it was just like a kid at Christmas. . . . I had lived all these years and had just never dreamed of all these kind of

material possessions in a classroom." But the Black high schools, focal points of the Black community, enjoyed more parental and neighborhood support before desegregation. That diminished, both in Columbus and across the South, when the student and teacher composition suddenly became 70% White.

Professional Crisis/Professional Growth

Many White teachers found that their training and experience had not prepared them for the unfamiliar setting. Orfield (1975, p. 318) noted that some teachers after desegregation faced a "professional crisis," realizing they were ineffective with diverse children: "[W]hen these 'good' teachers are confronted with large numbers of students who do not respond in the school's 'typical' way, they must either blame the newcomers or recognize and acknowledge a very serious professional inadequacy." Teachers competent in racially homogeneous classrooms often experienced discomfort and self-doubt in diverse settings. A survey of 215 Southern teachers in 1974 revealed that more than a third of White teachers acknowledged becoming "more conscious" of their "qualifications as a teacher" after integration (Buxton et al., 1974, p. 20). "[I] just didn't know how to reach them or to help them," said Delores Findlay, an experienced White teacher in Columbus. "And so there was a frustration with that grew each year until in 1974 I had an opportunity to go to a private school to teach where I could teach in the way that I felt comfortable. . . ." Some teachers simply couldn't deal with the changes in student composition. Generational differences seemed to be a key in the ease of teacher adjustment, especially among White teachers. When an older White male teacher retired after the first year of desegregation, Jane Stoddard, a younger White colleague, remembered that in his farewell speech the retiree said he couldn't stand the "noise" and was "beating around the bush" but "he just kept going on and on, and everybody knew what he was talking about." He meant he didn't want to teach Black children.

But for many teachers, desegregation fueled individual professional and personal growth. Black teachers and principals got more equitable treatment from the school district leadership and collaborative opportunities that previously would have been impossible. Career advancement opportunities especially opened for Black women, perhaps because they were seen as less threatening than Black men by the White district leadership. The mixed setting facilitated cross-cultural understandings, particularly for White teachers. Ted Little saw himself as a better person after desegregation because "I never would have known a lot of great and wonderful Black people if they hadn't done it. . . . . And I had some Black students I got real close to." Jane Stoddard, a social studies teacher, became more culturally sensitive. A couple of years into desegregation, she put up posters of Black people in her room, then realized "every one of them was very light-colored. And I thought, 'This isn't right. . . .' So I quit using them. Because it was almost like saying [to] the [students] that were a darker color, 'You don't fit in here.'"

Another White teacher, Susan O'Kelly, stated that desegregation led to her personal and professional growth: "I think it would have been very easy for me to be one of those little girls back in the White suburbs . . . which was essentially my upbringing. . . . It put me in a situation where I had to [adjust] and it was OK. So, you know, it worked." Karen Tate became more aware of inequity and more activist. She felt that if she had gone back in time, "instead of keeping my mouth shut because I didn't want to offend, I think I probably would have been a flaming liberal. . . ."

Curriculum and Pedagogy: "Business as Usual"

Teachers and students remembered little emphasis on academics during the first tumultuous years of desegregation. Rather, it seemed that "survival" was the goal. Charles Schenck, a White teacher in a former Black school recalled, ". . .[O]ur task was to get through the year and have reasonable order in the schools. Our task was not to look at curricular issues. Our task was to survive the year." Many students agreed. White student Kevin Johnson said, "[Desegregation] taught you more about survival than it did about school." Students recalled "minimal" learning, much disruption in class, so little teacher control that students could "do whatever you wanted to." In taking this approach, many teachers abdicated what efficacy they may have had in their classrooms and schools. Nearly three decades later, many White respondents cited a decline in academic standards as a result of desegregation, though several studies indicated White academic achievement in Southern schools during the first years of desegregation remained steady or rose slightly and that Blacks posted achievement gains (Mahard & Crain, 1983; Rodgers & Bullock, 1974, 1976). Fred Nix, a Columbus principal, lamented, "It was a lost educational time, in my opinion, for many students." Teachers also perceived a decline in the quality of education. Karen Tate explained that the original intent was to equalize education, "but we equalized them on the lower end instead of the higher end of education." The system's attempt to homogenize schools was a mistake, said Charles Schenck. The school district's goal

in the early 1970s was to offer the same curriculum in the same way at every school. Schenck argued, "I think if we make all schools the same, we would tend to make all schools move toward mediocrity."

Desegregation did not prompt major curriculum revisions in Columbus, even though "integrated" committees supposedly rewrote the K-12 curriculum in 1966 (*Lockett*, 1967, p. 51). A 1982 study of 200 Southern high schools indicated an inclusive curriculum had a positive effect on Black student attitudes (Crain, Mahard, & Narot, 1982, p. 141). The Muscogee County curriculum, however, remained essentially Eurocentric. Black teachers who had incorporated diverse content at formerly Black schools found few such curricular materials at the White schools. Those who added Black history or Black literature to their courses often met resistance. When Angela Burton, a Black English teacher, gave out a reading list of Black authors, a White student complained to the principal, who called her in for a conference. Another Black English teacher required a paper on a Black author one grading period. Her son, Bob Hartford, recalled that a White student took an F rather than do the paper, then spray-painted a cross on her door. White teachers resisted curricular change, too. Hartford pointed out, ". . . So many of the [White] instructors . . . did not want to teach Black history at all. . . ." He learned about Black literature only in a Black teacher's class. Students in the formerly White schools had no memories of Black history or literature. Asked if he remembered observing Black History Week, one White student, Leonard Anderson, replied, "It didn't exist." It had existed for decades, just not in Columbus' White high schools.

Fear of conflict steered both Black and White teachers away from discussions of race. Janice Yelverton, a White student, remembered "bantering" between "hard-core" White students when racial issues came up in class. Deborah Gaston, a Black social studies teacher, explained that she did not facilitate racial discourse but rather "went on with the business of teaching." Another Black teacher, John Porter, also tried to make race invisible: "Everyone expected that you were going to conduct yourselves as a good teacher or as a good student as if race didn't matter." A White teacher, Karen Tate, had the same approach: "You just ignore it and maybe everything will be all right." If race came up in class, Ted Little, a White science teacher, tried "to pacify, you know, tell them better days were coming." David Allen, a Black teacher who taught at the high school that had the most violence, thought that making race invisible was an ineffective approach. He explained, "[The attitude] was that we've got this thing worked out and we don't need to dwell on it. . . . Just pretend [racial conflict is] not there, or if it's there, we're going to resolve it before it gets out of hand. And that didn't work." Metz (1994, p. 72) linked this "color-blind" view to the maintenance of traditional racial relationships: "The taboo on the discussion of race, while it may seem designed to avoid hurt feelings or social awkwardness, serves to reinforce White power by stifling discussions of different perspectives."

Just as there were few curricular changes, teachers also held fast to their pedagogical approaches, even when faced with a very different class composition. Orfield (1975) found this to be the typical reaction of teachers across the South. Many Columbus teachers, Black and White, were determined to be "color-blind," as Black teacher Angela Burton noted: ". . . I made up my mind there were going to be students that were going to be neither Black nor White." Beth Henderson, a White teacher, echoed the sentiment: ". . . I was teaching to students. Not Black students nor White students. I was just teaching to students." With a strong White majority, however, being "blind" to race placed an institutional emphasis on the needs of White students (Busby & Barnett, 1988; Fancher, 1971; Foster, 1990; Hudson & Holmes, 1994; Lowe & Kantor, 1989; Nicholson, 1976). Teachers plunged ahead with "business as usual," but as Henderson admitted, "We essentially kept doing what we'd always been doing. . . . It didn't always work." In fact, as Serow and Solomon found in 1979, an impersonal business-like approach negatively affected intergroup relations and stifled student interactions.

Yet teachers and students perceived differences in the ways Whites and Blacks taught and learned. Delores Findlay, a White teacher at a former college-preparatory White high school, found she couldn't effectively teach the new mix of students with her traditional lecture, question-and-answer mode of instruction. Rather than adapt, she resigned. Black teachers observed that their White counterparts rarely facilitated discussion, relying heavily on worksheets and questions from the textbook. Black students alleged that one reason for White teachers' emphasis on quiet individual seat-work was control of behavior, an approach typical of other desegregated schools (Kritek, 1979, p. 22). Black teachers and students maintained that many White teachers feared Black children and used passive learning to avoid conflict. A Black student, Julie Blassingame, asserted that showed a lack of interest in Black students' learning: "[White teachers] just tolerated us to get us through the door, excuse me, and out."

Many Black teachers used more active learning to engage students, though behavior control may also have been a motive. David Allen admitted using experiments and student demonstrations in his science classroom to keep students busy in order to reduce the tension between himself and White students. In former Black schools,

teachers had used more individualized instruction because they taught so many academic levels in one classroom. Many continued that practice in their new settings, as did Black teachers in other Southern schools (Schofield & Sagar, 1983). Margaret Finch explained, "I did not try and march 30 students in the same row." Impressed with a Black math teacher's individualized instruction, Jane Stoddard, a White teacher, observed, "She would have five different preparations [lessons] going in her class at one time, you know. She was amazing."

Perceptions About Students and Teachers

Teachers of both races brought preconceptions of White privilege into their newly desegregated classrooms. Many Black teachers believed that White students were superior. Lawrence Thomas, a Black English teacher, found White students better-read and "able to live vicariously through characters [in literature]." He stated because their parents subscribed to newspapers, most White students could discuss current events. At the same time, Black teachers said White students took advantage of Black teachers' perceptions that they were "far, far advanced" and did less work, yet got better grades than Black students. Black students struggled to get good grades from White teachers. Steve Victor, a Black student in a former Whiter college prep school, recalled that in the first grading period his White English teacher gave him a D, but after she saw that he was a good student, he earned As and Bs the rest of the year. White teachers usually had lower expectations for Black students due to what Lawrence Thomas characterized as "years and years and years of programming." They would talk down to Black students, oversimplifying their vocabularies. A 1979 Texas study supported this perception. It found White teachers asked White students more "product" questions and directed lower-level questions to Black students (Barnes, 1979, p. 25). When White teacher Karen Tate complained to her principal that Black students were roaming the halls and popping into her room at odd times, he communicated his low expectations: "If we can teach each of the Black students . . . to stay in the room for an entire period, then we have done well." Capable Black students adjusted, but those who needed additional help got lost in what Black student Bob Hartford called "the free enterprise system" of instruction. The White teacher's attitude was, "There it is. You come and get it if you want it. I don't have time. You all are too old for me to have to push you to the information. There's a library in the school. Here's your textbook. Here's your syllabus. Take notes, bring tape recorders, whatever you want to do. May the best man win."

White teachers soon learned to grade Black students carefully. They couldn't have a lot of Black failures. When a White teacher failed 35 Black students in one term, the principal told him to write a letter about each student, detailing how he individualized instruction to help each succeed. Jane Stoddard, his White colleague, remembered, ". . . [T]o say the least, we got the message. [That teacher] didn't come back the next year." "[W]e were teaching at the same standard [as before desegregation]," recalled White social studies teacher Beth Henderson, "and we couldn't understand why these [Black] kids couldn't meet the criteria. . . ." She described how White teachers devised a "D/F" stratagem to avoid failing Black students: give the student a D the first six weeks, then an F, and alternate through the year so the student passes with the lowest D possible. Some White teachers spoke resentfully of inflating grades and of compromising standards to avoid failing Black students. They believed White students saw Black students producing poor quality work and being passed anyway, so they followed suit. Karen Tate explained, "We watered down education to such a point that a high school diploma didn't mean anything. We were so intent on everybody passing so that you could not be perceived as being prejudiced or discriminatory in any way. . . ." But White teacher Jane Stoddard saw an abdication of professional responsibility in this approach: ". . . The response, rather than let's try something different with these kids, is let's just pass them. That's easier than saying maybe I need to look at how I teach . . . or how the system is organized."

Black teachers, as well as Black students, were thought less capable by White principals, teachers, and students, even though a higher percentage of Black than White teachers in Muscogee County held Masters degrees (*Lockett*, 1964, p. 56). And Black teachers' degrees were often from prestigious institutions, such as Columbia University, New York University, Northwestern University, University of Chicago, and Indiana University because Georgia in the 1950s and 1960s paid to send Black teachers out of state rather than desegregate the University of Georgia. Their White colleagues sometimes discounted their degrees and experience and promoted the idea of Black teacher inferiority. Charles Schenck explained, "Generally speaking, our belief was that the quality of education available at [Black colleges] was not the same as what was available at [White colleges]." Historical patterns of segregation created self-doubt in some Black teachers facing a classroom of White students, as Lynn Carson recalled: ". . . It was verbally said that Black teachers were not qualified to teach White children. . . . You can hear these kinds of

things for so long until you wonder if . . . that's true." Another Black teacher, Margaret Finch, echoed that uncertainty: ". . . All of your life when you've been reminded by people that are non-minority that you are inferior, now you're thrown into the test to prove that you can or you can't teach all children. . . . You really kind of walk there not knowing what you can do. . . ."

White standards of language usage became the primary measure of Black teacher competence. White principal Fred Nix required standard English: "You just couldn't hire [a Black teacher] that didn't speak good English. . . . " Black English teacher Lawrence Thomas thought that White students were unfamiliar with Black speech patterns, particularly intonation and pronunciation, "so the perception could have been that [the teacher] might not really know this, just can't get this over to me. . . ." Black science teacher Margaret Finch put great effort into speaking "standard English," nightly practicing her lectures before her mirror. White students often ridiculed Black teachers' language usage in cruel and insensitive ways "as only teenagers can." Because she had trouble understanding her Black English teacher's "dialect," White student Julie Yelverton "immediately assumed that she was not as educated. . . ." Another White student, Lisa Tappan, said her English teacher didn't speak "standard English" and that other White students had little respect for her: "She was kind of like a joke to be in the English Department."

Challenges and Cultural Conflicts

Black teachers across the South constantly dealt with challenges from White students and their parents during the early 1970s (Foster, 1990). As Black teacher Valerie Trent remembered, "You're having to prove yourself over and over. . . . The kids didn't have any respect for you, the parents didn't have any." White students often were insubordinate: some even called Black teachers "stupid" or "dumb" to their faces. Trent, a science teacher, recalled that a White doctor's son took dated information from an old medical book to challenge something she had taught. She showed him the correct information in the textbook; the student refused to accept it. He responded, "I know that's right. My daddy is a doctor." "I don't care what your daddy is," Trent replied. "If your daddy is still dealing with that [information], I'd certainly hate to go to him because he's going to kill somebody if he's not careful." She recalled, "[That student] was furious with me." Trent and the father resolved the situation over the telephone, and the father later complimented her knowledge of biology. Not all parents were so supportive. Another Black teacher, Deborah Gaston, remembered frequent White parent visits to her classroom during the first year of desegregation. Others told of White parents removing children from their classes and of constant parent-teacher conferences.

White teachers likewise experienced conflict with Black students. Sometimes students were insubordinate, such as in a gym class recalled by White student Duke Tanner, where Black students ignored a White teacher's order to play on mixed teams. Occasionally, White teachers experienced violence at the hands of Black students. White student Kevin Johnson said when his White male teacher answered a knock on his classroom door, a Black student punched him in the face, breaking his glasses and his nose. Though the infrequent attacks on teachers reinforced a White stereotype of Black brutality, most conflict was more subtle. White teacher Karen Tate remembered thinking, "Will they ever learn to behave and act like White folks?" Most schools tried to force Black children to adapt to how education had "always" taken place (Kritek, 1979, p. 22). The cultural gap between White teachers and Black students seemed wider than between Black teachers and White students, possibly because Black people in the South historically were more involved in the White community (Buxton et al., 1974; Kron & Faber, 1973). Some White teachers exacerbated conflict with racist remarks. For example, vocational teacher Angela Burton remembered that her White supervisor told Black students that if they didn't behave, "he was going to bundle them up and send them all back to Africa." Some White teachers resisted correctly pronouncing "Negro," and Black students (and teachers) took offense at the Southern White pronunciation, "Nigra" (Johnson & Hall, 1968).

Cultural ignorance caused many conflicts. Black basketball coach John Porter bemoaned losing his top player because of "overreaction" by a White female teacher. When the Black student entered the teacher's homeroom to visit a friend, the teacher felt intimidated, as did many White teachers in desegregated settings (Kritek, 1979). To ease the White teacher's fears, the Black principal transferred the student to another school. Porter recollected, "I saw a lot of Black kids, especially Black male kids, who perhaps were just being themselves in terms of their culture [and] may have been penalized because some of the White teachers felt a little threatened by their mannerisms or by some of the things they would do in a predominantly Black situation. . . ." Black students across the South were similarly punished in disproportionate numbers for "cultural" behaviors (Eyler, Cook, & Ward, 1983; Kritek, 1979). It took time for White teachers to adjust. Karen Tate recalled her discomfort when a Black male student put his arm around her: "I remember looking down and seeing this Black hand patting my shoulder and I thought 'What in the

world?' . . . I wasn't used to Black students and this hand was patting my shoulder."

White teachers' racial insulation influenced content and pedagogy as well. White social studies teacher Jane Stoddard asked students to draw plans for their "Dream House." She was stunned to see Black students draw shotgun houses. She explained, ". . . It never dawned on me. They had never been in a house that wasn't like that. . . ." Black students, including Julie Blassingame, Katherine Paley, and Bob Hartford, commented on White teachers' fear and anxiety and felt teachers "didn't care" about Black students. Black students remembered a few sympathetic White teachers, but as Hartford noted, "We saw a lot of [White] teachers that wanted to help, but . . . they didn't want to step outside their comfort zone to help us, because they would probably be isolated." Some Black students thrived in a desegregated setting, but Blassingame and Paley characterized many of them as more "like [Whites]." They saw teachers of both races as favorably inclined toward Black students who had straighter hair, lighter skin, were middle-class and "talked real proper, very distinct. . . ." Similarly, White teachers described many successful Black students as coming from elite families and neighborhoods.

Both Black and White teachers felt constrained in disciplining students, fearing parental reprisals and charges of racism. Tim Martin, a Black student who remained at a formerly Black school, said that Black parents didn't want White faculty using corporal punishment, seeing parallels to slave punishment. Likewise, Black teachers felt White parents would not support their disciplining White children, and were careful not to touch White students. Several students noticed that teachers ignored challenges from students of the opposite race. Bob Hartford and White student Lisa Tappan both remembered that if a student started to argue or refused to work, the teacher often just walked away. White student Janice Yelverton remembered White teachers being tougher on Black students, but alleged perhaps "the Blacks instigated more [trouble]" than Whites. She also stated that if Black and White students had a dispute, "the Whites usually [won] out with the adults as mediators. They just seemed to have more credibility."

White teacher Karen Tate tried a creative method of discipline with a "hyper" Black male. She placed his desk facing the front wall of her classroom and arranged her standing maps around him. He could hear the class but not "entertain" other students. Or so she thought. She recalled, ". . . All of a sudden I see the students begin to snicker and to look at each other. . . . I turned around, and [the Black student] has gotten out of his desk and is standing on his head so that the only thing you can see is his head and his shoulders and they are upside down. And I just lost it. I laughed. That's all I could do."

Similarly, Black teacher Angela Burton found an unorthodox, though effective, way to discipline a White student. She saw a White student had written "KKK" on his desk. When he admitted writing it, she instinctively turned the fear of racial reprisal on the student. She asked, "Do you think you'll frighten me? Because I know some people in the Black Panthers! Would you want me to go tell them that [you threatened me]?" The student immediately apologized and insisted he was "just playing." Burton laughed years later as she told the story: ". . . I was so tickled, because I wanted to tell him, I'm scared of [the Black Panthers] myself! . . . [But] I never saw 'KKK' written right on the desk any more. . . ."

Inservice Opportunities

Given the difficulties teachers faced in forced desegregation, it seems that the Muscogee County School District would have tried to provide more support for them. The district received federal funding from the Emergency School Assistance Program (1970-1973) and the Emergency School Aid Act (1973-1975) for inservice training and curriculum development (MCSD Board Minutes 1970-1975; P. Smith, September 11, 1971, p. 6). A few summer programs were offered, but most teachers interviewed recalled no inservice activities to help them adjust to a desegregated setting. At only one high school did the administration try from the outset to confront issues of race, regularly bringing in speakers from nearby Fort Benning to conduct sensitivity training. Both Black and White teachers, however, expressed doubts about the effectiveness of such training. After a riot at a former White working-class high school, local police officers and race relations experts gave a mandatory workshop for faculty. Ted Little, a second-year White teacher, remembered it as very confrontational, with older White teachers "standing up screaming, hollering at [the workshop leaders], and trying to walk out." The veteran White teachers believed that the workshop was trying to "stuff some stuff down our throats," that "they thought we were having to make too many concessions [to Blacks]." Generational differences in attitudes again were clear. Though Little felt the workshop was divisive rather than unifying, he disapproved of the behavior of the senior faculty.

In 1971, the School Board established both an adult and a student Biracial Committee to advise the district on priorities for the federal funds. Those Committees during the first years of desegregation consistently recommended more teacher training in content and pedagogy responsive

to "minority" students (MCSD Board Minutes, 1971, pp. 5571, 5599, 5637). In June 1972, an adult representative of the Biracial Committee reported to the Board that "students are involved in live issues, [but] it is too much business as usual in the classroom." He recommended the teaching of more Black history and literature and advocated a race relations program required for all school personnel. But the Assistant Superintendent for Instruction did not receive the report favorably. He remarked that "successful teachers of both races seemed to experience little difficulty with children of the same or opposite race," a view quite different from the accounts of teachers in high school classrooms at the time. The Board did not require the training (MCSD Board Minutes, 1972, pp. 5951, 5854-5855).

Implications for Educators

Few of the Columbus teachers interviewed remember the years 1968-1975 fondly. Their descriptions evoke images of conflict and anarchy, and many still are bitter. They recall internal and external constraints on their professional efficacy, their powerlessness, their frustration with the changes desegregation brought. Black teachers say their effort and sacrifice may have been in vain because many schools in Columbus resegregated by the early 1980s (*Lockett*, Individual school reports, 1975-1982). Schools in Columbus and across the nation are becoming more and more racially segregated, and the courts have shown little interest in judicial remedies (Houston & Hyatt, 1997; Orfield & Eaton, 1996). But most White respondents judge desegregation in Muscogee County "successful," despite resegregation, citing improved resources for formerly Black schools and greater options and opportunities for Black children. Interestingly, few Whites saw that "success" extending to White students.

Many Blacks and Whites in Columbus and elsewhere argue that mandated racial ratios alone do not lead to improvement in the quality of education (Bivens, 1997; Holmes, 1997; Shepard, 1997). Most of those interviewed, as well as leading scholars who study cultural diversity issues in education, however, see value in children being exposed to people different from themselves, particularly with the growing multiculturalism of the global society and workplace (Banks, 1997; Gay, 1994). But resegregation highlights continuing inequities. Test scores in majority Black schools in Columbus remain the lowest in the district, falling well below the district and national averages, a problem common to school districts across the country (Muscogee County School District, 1997). Many majority Black schools are not well-equipped due to the inability of their lower-income parents to support "booster clubs" that provide library books and computers, athletic and band equipment, and other "extras."

Even as student populations diversify, teacher populations today are growing more White and middle-class, a national trend reflected in Muscogee County (Hodgkinson, 1991; Lockett, Individual school reports, 1990-1997; Protheroe & Barsdate, 1991). The schools in Columbus, even those that have become majority Black, are still controlled by Whites. No Muscogee County school, for example, had a majority of Black teachers after 1970 even though White students were in the minority in the district by 1987 (*Lockett*, Individual school reports, 1971-1997). As Busby and Barnett (1988, p. 58) argue in studying the effect of desegregation across the South, schools remain within the dominant White culture. In this unfortunately familiar context, the desegregation experiences of teachers from three decades ago still resonate.

Perhaps most disturbing to Black parents, White teachers in majority Black schools often have little cultural understanding of their children and hold lower expectations for their achievement (Delpit, 1995; Irvine, 1990). Black veteran teacher Lynn Carson commented that Columbus had all-Black schools before desegregation, and it had all-Black schools soon after desegregation. The crucial difference is before desegregation, Black teachers taught Black students and had more power to influence their lives. That was no longer the case after Black teachers were dispersed and teachers of both races became "color-blind" (Foster, 1990; Lowe & Kantor, 1989).

There are no simple answers to questions of desegregation and resegregation. The remedies of a quarter century ago led to a loss or abdication of teacher efficacy in significant ways. In Columbus, ingrained notions of White power and privilege were, and perhaps remain, the overt and underlying philosophical bases of the desegregation "revolution," and both Black and White teachers were drafted to fight for the cause, a cause whose outcome remains unclear even after thirty years.

References

Anderson, R. E., Jr. (1971). Mobile, Alabama: The essence of survival. In [no editor] *The South and her children: school desegregation 1970-71* (pp. 38-49). Atlanta, GA: Southern Regional Council, Inc.

Appeal due in Bootle's ruling. (1970, January 23). *Columbus Ledger*, pp. 1-2.

Banks, J. A. (1997). *Educating citizens in a multicultural society*. New York: Teacher's College Press.

Barnes, W. J. (1979). Teachers in desegregated high schools in Texas. *Integrated Education, 17*(1-2), 25-26.

Battle, L. (1970, January 17). Jordan quiet after brief walkout. *Columbus Ledger*, p. 9.

Bell, D. (1983). Time for the teachers: Putting educators back into the *Brown* remedy. *Journal of Negro Education, 52*, 290-301.

Bivens, L. (1997, July 15). Education shaping agenda of NAACP at convention. *Detroit News*, p. A-5.

Bosma, B. (1977). The role of teachers in school desegregation. *Integrated Education, 15*(6), 106-111.

Busby, D. A., & Barrett, C. A. (1988). Impact of integrated education on Blacks in the South: A participation crisis. *Equity and Excellence in Education, 23*(4), 57-62.

Buxton, T. H., Pritchard, K. W., Bingham, C. M., Jackson, C. E., & Talps, L. (1974). Black and White teachers and desegregation. *Integrated Education, 12*(1-2), 19-22.

Crain, R. L., Mahard, R. E., & Narot, R. E. (1982). *Making desegregation work: How schools create social climates.* Cambridge, MA: Ballinger Publishing Company.

Deever, B. (1992). Desegregation in a southern school system, 1968-1974: Power/resistance and the discourse of exclusion. *Journal of Education, 174*(3), 66-87.

Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom.* New York: The New Press.

Dickinson, G. E., & Freeland, K. (1981). Student, teacher, and administrator attitudes toward school desegregation, 1964-1981. *Integrated Education, 19*(3-6), 40-43.

Eyler, J., Cook. V. J. & Ward, L. E. (1983). Resegregation: Segregation within desegregated schools. In C. H. Rossell & W. D. Hawley (Eds.), *The consequences of school desegregation* (pp. 126-162). Philadelphia: Temple University Press.

Fancher, B. (1971). Beaufort, South Carolina: The new breed and the old. In [no editor] *The South and her children: School desegregation 1970-71* (pp. 50-55). Atlanta, GA: Southern Regional Council, Inc.

Foster, M. (1990). The politics of race: Through the eyes of African-American teachers. *Journal of Education, 172*(3), 123-141.

Fuchs, E. (1969). *Teachers talk: Views from inside city schools.* Garden City, NY: Doubleday.

Gaston, P. M. (1971) The region in perspective. In [no editor] *The South and her children: School desegregation 1970-71* (pp. 5-20). Atlanta, GA: Southern Regional Council, Inc.

Gay, G. (1994). *At the essence of learning: Multicultural education.* West Lafayette, IN: Kappa Delta Pi.

Hodgkinson, H. (1991). Reform versus reality. *Phi Delta Kappan, 73*(1), 9-16.

Holmes, S. A. (1997, June 23). At N.A.A.C.P., talk of shift on integration. *New York Times*, p. A-1, B-7.

Houston, J., & Hyatt, R. (1997, May 7). School system wins case. *Columbus Ledger-Enquirer*, pp. A-1, A-4.

Hudson, M. J., & Holmes, B. J. (1994). Missing teachers, impaired communities: The unanticipated consequences of *Brown v. Board of Education* on the African American teaching force at the precollegiate level. *Journal of Negro Education, 63*, 388-393.

Irvine, J. J. (1990). *Black students and school failure: Policies, practices, and prescriptions.* Westport, CT: Greenwood Press.

Johnson, C. (1970, January 5). Massive teacher shifts expected. *Columbus Ledger*, p. 1.

Johnson, C. (1970, January 6). Board begins teacher shift. *Columbus Ledger*, p. 1.

Johnson, C. (1970, January 28). Injunction sets stage for battle. *Columbus Ledger*, pp. 1-2.

Johnson, C. (1970, February 1). Maddox calls for resistance. *Sunday Columbus Ledger-Enquirer*, pp. A1, A4.

Johnson, C. (1970, February 25). State law prompts suit dismissal. *Columbus Ledger*, pp. 1-2.

Johnson, H. T., & Hall, M. M. (1968). *School desegregation, educational change, and Georgia.* Athens, GA: Georgia Association for Supervision and Curriculum Development and the School Desegregation Educational Center, University of Georgia.

Kritek, W. J. (1979). Teachers' concerns in a desegregated school in Milwaukee. *Integrated Education, 17*(1-2), 19-24.

Kron, K. N., & Faber, C. F. (1973). How does the teacher cope with the culture shock? *The Clearinghouse, 47*, 506-508.

Laney, M. (1971, August 2). Teacher reassignments well-accepted. *Sunday Columbus Ledger-Enquirer*, p. 13.

Lockett, J. L. et al v. Board of Education of Muscogee County School District. (1964, April 1). Civil Action 991. Middle District of Georgia, Columbus Division.

Lockett, J. L. et al v. Board of Education of Muscogee County School District. (1967, June 15). Civil Action 991. Middle District of Georgia, Columbus Division.

Lockett, J. L. et al v. Board of Education of Muscogee County School District. (1968; 1971-1997). Individual school reports. Civil Action 991. Middle District of Georgia, Columbus Division.

Lockett, J. L. et al v. Board of Education of Muscogee County School District. (1968, August 8). Civil Action 991. Middle District of Georgia, Columbus Division.

Lockett, J. L. et al v. Board of Education of Muscogee County School District. (1970, July 30). Answers of the defendant Board of Education of Muscogee County School District to the plaintiffs' first 1970 interrogatories. Civil Action 991. Middle District of Georgia, Columbus Division.

Lowe, R., & Kantor, H. (1989). Considerations on writing the history of educational reform in the 1960s. *Educational Theory, 39*, 1-9.

Mahard, R. E., & Crain, R. L. (1983). Research on minority achievement in desegregated schools. In C. H. Rossell & W. D. Hawley (Eds.), *The consequences of school desegregation* (pp. 103-125). Philadelphia: Temple University Press.

Metcalf, G. R. (1983). *From Little Rock to Boston: The history of school desegregation.* Westport, CT: Greenwood Press.

Metz, M. H. (1994). Desegregation as necessity and challenge. *Journal of Negro Education, 63*,64-76.

Muscogee County Board of Education. (1968-1975). *Minutes [of the] Muscogee County Board of Education, Muscogee County School District, Columbus, Georgia.* Superintendent's Office, Muscogee County School District Administration Building, Columbus, GA.

Muscogee County School District. (1997). Norm-referenced achievement test data summary, 1989-1997.

National Education Association. (1980). *School desegregation guidelines for local and state education associations.* Washington, D.C.: National Education Association of the United States.

Nicholson, H. (1976). The way I see it: Reflections on school desegregation by teachers from four cities. In F. H. Levinsohn & B. D. Wright (Eds.), *School desegregation: Shadow and substance* (pp. 187-191). Chicago: University of Chicago Press.

Orfield, G. (1975). How to make desegregation work: The adaptation of schools to their newly-integrated student bodies. Law and Contemporary Problems, *39*, 314-340.

Orfield, G., & Eaton, S. E. (1996). *Dismantling desegregation: The quiet reversal of Brown v. Board of Education.* New York: W. W. Norton.

Robinson, S. P., Robinson, A., & Bikel, F. (1980). Desegregation: A bibliographic review of teacher attitudes and Black students. *Journal of Negro Education, 31*, 48-59.

Rodgers, H. R., & Bullock, C. S., III. (1974). School desegregation: Successes and failures. *The Journal of Negro Education, 43*, 139-154.

Rodgers, H. R., & Bullock, C. S., III. (1976). *Coercion to compliance.* Lexington, MA: Lexington Books, D.C. Heath.

Schofield, J. W., & Sagar, A. H. (1983). Desegregation, school practices, and student race relations. In C. H. Rossell & W. D. Hawley (Eds.), *The consequences of school desegregation.* (pp. 58-102). Philadelphia: Temple University Press.

Serow, R. C., & Solomon, D. (1979). Classroom climates and students' intergroup behavior. *Journal of Educational Psychology, 71*, 669-676.

Shepard, S. (1997, July 11). NAACP at a crossroads: Group struggles with future direction. *Atlanta Constitution*, p. A-11.

Smith, P. (May 29, 1971). Muscogee ordered to integrate fully. *Columbus Enquirer*, pp. 1, 6.

Smith, P. (June 24, 1971). District court to get 70-30 ratio plan. *Columbus Enquirer*, pp. 1, 6.

Smith, P. (September 11, 1971). School schedules will be staggered. *Columbus Enquirer*, pp. 1, 6.

Wey, H. (1965). *Planning and preparing for successful school desegregation: A guidebook.* Bloomington, IN: Phi Delta Kappa International.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 2, 25-32

The Role of Technology Specialists: Case Studies of Change Agents

Kenneth M. Proctor
Henry (GA) County Schools

Martha J. Livingston
Valdosta State University

Teachers, despite their desire to use computers and other interactive technologies in their teaching, lack sufficient or appropriate education and training to enable them to use technology effectively (Lee, 1996). Technology specialists are becoming the change agents for technology underutilization and a technologically ill-prepared population. The purpose of this study was to investigate the impact of technology specialists on the use of technology among teachers in rural school districts in Georgia. Qualitative strategies were employed to complete this study. Five technology specialists were interviewed and observed in the field to determine the roles they have in increasing technology usage by teachers. Technology specialists were found to provide assistance to teachers in three roles of technology support which include: staff development, technical assistance, and administrative support.

In the last decade, computer use has dramatically changed society. Americans have become enamored with computers, just as they have been impressed with automobiles and television sets since their invention. Educators, although slow to recognize many societal trends, have reluctantly, over the last decade, begun to forge into the information age and experiment with emerging technologies. The authors of the 1997 report of the Southern Technology Council stated that emerging educational technologies were key to solving issues of educational performance, as well as issues of regional and demographic disparities evident in our society.

The purpose of this study was to investigate the roles of technology specialists and the utilization of technology among teachers in rural school districts in Georgia. Georgia has become a leader in the area of supporting technology for teachers as monies were appropriated by the governor and state legislators in fiscal year 1996 to fund technology specialist positions in every school district within the state. Technology specialists are individuals who are employed for the purpose of technology coordination, teacher and staff training, as well as equipment repairing and upgrading in school districts. School district administrators have had wide discretion regarding the personnel hired to fill these positions. Some administrators have chosen to place a part-time technology specialist in each school, whereas other administrators have employed technology specialists who serve several schools.

As they have been employed in many states to support teachers, technology specialists are becoming the change agents for technology underutilization and the technologically ill-prepared teacher population (Southern Technology Council, 1997; Strudler, 1996). Teachers, despite their desire to use computers and other interactive technologies in their teaching, lack sufficient or appropriate education and training to enable them to use technology effectively (Lee, 1996). The support that can be provided by technology specialists will help to smooth the transition as technology use becomes more prevalent (Strudler, 1996). Within the scope of this study the research question examined was as follows: Since the funding of technology specialist positions in Georgia during the 1995 - 1996 school year, what has been the role of technology specialists in the utilization of technology by teachers in the instructional setting?

Technology Coordinators

Over the past 10 years, a change in computer use in schools has occurred. From an early emphasis on computer programming, to computer assisted instruction and computer applications, to the current day classrooms with access to a full range of these tools and more, including hypermedia, technology has changed the way instruction has been delivered in some schools. The hypermedia classroom has provided teachers and students access to a wide range of instructional materials, contained either inside the classroom using readily accessible new storage

Kenneth M. Proctor, Ed.D., is Principal in the Henry County, Georgia school district. Formerly, Dr. Proctor worked as a technology specialist in the Ware County, Georgia school district. Martha J. Livingston, Ed.D., is Professor in the Department of Educational Leadership at Valdosta State University. Before beginning her career in higher education, Dr. Livingston was a school Superintendent in Alabama. Correspondence regarding this article should be addressed to Kenneth M. Proctor, Ed.D., Principal, Henry County Schools, 412 Waterford Landing, McDonough, GA 30253 or by e-mail at kproctor@henry.k12.ga.us.

media or outside the classroom through means which have included electronic mail, electronic bulletin boards, and local and wide area computer networks (Bull, Bull, & Sigmon, 1996).

In this environment with complex interactions among the hardware and software, substantial support from highly knowledgeable persons has become essential. Thus in many school districts, technology coordinators have been employed (Bruder, 1990; Kohler, 1995). Similar to the proliferation of terms for professional development, technology coordinators have been referred to by many titles, such as computer coordinator, technology leader, and technology specialist. These terms have been used interchangeably by this author.

Strudler (1996) conducted a qualitative research project involving case studies of three school-based technology coordinators as a follow-up study to an analysis which had been conducted seven years earlier. The longitudinal nature of this research provided some valuable insight into the role on-site technology coordinators played in three elementary schools. As Strudler's (1996) case studies were being analyzed, one of the main barriers to the integration of technology was the lack of staff development and training available on-site. Shelton and Jones (1996) suggested each school campus have an on-site coordinator who would provide technology staff development in which the school's specific objectives were addressed. Lee (1996) stated that along with on-site professional development in the area of technology, teachers must be provided training integrated into teachers' work schedules. Some teachers viewed the scheduling of release time during the school day as an indicator of administrative support and importance (Shelton & Jones, 1996; Zeitz, 1995). Release time for teachers, or the lack of release time, further provided evidence for the employment of technology coordinators available during the teachers' work day (Harris, 1995; Lee, 1996; Zeitz, 1995).

Finkel (1993) and Harris (1995) discussed the importance of these coordinators being colleagues who are a part of the existing social system of the school and who are viewed as being available for long-term support. Although technology coordinators have worked with computer networks and the technical aspect of the equipment, their most important roles have involved the continuous training they provided teachers. Modeling lessons, helping to choose and use new software, and providing follow-up support are activities in which technology coordinators have been directly involved (White, 1995). Meltzer and Sherman (1997) stated that the provision of this on-going support for teachers was the primary role of the technology coordinator and a much more important role than supervising students or selecting and maintaining software and equipment.

In a doctoral dissertation, Baird (1995) examined on-going support and technology coordination to identify a set of competencies which involved four domains for the emerging role of the technology coordinator. This researcher found that technology coordinators needed skills in the functional domain which represented the knowledge that helped educational leaders function effectively in their jobs. Skills necessary in the programmatic domain were those skills which enabled technology coordinators to plan and to implement effective programs of instruction. Dealing effectively with the school and community environment were the skills and knowledge necessary in the contextual domain. Finally, the personal/interpersonal domain was described by the researcher as the skill a technology coordinator must possess to be effective in working with others.

Baird (1995) concluded that the task of coordinating technology required a balance of technical, organizational, and interpersonal skills. Limited budgets often were cited by educators as forcing them to do without technology coordinators altogether or to choose between hiring technicians who could keep the equipment running and curriculum specialists who could address technology integration (Harrington-Lueker, 1996). Baird (1995) further described the need for quality preparation programs for coordinating instructional technology in which these two areas of technical support and curriculum integration could be brought closer together.

Because the employment of a technology coordinator for each school in a district was generally not financially possible, many district administrators have employed coordinators to serve several schools (White, 1995). These coordinators have been a valuable asset to school district instructional programs. District technology coordinators have helped to create, implement, and direct a district-wide vision for educational uses of technology by aligning educational technology with instructional goals and objectives and ensuring equity and consistency in technology programs among school buildings and across grade levels (Kinnaman, 1990).

An equally important role for these district-level coordinators was serving as the district's primary staff developer. Kinnaman (1990) and Orwig (1994) discussed the idea of the district coordinator training a core group of technology users comprising a cross section of the district's subject areas and grade levels. These teachers became the in-house technology experts in each school building. In this manner, one district technology coordinator had an influence on many teachers in the district. Shelton and Jones (1996) discussed this trainer-of-trainer model in terms of an implementation effort in which trainers were given various levels of certification. These site-based trainers were able to provide their staff members with training which ranged from the basics to more

advanced application-specific training depending on the teachers' readiness levels.

Technology and School Culture

November (1990) investigated the concept of school culture and offered information to guide technology coordinators toward developing a technology culture. Through framing a vision in which technology coordinators could serve as catalysts in the restructuring process, change was brought about in the cultures of schools. The need for a solid research base wherein computer coordinators helped teachers serve as action researchers was discussed by this researcher. November (1990) also suggested that computer coordinators work as interns at companies to see firsthand how major industries used and applied technologies to solve problems and to improve productivity. Finally, this investigator advised that educators become involved in technology institutes in which teachers developed long-range plans to cultivate a culture where higher-order thinking skills were used by students and educators.

The work of culture restructuring by the technology coordinator is not always easy and often leads to feelings of discouragement on the part of the coordinators. Technology coordinators often felt frustrated and disturbed by the work they had been charged with undertaking (Bruder, 1990). The five most common grievances discussed by these coordinators in a national survey were (a) lack of time and staff, (b) lack of administrative support, (c) lack of teacher interest, (d) lack of budget, and (e) lack of additional compensation and appreciation. Dilemmas such as these were often cited by technology coordinators as impediments to their work being performed in a desirable manner (Bruder, 1990). Most coordinators, however, found ways to cope with these frustrations and to support teachers and students in making technology innovations and computer-based instruction successful (Meltzer & Sherman, 1997).

Methods

Participants

Five experienced technology specialists employed in the 41 school districts served by Valdosta State University (VSU), which is a regional university located in rural south Georgia, were the population for this study. Selection of the technology specialists was based on reputational nominations (Glaser & Strauss, 1967; Merriam, 1988). Phone referrals were conducted with technology consultants employed by administrators of the five Regional Educational Service Agencies located in the VSU service area. The purposive sample (Patton, 1980) selection was based on identification of technology specialists who were demonstrating exemplary practices in the area of instructional support for teachers. Attention was given to choose technology specialists who have provided technology training for teachers in their school districts.

Three of the participants were female and two were male. Regarding ethnicity, all of the subjects were Caucasian with an average age of 45 (SD = 4.14). Two specialists held Master's degrees; the remaining three held Specialist's degrees. The subjects' mean years of experience within the field of education was 21 years, and the average number of years each specialist had worked with technology ranged from 2 years to 15. Because the state of Georgia only made funds available for these positions in fiscal year 1996, some technology specialists selected for this study may have been employed for a period of two and one half years in technology positions. Some specialists may have been employed for a shorter period of time, while others may have been employed longer, if their schools districts made funding for technology support a priority from local funding allocations prior to the 1995 - 1996 school year.

Instrumentation and Procedures

A researcher-developed interview guide (See Figure 1) entitled Technology Specialist Interview Guide was used to direct the interviews conducted by the researcher with the technology specialists in the five school districts. The nature of qualitative research indicates that interviewees may digress onto other topics; therefore, the interviewer exercised discretion in the utilization of the instrument to direct the discussion. In most instances as the interviews progressed, questions were raised by both the interviewer and the interviewee and their answers were pursued (Miles & Huberman, 1994).

1. Discuss your background in education or prior to coming to the field of education.
2. Describe your typical work day.
3. If you could design an ideal work day, in what ways would it differ from your typical work day?
4. In what ways have you helped teachers integrate technology into their instructional practices?
5. In what ways have you helped teachers integrate technology into their record keeping practices?
6. Discuss the types and formats of the teacher training sessions you have conducted.
7. In what ways have teachers indicated to you that your training has been helpful to them?
8. Many researchers indicate that technology training is most effective if conducted in the actual workplace. What has been your experience with this phenomena?
9. Give me some specific instances of which you are aware that teachers have used training you provided for them to benefit students or themselves.

Figure 1. Technology Specialists Interview Guide

In the five counties in which this study was conducted, a letter was sent to superintendents requesting permission for the research to be conducted. These letters were followed up with a phone call to confirm receipt of the letter and obtain final permission of the superintendent. Each technology specialist chosen for this study was then contacted by phone to discuss the nature of the research project, to seek the specialists' approval to participate in the study, to discuss the method in which they were referred for this study, and to establish an interview date. All participants were asked to sign informed consent statements with specific concerns concerning confidentiality and rights to withdraw throughly explained.

The superintendents were asked to allow the researcher to spend a day with their technology specialists during the course of which an in-depth interview was conducted. This day was used to get a sense of a typical day for the technology specialist as well as to ascertain the context of the school system. Follow-up phone discussions may have been needed to complete the case studies.

Interviews were conducted with the technology specialists during November and December of 1997, as well as January of 1998. These interviews were audio taped so that all information could easily be analyzed. The audio tapes of each interview were transcribed by the researcher. Field notes were also taken by the researcher. Triangulation between the interviews and field notes was conducted to ensure that information had come from more than one source in the construction of the case studies (Miles & Huberman, 1994). Responses of the interviewees were examined and coded to determine if consistent themes or underlying assumptions were present (Miles & Huberman, 1994). Coding techniques involved the use of a series of symbols written on the interview transcriptions to label reoccurring themes as technology specialist utilized common words or phrases to describe their work and answer the interview questions. Findings regarding technology specialists' perceptions of their roles in helping teachers use technology were synthesized and presented in narrative form.

Results

Synthesis of the Case Studies

Although technology specialists who were interviewed had various experiences working in education and diverse technical expertise, common themes were apparent in the information provided in their interviews. Three of the five specialists interviewed indicated they were the only employee in their county employed in the area of technology. The need for more personnel was emphasized by all of the specialists because not enough time or people were available to address all of the needs and requests made by administrators, teachers, and students. Staff development, technical assistance, and administrative support were the three areas of need most often discussed by the technology specialists during the course of the interviews and field observations.

Staff development was emphasized as the major focus of the technology specialist positions when first implemented by the administrators at the Georgia Department of Education. Only two of the specialists interviewed thought adequate staff development in technology had been conducted in their systems. Although the specialists had offered professional development sessions on basic computer operations (*Windows* functions, grade book programs, and the Internet), three specialists were not satisfied with the quantity of sessions held.

Inadequate facilities and computer equipment coupled with the lack of time were the most often discussed reasons for staff development being limited. One specialist said "there is so little time during the day to prepare or plan for staff development sessions that I find it difficult to justify offering them. Originally the technology specialist's job was for the integration of technology into the curriculum, if the computers are not running, it is hard to use it (technology)." The specialist who made this statement summarized the feeling of all five specialists and helped to disclose a certain level of frustration experienced in their positions.

Training has been offered by the specialists for teachers and administrators at varying times through a variety of formats. After school sessions were reported as the time most often set aside for technology training. School system administrators have provided release time for teacher training in all of the districts involved in the case studies. Summer sessions were provided by four of the specialists interviewed. The most effective training, as stated by all technology specialists, has been the hands-on approach with software the teachers believed was useful to them. Some examples of this software included grade book programs and word processing packages which helped the teachers with record keeping.

Because each teacher was at a different level of technology literacy, these five technology specialists indicated that individualized informal sessions were the most beneficial for many teachers. One specialist stated, "most teachers were very interested in learning although some were hesitant. They like the fact that I come to their school on a certain day, and they don't feel as embarrassed to ask questions when I am in their classrooms. Many times, in front of their colleagues, they feel as though their questions are stupid--especially if some of the teachers are computer literate in the room." Another technology specialist concurred by stating that "the individual training with teachers to show them how to get past a problem so that they can move forward has been the most rewarding and beneficial both for me and them."

An elimination of fear of the equipment was seen as the greatest accomplishment by some of the specialists through their training efforts. Experienced teachers who had little experience in working with technology were extremely fearful that they would do irreparable damage to the equipment. A technology specialist stated "teachers have gotten more comfortable with it, and they are allowing the children to explore more--as they learn more, they are more willing to allow the children to show them how to do things on the computer."

As reported earlier, staff development is only a part of the job responsibilities of the technology specialists. Technical expertise was also a necessary component of the technology specialists' jobs. Two of the specialists revealed that more than 90% of their work was devoted to technical assistance. Repairing malfunctioning equipment, solving software glitches, and providing assistance with computer networks were responsibilities assigned to the specialists. Media specialists were identified by four technology specialists interviewed as the contact person at the school site to whom technical problems were reported. Frequent contact was made with the media specialists by the technology specialists to ensure that problems were handled in a timely and efficient manner.

All specialists reported that ideally there should be personnel who work with staff development and curriculum issues and others who dealt with the technical problems. One specialist stated, "The best way for the day to be set up is to have curriculum specialists who help with the integration--lesson plans--record keeping, and technicians who deal with the hardware problems." Unfortunately, only two school systems in this study had more than one employee in the technology department. Two technology specialists reported feeling uncomfortable and lacking confidence in working with most technical problems. These two specialists frequently used computer vendors to solve technical problems.

Administratively, some activities must be completed for technology to function properly. Purchasing and ordering equipment which met certain specifications and the technicalities of bid requirements were discussed by most of the specialists as frequent job related tasks. A considerable increase in administrative time allotted to administrative technology had been noticed by all of the technology specialists as the department of education implements the *REGIIS XXI* computing system. *REGIIS XXI* is an upgrade project for administrative computing in the Georgia Department of Education primarily for accounting purposes to offset the year 2000 anticipated computer glitches. In three of the school systems, the technology specialist was the only full-time person in the department and was responsible for all paper work related to technology for the local board of education as well as the Georgia Department of Education.

Small districts have often added additional job assignments to specialists responsible for technology. These additional assignments were required for four of the five technology specialists interviewed for this study. Title I programs, staff development programs, and personnel were examples of some of the additional responsibilities assigned to the technology specialists in this study. Time and balance, therefore, have become issues for these four technology specialists as they continue a high level of technology support for classroom teachers, while providing administrative technology support.

Conclusions and Discussion

Findings of this study revealed examples of how technology specialists provide assistance to teachers and administrators in the roles of staff developer, technology technician, and technology administrator. Results from the present study are consistent with findings of previous researchers (Bruder, 1990; Gilmore, 1995; Strudler, 1996) who found that those technology specialists who worked with teachers in these areas were effective in improving the technology skills of teachers as well as the use of technology by teachers in the instructional setting.

Staff Development

When the technology specialists positions were funded by the Department of Education for the 1995-1996 school year, teacher staff development in technology was to be a major component of the job description. Technology specialists interviewed for this study expressed concerns about their job performances in the realm of staff development as well as meeting the staff development expectations of teachers, administrators, and state department officials. Just as the authors of the Southern Technology Council's report of 1997 suggested, the primary concern expressed by the technology specialists in the present study was that of insufficient time for teacher training. Technology specialists in this study expressed that teacher training in technology was not one of the tasks on which they were able to devote the necessary time. In fact, teacher training was often the area which received the least attention.

According to the responses of technology specialists interviewed for this study, reasons for this lack of staff development were sundry. First, a number of school districts across the state were not ready to begin a major staff development program in technology because hardware was not in place or was in bad repair. Second, funding for the needed hardware and staff development

release time has been an impediment in a number of districts. Third, lack of time has been a problem, because several technology specialists in small districts have more than one job title and must balance their time between the important tasks for each position.

Officials in the Technology Division of the Georgia Department of Education have realized the constraints on technology specialists in terms of providing staff development as well as the importance of peer-to-peer training and have developed a model for ensuring that teachers at the school-level have the skills to provide this training for each other. The *InTech* program (Georgia Department of Education [GADOE], 1998) was designed in three phases. Phase I, entitled Integration, was designed to support and enhance the existing curriculum where technology will drive change in teaching and learning. Innovation is emphasized in Phase II. Experiences in Phase II offer challenging options for classroom teachers to make, do and create with productivity packages, multimedia authoring tools, Internet applications and instructional software. In Phase III, skilled users of instructional technologies are inspired to create integrated, multidisciplinary, project-based lessons which encourage students to use modern technologies to solve real-world problems. Successful implementation of the *InTech* program will do much to provide staff development for teachers on the processes and procedures of integrating instructional technology into instructional practices through one-on-one or small group interactions.

The findings from this study support the state's justification for the implementation of the *InTech* (GADOE, 1998) program. Hopefully, programs such as *InTech* will help to increase the time for staff development for teachers in the area of technology. Because school teams will be responsible for the delivery of this training to the teachers in their schools, technology specialists should be relieved of some of their staff development responsibilities.

Most of the technology specialists in this study worked with teachers in the area of technology staff development and curriculum integration to help them increase the use of technology in their classrooms. All of the technology specialists interviewed discussed the need for more time for teacher training and staff development. In order for additional training to occur, most of the specialists expressed the need for more personnel in the technology department and the need for more release time for teacher staff development.

Researchers (Gilmore, 1995; Hurst, 1994; Lauro, 1995; Strudler, 1996) indicated that technology training was most effective if conducted in the actual workplace. The specialists interviewed for this study agreed with this research; however, some reported that, because of equipment and personnel limitations, site-based training had not been possible for them. "Logistically, this is not possible for us because there is only one person in our system--time is a problem," was the response of one specialist. "Although individual training is best if offered at the school site, county training has been more effective at the training lab over a longer period of time so that the depth and scope of the training can be greater," were the thoughts of another specialist. Two technology specialists reported that as long as the hardware was similar and the software was the same, the training did not have to actually be conducted at the teachers' school site on the teachers' actual equipment. On the other hand, if the hardware was different, it seemed to be difficult for the teachers to transfer the concepts to their own equipment.

Technical Assistance

Technical assistance was the second area technology specialists discussed as critical for technology advancement in classrooms in this study. This finding was congruous with Baird (1995) who concluded that on-going technical support was required as a part of the balance of technology coordination. If the hardware is not functioning properly, it is of little use to teachers and students. Previous researchers (Bennett, 1996; Niederhauser, 1996) have concluded that technical support was often cited as a major need by teachers in technology-rich environments. Therefore, highly skilled technology specialists were needed.

Administrative Support

The findings from this study revealed that administrative tasks must be performed by technology specialists in order for an educational technology program to be successful. These tasks suggested by technology specialists and teachers in interviews and surveys involved scheduling and arranging the logistics of technology training. Additionally, tasks such as completing purchasing agreements, ordering materials, complying with Georgia Department of Education standards as well as standards for local boards of education were discussed by technology specialists.

Technology planning is another administrative task in which technology specialists indicated that they should be involved. Short and long range comprehensive technology plans should be developed by technology specialists, administrators, and teachers for school districts and individual schools. These plans should include technology hardware and software acquisition as well as plans for technology professional development. These findings were consistent with Baird (1995) who discussed the importance of a balance of skills including organizational administrative matters in the area of technology.

As a part of this planning, technology specialists indicated that they must be deliberate in their efforts to

change the culture of schools in terms of technology usage. November (1990) suggested that technology planners develop a vision in which technology specialists could serve as catalysts in the change process. Technology specialists must be constantly diligent in their efforts to encourage teachers and administrators regarding the benefits of technology usage. The findings of Kinnaman (1990) are consistent with the findings in this study that teachers must individually become convinced that their students will benefit from the infusion of technology into the classroom.

Technology specialists were found to engage teachers in a process of learning about technology and exploring how it might be used in the classroom. It was suggested by Muir (1994) that the focus should be on methods and strategies in which technology allows teachers to do that which they were never able to do before, such as instantaneous worldwide communication, simulations, and hypermedia presentations. Muir further indicated that teachers should become creatively involved in technology planning as they learn the effects technologies can have on their students' learning.

In accordance with the findings of this study, White (1995) also concluded that technology coordinators have been a valuable addition to school districts' instructional programs. Consistent with Kinnaman (1990), technology coordinators helped to create, implement, and direct a vision for educational uses of technology in this study. Staff development, technical, and administrative assistance were found to be provided by technology specialists for teachers and administrators to ensure equity and consistency in technology programs among school buildings and across grade levels.

From the findings in this study, it is concluded that technology specialists in rural south Georgia have impacted the use of educational technology among teachers since they were funded during the 1995 - 1996 school year. Technology programs and plans must be comprehensive in nature so that everyone's need are met--beginners as well as more advanced users. Technology specialists were found to be essential persons in the development and execution of these plans according to the responses of technology specialists.

It is further concluded from the findings of this study that technology specialists are aware of the time constraints of their positions, and recognize the areas to emphasize and de-emphasize. The lack of time was discussed by all of the technology specialists interviewed. Evidence in this study points toward the need for technology specialists to prioritize daily tasks that must be accomplished. Unfortunately, teacher training was discussed by the technology specialists as the area that often received the least amount of the technology specialists' attention. Technical assistance and equipment repair were the tasks reported to take the largest portion of the technology specialists' time. Administrative duties were often revealed by the technology specialists to be interspersed throughout the work day.

Lack of personnel in the technology department was discussed by all of the technology specialists interviewed. Interview data also suggested the ability of all of the technology specialists to adequately provide technology expertise in the areas of hardware and software acquisition and maintenance. Therefore, it is concluded that technology specialists are able to articulate the personnel, hardware, and software needs of their districts.

In small rural districts in south Georgia, it was concluded that the technology specialist is often assigned duties that conflict with the role of the technology specialist. Because of a lack of funds to employ individuals for each job title, technology specialists reported that they often had job responsibilities other than technology. Three technology specialists had two or three job titles in addition to technology specialist. Time was taken away from the mission of developing technology literate students, teachers, and administrators when the technology specialists had to be assigned multiple job responsibilities.

The search for a model of teacher training and support involving technology specialists is a continuing process. Although an exemplary model may be difficult to construct, further research into how technology specialists can help improve technology skills for technology practitioners is important. The recommendations for further study include:

1. Conduct studies focusing on curriculum integration in computer-intensive classes, because the implementation of computers into the curriculum requires a fundamental change in how teachers teach.

2. Conduct studies comparing the efficiency of on-site technology specialists with technology specialists who are responsible for multiple school technology programs.

3. Conduct a longitudinal study of other technology programs and technology specialists to determine if staff development, technical assistance, and administrative support continue to be the roles in which technology specialists function.

4. Investigate exemplary technology practices to identify the job descriptions of technology specialist that will strengthen technology programs.

5. Replicate this study with schools in which teachers have been involved in the *InTech* (GADOE, 1998) program to see if technology staff development for teachers has improved.

6. Investigate the administrative roles of technology specialists for efficiency so that sufficient time is invested in providing teachers with instructional support.

Until further studies have been conducted, readers should be judicious in generalizing or transferring the findings of this study. Research based on these suggestions may confirm the findings of the present study and continue to add to the existing body of knowledge focusing on teacher training in technology and computers as instructional tools.

References

Baird, N. T. (1995). Competencies for the emerging role of instructional technology coordinator (Doctoral Dissertation, University of Georgia, 1995). *Dissertation Abstracts International, 56*, 2942.

Bennett, C. K. (1996). Schools, technology, and educational leadership. *NASSP Bulletin, 80*(57), 53-59.

Bull, G., Bull, G., & Sigmon, T. (1996). An electronic journey. *Learning and Leading with Technology, 23*(5), 17-35.

Bruder, I. (1990). The third computer coordinator survey. *Electronic Learning, 9*(4), 24-29.

Finkel, L. (1993). You get what you pay for. *Electronic Learning, 12*(8), 18.

Georgia Department of Education Educational Technology Training Center. (1998). Course catalog: Winter 1998. [Brochure]. Atlanta, GA: Technology Services.

Glasser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Chicago: Aldine.

Gilmore, A. (1995). *Turning teachers on to computers: Evaluation of a teacher development program.* [WWW document]. URL http://www.elibrary.com

Harrington-Lueker, D. (1996). What makes a good professional development program? *Electronic Learning, 16*(1), 35.

Harris, J. (1995). Teaching teachers to use telecomputing tools. *The ERIC Review, 4*(1), 2-4.

Hurst, D. (1994). Teaching technology to teachers. *Educational Leadership, 51*, 74-77.

Kinnaman, D. E. (1990). Staff development: How to build your winning team. *Technology and Learning, 11*(2), 24-30.

Kohler, A. P. (1995). Perceptions of the position of building level technology coordinator within the state of Florida (Doctoral Dissertation, University of Central Florida, 1995). *Dissertation Abstracts International, 56*, 3925.

Lauro, D. (1995). *Five approaches to professional development compared.* [WWW document]. URL http://www.elibrary.com/cgi-bin/hhweb/hhfetch.html

Lee, C. (1996). *Teacher training for technology.* [WWW document]. URL http://www.elibrary.com.html

Meltzer, J., & Sherman, T. M. (1997). Ten commandments for successful technology implementation and staff development. *NASSP Bulletin, 81*(585), 23-32.

Merriam, S. B. (1988). *Case study research in education: A qualitative approach.* San Francisco: Jossey-Bass Publishers.

Miles, M. B., & Huberman, A. M. (1994). *An expanded sourcebook: Qualitative data analysis.* Thousand Oaks, California: SAGE Publications.

Muir, M. (1994). Putting computer projects at the heart of the curriculum. *Educational Leadership, 51*(7), 30-32.

Niederhauser, D. S. (1996). Using computers in an information age classroom: What teachers need to know. *NASSP Bulletin, 80*(58), 71-80.

November, A. C. (1990). The emerging role of the computer coordinator. *Electronic Learning, 9*(4), 8-9.

Orwig, A. H. (1994). Begin with teachers and watch students benefit. *Technology and Learning, 15*(1), 74-76.

Patton, M. Q. (1980). *Qualitative evaluation methods.* Beverly Hills, CA: Sage Publications.

Shelton, M., & Jones, M. (1996). Staff development that works! A tale of four t's. *NASSP Bulletin, 80*(582), 99-105.

Southern Technology Council. (1997). *Making technology happen: Best practices and policies from exemplary K-12 schools.* Research Triangle Park, NC: Author.

Strudler, N. (1996). The role of school-based technology coordinators as change agents in elementary schools. *Journal of Research on Computing in Education, 28*(2), 234-254.

White, J. (1995). Whenever you call: How to provide teachers with training and support when and where they need it. *Electronic Learning, 15*(3), 16.

Zeitz, L. E. (1995). Developing a technology workshop series for your faculty and staff. *The Computing Teacher, 22*(7), 62-64.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 2, 33-43

Transforming Graduate Teacher Education Classes: Lessons from Educational Psychology

Carol VanZile-Tamsen
State University of West Georgia

Recent findings on the nature of human learning indicate that it is constructivist in nature, situated, and socially mediated. This study qualitatively examined students' reactions to an educational psychology course designed according to these three principles. After experiencing the course, 34 students responded to an open-ended questionnaire. Responses were analyzed using theme analysis. The findings suggest that most students believed class activities to be helpful to their understanding of course content and in learning about new teaching strategies. Negative reactions related to repetition/boredom, length, and the nature of the people in the group. When students discussed what should be covered in such a course, they indicated that they would prefer more hands-on activities in which they could apply course concepts to classroom situations.

A recent trend in education at all levels has been a move away from traditional teacher-centered instruction toward student-centered instruction (O'Banion, 1997; Tishman, Perkins, & Jay, 1995). This transformation in educational practice is related to recent findings about the nature of learning which suggest that meaningful learning is based on knowledge construction which occurs within a particular context and is socially mediated (Anderson, Blumenfeld, Pintrich, Clark, Marx, & Peterson, 1995; Doyle & Carter, 1996; Fosnot, 1996; O'Banion, 1997; Shuell, 1996; von Glasersfeld, 1996). Although much has been discovered about what meaningful learning involves, specific teaching methods which promote such learning are still being developed (Fosnot, 1993a; 1996). The primary purpose of the present study was to examine graduate students' perceptions of the use of structured collaborative groups as a method to foster knowledge construction and meaningful learning. Since much of the research on knowledge construction and meaningful learning has been conducted by educational psychologists (Anderson et al., 1995; Shuell, 1996), it seemed appropriate that the study take place within the context of an educational psychology course for in-service teachers pursuing a masters degree.

Carol VanZile-Tamsen is Assistant Professor, Department of Counseling and Educational Psychology, the State University of West Georgia. The author wishes to thank Mary Ruth Reynolds, Ph.D., for sharing her expertise as a peer debriefer. Correspondence relating to this article should be directed to Carol VanZile-Tamsen, Department of Counseling and Educational Psychology, State University of West Georgia, Carrollton, GA 30118-5170 or by e-mail to cvanzile@westga.edu.

The Nature of Meaningful Learning

Anderson et al. (1995) have suggested that it is important to keep new findings in the study of human learning in the forefront when designing instruction. They identify three aspects of human learning in particular which need to be addressed: people learn by constructing meaning based on prior knowledge and experience, learning is situated, and learning is socially mediated. These three aspects of meaningful learning are outlined below.

People learn by constructing meaning. The constructivist model grew from work done by researchers in the area of cognitive development, specifically Piaget and Vygotsky, as well as from the areas of philosophy and anthropology (Fosnot, 1993a). Constructivism is in direct contrast to early information processing models which suggested that there is an objective knowledge which can be known and understood in the same way by everyone and can be transmitted from teacher to student (Davis & Sumara, 1997; Mayer, 1996). Constructivism suggests, however, that there is no objective knowledge. Instead, knowledge is subjective, created by each individual through the process of equilibration. Knowledge structures are created or reorganized based on the reconciliation of personal experiences with new or discrepant information (Davis & Sumara, 1997; Fosnot, 1993a, 1993b, 1996; von Glasersfeld, 1996). The teacher's role is to facilitate this process for the students. Since there is no objective knowledge which can be transmitted directly from teacher to student, teachers must structure activities in which students are encouraged to actively make sense of information and restructure their knowledge based on the new information.

Learning is situated. Brown, Collins, and Duguid (1989) have suggested that learning is tied to the situation in which it occurs. When students learn information in a class, that information is situated within a class setting, and is not very easily applied to other contexts. They suggest that this application problem relates to the difference between declarative knowledge (knowing what) and procedural knowledge (knowing how). Transfer of learning to a new context will occur more easily if experiences are provided that allow *practice* of the task, rather than just information *about* the task. These practice experiences are called authentic activities and allow students to develop procedural knowledge by encouraging practice of the skills which are required in other settings. O'Banion (1997) agrees with these authors and suggests that learning by doing is best. In addition, it is important to create a culture in the classroom which values the skills being learned and maintains and fosters the development of these skills (Brown, Collins, & Duguid, 1989; Fosnot, 1996; O'Banion, 1997; Tishman, Perkins, & Jay, 1995).

Learning is socially mediated. In addition to the process of making meaning as an individual, personal knowledge construction is further facilitated by social interaction within a community of learners (Fosnot, 1993b, 1996; O'Banion 1997; Tishman, Perkins, & Jay, 1995; von Glasersfeld, 1996). Through the process of social interaction, individuals can become aware of others' understanding which can further encourage equilibration. In addition, this interaction allows the individuals to come to a consensus about knowledge, negotiating a knowledge base that is "taken as shared" (Fosnot, 1996). Consensual knowledge is important since it allows the members of a community of learners to communicate with ideas that are agreed upon and can be taken as givens (Fosnot, 1996; Tishman, Perkins, & Jay, 1995). In addition, Vygotsky (Rogoff, 1990) has suggested that this social interaction provides important scaffolding interactions in which the teacher and advanced students can help other students develop deeper understanding.

Fostering Meaningful Learning

Authors familiar with the findings concerning the nature of learning outlined above have suggested that instructional methods need to change to facilitate meaningful learning (Anderson et al., 1995; Baxter Magolda, 1998; Davis & Sumara, 1997; Fosnot, 1993a, 1993b, 1996; O'Banion, 1997; Shuell, 1996; Tishman, Perkins, & Jay, 1995; von Glasersfeld, 1996). The first suggestion offered is that the instructor develop a community of learners in each classroom, in which a culture that values learning and discovery is established and students are engaged in discussion and debate with classmates and the instructor (Fosnot, 1993b; O'Banion, 1997; Tishman, Perkins, & Jay, 1995). Through this dialogue within a learning culture, students will be encouraged to equilibrate what they are learning with what they already know and understand, thus promoting reorganization of knowledge structures (Davis & Sumara, 1997; Fosnot, 1996). Because this dialogue, or social mediation, is so very important to the learning process, cooperative learning is one method which is seen to be helpful in fostering meaningful learning (Jost, Havard, & Smith, 1997; Nelson, 1990; O'Banion, 1997).

Another method which is seen as useful in promoting meaningful learning is the use of authentic or meaningful tasks within the classroom setting (Baxter Magolda, 1998; Brown, Collins, & Duguid, 1989; Jost, Havard, & Smith, 1997; O'Banion, 1997). The use of these authentic, or real-world, tasks can encourage students to take the concepts and skills they learn in the classroom and more readily apply them in practical settings. Evaluation of learning would involve evaluating students' understanding of concepts, as well as their performance of important tasks.

Several undergraduate and graduate programs have attempted to incorporate the use of these methods to encourage students to become knowledge constructors and engage in meaningful learning (Baxter Magolda, 1998; Fosnot, 1993a; Holt & Johnston, 1989). The usefulness of these programs for individual students, however, often depends on the beliefs that the students have about the nature of learning (Baxter Magolda, 1998; Holt & Johnston, 1989; King & Baxter Magolda, 1996). King and Baxter Magolda (1996) found that students' beliefs about the nature of knowledge (i.e., absolute or context dependent) influence their ability to learn in a classroom or program that is based on knowledge construction. In addition, Holt and Johnston (1989) found that in-service teachers achieve differently in a constructivist-based masters program based upon their philosophies of teaching and learning prior to entering the program. These findings suggest that students who believe that knowledge must be disseminated from an all-knowing teacher will probably not be as open to, or as successful in, constructivist-based classes or programs. This caution is especially important for teacher education since many teachers enrolled in graduate programs have strongly ingrained beliefs about what teaching and learning involve (Fosnot, 1993a).

Since students' beliefs about what constitute teaching and learning are so important to what is learned within a class or program, these constructivist-based programs also emphasized identification of students' prior knowledge and beliefs (Baxter Magolda, 1998; Fosnot, 1993a; Holt & Johnston, 1989). Once beliefs are made explicit, they can be questioned. In addition, this identification process allows students and instructors to identify content areas which the students want and/or need to learn. Allowing

mature students to assist in determining what they learn enhances their motivation and interest in a program, as well as their perceptions of the program's value (Knowles, 1990).

Reforming a Graduate Educational Psychology Course

Based on constructivist ideas about the nature of meaningful learning, educational psychologists have exchanged much dialogue concerning changes in instructional design for educational psychology courses for pre-service teachers (Anderson et al., 1995; Doyle & Carter, 1996; Shuell, 1996; VanZile-Tamsen, 1997). The suggested changes are centered around incorporating student-centered methods which emphasize authentic activities to help pre-service teachers practice teaching skill. Anderson et al. (1995), Shuell (1996), and VanZile-Tamsen (1997) have indicated that educational psychology instructors should use more authentic classroom activities as teaching tools so that courses move away from what pre-service teachers know to focus on what they can *do*. In addition, evaluation methods in these classes should move away from paper-and-pencil tests of declarative knowledge toward methods which evaluate both understanding of course concepts and performance of important skills. Examples of authentic activities which have been proposed for pre-service teachers in educational psychology courses have included the development of lesson plans which incorporate findings from recent cognitive research (Shuell, 1996) and the analysis of "rich" cases that allow pre-service teachers to "practice" effective classroom management techniques (Anderson et al., 1995).

Pre-service teachers also believe that changes in class activities are essential to enhancing the usefulness of educational psychology courses for teacher preparation. In a qualitative examination of students beliefs about the value of an educational psychology course before and after student teaching, VanZile-Tamsen (1997) found that pre-service teachers see declarative knowledge about students and teaching as important before student teaching. However, after they have experienced a real classroom for the first time, these pre-service teachers call for more hands-on activities which provide practice in the following areas: managing the classroom, dealing with discipline problems, making modifications to individualize instruction, and motivating students.

Since a literature search revealed no articles addressing reform in the design of graduate educational psychology courses for practicing teachers, the purpose of the present study was to examine in-service teachers' perceptions of collaborative activities in an educational psychology course designed for graduate students who plan to remain teachers once they earn the advanced degree. The main research question addressed in the study was, from students' perspectives, what are the advantages and disadvantages of using social mediation to foster meaningful learning and the development of usable knowledge? There were two additional research questions related to the educational psychology concepts covered in the course. Adult learning researchers (e.g., Knowles, 1990) have suggested that adult learners should have a say concerning which concepts they learn in a course based on what they believe they need to know. Topics covered in the present course were determined before the course began with no input from students. To examine which topics were fulfilling students' needs and which topics should be included in the course, the following two research questions were asked: what topic that was covered in the course do practicing teachers believe will have the greatest impact on their teaching; and what topics do practicing teachers believe should be included in an advanced educational psychology course for in-service teachers.

Method

Participants

The participants were 34 practicing teachers enrolled in graduate programs in teacher education at a mid-size, Southeastern university and were recruited from the investigator's Psychology for Classroom Learning Class. Of these students, 28 were enrolled in masters programs, while 3 were Ed.S. students. One student was enrolled in a certification-only program, and two did not indicate their degree program. There were three males and 31 females in the class; one of these women was Black, and the rest of the students were White. The average age of the participants was 36, while, on average, they had completed 33.44 quarter hours in their programs before taking this class. The average grade point average (GPA) of these students was 3.87. Their majors were early childhood education, middle grades education, secondary education, special education, and speech and language pathology; and there was one administration major.

Procedure

The participants responded to an open-ended questionnaire at the end of the Fall quarter after they had experienced Psychology for Classroom Learning, an advanced educational psychology class. This course was designed to be constructivist in nature and centered around a course packet containing six articles from recent journals, each related to current issues in educational psychology (Brown, Collins, & Duguid, 1989; Covaleskie, 1992; Gindis, 1995; Mayer, 1996; Moran, 1991; Shuell, 1996), and two chapters from recent

educational psychology texts (LeFrancois, 1997; Ormrod, 1997).

For six of the ten quarter classes, students were to come to class having read the assigned reading. They were to turn in a summary in which they identified five aspects of the reading which they already knew, and five questions that still remained about the reading. When they arrived at class, they were divided into groups of five or six and were to discuss the things they understood from the reading and attempt to answer each other's remaining questions. Once they had reached the limits of their understanding in groups, they were to identify remaining questions. Each group was then to briefly summarize their discussion for the rest of the class and list these questions. Student identification of things they knew and did not know about each topic was meant to encourage the process of equilibration and deep understanding as students read course materials. Group discussions of these articles in class were designed to offer social mediation to further foster equilibration, clarify understanding, and promote knowledge construction.

The instructor recorded the remaining questions for each group as the groups presented them and then presented a lecture designed to answer these specific questions and concerns. In several cases, the questions students had concerning the readings were not anticipated by the instructor, and the instructor's presentation went in a completely different direction from what had been anticipated. The focus of the lecture presentation and additional small group activities was on applying the material covered in the readings to actual classroom situations. Students in the class were constantly reminded to look for implications for the classroom as a way to promote transfer. In addition, one of the goals of the graduate educational psychology class is to encourage reflective practice. Group discussions and activities, as well as written assignments, were designed to encourage students to become more reflective in their practice by having them reflect on their own teaching and relate what they actually do to concepts discussed in course reading materials.

For two of the class periods, students were to bring to class an empirical research article which addressed a particular topic chosen by the instructor. The first topic was classroom management and discipline, and the second topic was mainstreaming and inclusion. For each of these class periods, students were to come to class having read the article and with an article summary sheet completed. This summary sheet was a worksheet which led them through the creation of a summary of key points in the article (theoretical background, hypotheses, participants, procedure, materials, design, results, conclusions, strengths, and weaknesses). Once students arrived in class, they were grouped according to certification area. In these groups, they were to share the main findings and implications of their articles. After each person had shared, they were to find common themes and overarching implications for practice of all of the articles summarized in the group. They were then to present to the whole class the main ideas to come out of their small group discussions. Their presentation to the class was evaluated, and points earned contributed to the final grade. At the end of class, they handed in the article summary sheet, which was graded and contributed to the final grade. The final project was a formal critique of a research article of their choice.

There were two evaluations of their learning progress during the term, the goal of which was to assess their understanding of course concepts and to encourage active reflection on their own teaching. These consisted of a mid-term and final integrative essay in which they were to summarize important points from all course materials, including the class research presentations, identify the key implications of these materials, discuss how these materials related to their own beliefs about teaching and learning, and discuss how these materials had impacted or would impact their own teaching. Six of the articles from the course packet were to be addressed in the final essay; students who responded to the open-ended questionnaire were told that they need only address three of these articles in the final essay, and they could choose the articles which they felt to be most relevant. Only one student chose to address all six articles rather than complete the questionnaire. Also contributing to the final grade was class participation.

Data Analysis

The responses to questions on the questionnaire for each participant were separately entered into tables in a word processing package. Participants were then grouped according to certification area. Case ordered matrices were formed (Miles & Huberman, 1994) in which the participants were listed in one column of a large chart. Their responses to the research questions were also listed in this chart; each research question constituted one column of the chart. Key themes were identified within each person's responses and compared across certification groups. The key themes are summarized into the taxonomies shown in Tables 1-3.

The main strategy used to enhance the validity of the findings was the use of participant language and verbatim accounts (McMillan & Schumacher, 1997). Where possible the words of participants were used to label themes. In addition, in Tables 1 and 2, each identified theme is presented with a statement from a participant which seems to exemplify that theme. In this manner, the reader is able to see how the themes were derived. A secondary method of enhancing validity was to actively search for discrepant

data during data analysis (McMillan & Schumacher, 1997). Once discrepant data were identified, themes were further delineated to account for these discrepancies.

A major drawback of qualitative data analysis relates to the role that investigator bias can play in determining results. The greatest source of bias in the present study relates to the fact that the instructor was also the investigator. This fact could result in the identification of only those themes thought to be relevant by the instructor or identification of only those themes with which the instructor agrees. To help minimize bias in the data analysis, a peer debriefer was used to examine raw data and resulting themes (McMillan & Schumacher, 1997). The peer debriefer was an objective colleague who also conducts qualitative research and is from the same department as the investigator. After reviewing raw data and resulting themes, the peer debriefer found no apparent bias in the data analysis.

An additional problem related to the fact that the investigator and the instructor were the same person concerns the honesty of the participants. To try to minimize the problem of participants responding with comments they believe the instructor wanted to hear, the questionnaire was completely anonymous. In addition, the investigator explained to participants that the comments would not be read or analyzed until after final grades were turned in and that a graduate assistant would type all comments into a word processing file before they were seen by the instructor.

Results

Reactions to Small Group Discussions

When respondents were asked to comment on the group discussions used in class, positive and negative themes emerged. Thirty-three of the students had positive things to say about the group activities, while only one student had nothing positive to say. Thirty students qualified their positive statements with something negative about the groups, while three students had only positive things to say. Within the positive statements, eight different themes emerged (shown in Table 1).

There are three positive themes which seem to dominate the data: application, sharing/support, and educators from other areas. Each of these themes was mentioned by approximately one-third of the respondents. Many of the students found the groups to be beneficial because the groups facilitated the application of class concepts to actual classroom settings. There was a level of sharing and support in the groups which most teachers, due to the nature of the profession, do not experience on a regular basis; and hearing people from other schools and different educational backgrounds helped broaden students' perspectives.

Table 1
Positive Aspects of Group Discussions

Theme	Frequency (%)	Illustrative Comment
Application	10 (29%)	"Overall, the discussions were helpful and I have tried strategies that were used successfully by my classmates."
Sharing/Support	10 (29%)	"The thing I liked the best was when everyone shared what had happened to their classroom that week. . . . Also, it was amazing to see similar problems in second and fourth grade, which I also have in sixth grade. I didn't feel like I was alone! Comradeship!"
Educators from Other Areas/ Levels	9 (26%)	"I enjoyed having time to share with other teachers because of the variety of subjects taught and because they represented different counties. Teachers don't often get that kind of time to talk with teachers other than ones they work with."
Clarify Understanding	6 (18%)	"Discussing helped me understand the subject better."
Different Points of View	6 (18%)	"I enjoyed seeing a topic from many different viewpoints."
Educators from Same Areas	5 (15%)	"The discussions seemed to 'flow' more easily when the members of my group taught a similar grade and/or had similar educational interests."
Personal Comfort	2 (6%)	". . . I'm usually intimidated in large groups, so I don't ask questions, but I was comfortable with the small groups."
Feedback	1 (3%)	"What I liked best about in-class group discussions was using other professionals as a 'sounding board' when expressing my own opinions. The instant feedback allowed me to either challenge or affirm my teacher beliefs, which helped me 'polish' my present methods used in the special education setting."

A secondary set of themes emerged from the positive comments: clarify understanding (18%), different points of view (18%), and educators from same areas (15%). Students felt that small group discussions helped facilitate their understanding of class readings. In addition, they believed it was helpful to hear people with differing points of view add their own perspectives on the readings. Finally, there were individuals who most appreciated discussions that involved people from similar educational areas. They felt that it was easier to apply topics when they were all basing the application on experiences in similar settings.

There were two positive themes which seem more idiosyncratic since they were mentioned by so few people. First, two students mentioned the idea of personal comfort; they liked the group discussions because they are more comfortable expressing ideas in a small-group setting. Second, one student mentioned feedback. The group members provided instant feedback to individual comments which allowed the individual to better understand and clarify his/her own beliefs, attitudes, understandings, methods, etc.

Based on the comments from class members and the themes which these comments represent, it seems that the small group discussions were a valuable part of class activities. The only difference in themes expressed between certification areas came from the preference to be grouped with educators from the same versus different areas of education. Early childhood and middle grades teachers seemed to like sharing experiences with individuals from different areas, while secondary and special education teachers seemed to have a preference for working in groups of people from the same educational areas.

The ten negative themes are shown in Table 2. The most overwhelming negative theme involved repetition/boredom; about 24% of the students mentioned this in their statements. This repetition/boredom had two aspects. First, each class meeting except for two began with small-group discussions of the assigned reading(s) for that night. Students felt that this became repetitive and unhelpful over the course of the term. The second aspect relates to the two group presentations students did in which they summarized empirical research articles. Students felt that when the small groups presented to the whole class, the information was repetitive and boring and wasted a great deal of time.

The next two negative themes which were represented most frequently in students' comments related to students in the groups. First, 18% of the students indicated that they did not like having unprepared students in their groups. Second, 15% of the students said that having one or more students in a group who were opinionated and domineering decreased the usefulness, and their enjoyment, of the small group discussion.

Table 2
Negative Aspects of Groups

Theme	Frequency (%)	Illustrative Comment
Repetition/ Boredom	8 (24%)	"The small group meetings every night of class were a bit much."
Slackers	6 (18%)	"The major downfall of the discussion groups was lack of participation by some unenthused educators."
Domineering Students	5 (15%)	"The other concern was a dominance of the group by one member's opinion and very vocal nature."
Length	4 (12%)	"The thing I liked least about them is that occasionally they lasted too long."
Mixed Groups	4 (12%)	"The thing I liked the least was being in groups with teachers who did not share any of the same experiences because our teaching assignments were so different."
Lack of Conciseness	3 (9%)	"When we moved from small group to large group I got frustrated when teachers were unable to be concise with what they were sharing. This just means we need more practice in this area."
Off-Task	2 (6%)	"I least liked the tendency of some groups to get off the subject or degenerate into 'gripe-sessions.'"
Lack of Applicability	2 (6%)	"The discussions about the articles themselves were not beneficial, because it was not applicable to my classroom."
Lack of Group Preparation	1 (3%)	"I think my least favorite thing was giving empirical research presentations as a group without any preparation as a group."
Personal Comfort	1 (3%)	"Overall, they were not helpful to me as a teacher. I know people learn in different ways-- I prefer the lecture method. I really get more out of what the professor says than what other people who are taking the class have to say."

Relating again to the empirical research presentations was the negative theme concerning length. Students (12%) felt that these lasted too long and were not useful. Another 12% indicated that they did not have positive group experiences when the groups were composed of individuals with different teaching backgrounds and assignments. As with the positive themes concerning group composition, the individuals who did not like mixed groups were either secondary or special education teachers.

Relating to the theme of repetition/boredom was the theme of lack of conciseness. Nine percent of the students felt that some groups and some group members were not concise when speaking, which made the discussions last longer than necessary. These verbose students went on and on and repeated points that were already mentioned, which caused other class members to lose patience. Groups getting off task and lack of applicability of reading assignments were represented in the statements of 6% of the students. Interestingly, both students who found the reading to lack applicability to the teaching situation were special education/speech language pathology teachers. One student mentioned that the lack of group preparation before class for empirical presentations was a negative aspect of the group experience, while one student mentioned that the groups were not personally comfortable, preferring lecture instead.

From these negative themes, it becomes apparent that three (repetition/boredom, length, and lack of group preparation) relate mainly to the empirical research presentations and suggest that the format of these needs to changed in some manner. The first two of these three themes also relate to the weekly discussions of assigned readings and seem to suggest that students need these less as the quarter proceeds, since they are becoming more proficient at reading and understanding the content of readings and would prefer to use group discussions to apply concepts to classroom settings.

Three negative themes ("slackers," "domineering students," "off-task") seem to relate to problems that will exist in any group situation in a classroom. Making students more accountable for group activities may be one way to help eliminate "slackers" and groups getting off task. However, in a setting like this with students who are already teachers, being off-task may not be that bad as a lot of learning may actually be occurring as they share what happened at school today and what they did about it. In fact, many of the students who mentioned the positive theme of sharing/support seem to be suggesting exactly that. The domineering students will always be present in any type of class, and there is not a great deal that an educational psychology instructor could do to eliminate them; perhaps it is even beneficial that they are present since interacting with such domineering individuals may provide practice for dealing with opinionated and domineering people.

In spite of the negative themes which emerged, most of the students had positive group learning experiences. With some minor changes in class format, it seems that such group experiences can be used effectively to promote the learning of in-service teachers. These findings suggest that incorporating the social nature of learning is very important in a graduate class for in-service teachers.

Topics Which Have Impacted Teaching

Table 3 summarizes the topics covered in class, through readings, lecture, and small group discussions, that the participants believed will or have already had the greatest impact on their teaching. The most commonly mentioned topic, mentioned by 47% of the respondents, is information processing/cognitive learning. This includes 9 people who mentioned theory in general, 2 people who mentioned learning styles, scaffolding, and situated/authentic learning, and 1 person who mentioned prior knowledge. The importance that these students seem to place on topics related to cognitive learning seems to be summarized in the following statement:

> I think our discussion of information processing will have the greatest impact on my teaching. I was surprised by some things, like the length of time information stays in short term memory. I understand the physiology of the process much better than ever before and will work smarter and harder to get things into students' long-term memories.

The second most frequently mentioned topic was the combined topic of mainstreaming and inclusion, mentioned by 26% of the respondents. It seems that mainstreaming and inclusion were seen as very important topics because they occur so frequently in local school systems. Few of these teachers have ever had any training dealing with the special needs of students or in making accommodations to individualize instruction.

The related topics of classroom management/ discipline/morality/and character education were mentioned by 18% of the respondents. Coming in last was motivation, which was mentioned by 12% of the respondents. Taken as a whole, these results seem to suggest that experienced teachers are more interested in the cognitive learning of their students than in other aspects. In addition, because mainstreaming and inclusion is becoming more prevalent and they need strategies to deal with it, this is an important topic.

Table 3
Suitable Topics for Graduate Educational Psychology

Topics that have the Most Impact on Teaching	Frequency	Topics that should be Addressed	Frequency
Information Processing/ Cognitive Learning Theory	16	Behavior/Classroom Management/ Discipline/Morality	14
Learning Styles	2	Motivation	14
Scaffolding	2	Dealing with Personal Difficulties/Issues: Self-Esteem/Emotional/Social/Family	10
Situated/ Authentic Learning	2	Needs of special children	10
Prior Knowledge	1	Student Diversity: Learning styles/ intelligences, ethnic, ses	8
Mainstreaming and Inclusion	9	Cognitive Learning	8
Classroom Management/ Discipline/ Morality/		How to better deal with the parents	6
		Inclusion and Mainstreaming	5
Character Education	6		
		Testing/Psychological Reports	1
Motivation	4		
		Instructional Strategies	1
		Setting Expectations	1

Topics Which Should Be Covered in an Educational Psychology Course for In-Service Teachers

Findings concerning the topic of greatest impact seem to be contradicted by the list of topics students felt should be included in an educational psychology course (shown in Table 3). In terms of topics that in-service teachers feel should be addressed, the related topics of classroom management/discipline/morality/and character education were mentioned by 41% of the respondents. This represents an increase of over 100%. Motivation was tied for number one, also being mentioned by 41% of the people, which is an increase of over 200%.

Coming in second were two new topics which were each mentioned by 29% of the respondents. These two topics were dealing with personal difficulties/issues and needs of special children. Student diversity (on a variety of dimensions), another new topic, was mentioned by 24% of the respondents. As a result of these new topics, cognitive learning was pushed down to third place, going from 47% to 24%, a 50% decrease. Strategies for relating to parents was another new topic mentioned by 18% of the respondents. Mainstreaming and inclusion were now only mentioned by 15% of the respondents, almost a 50% decrease. Three additional new topics were each mentioned by one respondent: testing/psychological reports, instructional strategies, and setting expectations.

Although these findings about topics which should be covered are inconsistent with the findings concerning which topics have the most impact on teaching, there are at least two possible explanations. First, when listing the topic with the greatest impact, the respondents were choosing from a finite set of topics which had been addressed in a ten-week course. When asked what topics should be covered, they could choose from any of these topics in addition to any other topics in the areas of education and psychology which they could possibly generate.

However, this first explanation does not address why the related topics of classroom management/discipline/ morality/and character education and motivation were at the bottom of the first list and at the top of the second. Therefore, another explanation is necessary, and it can be inferred from the following statement:

> There are a few things I do wish I knew more about like: How to better help students with special needs (abused emotionally). I want to be more effective with parents-- how do I get them to see the value of time spent with a child? Motivating the one or two kids that don't fit with all the others for one reason or another, behavior management strategies for those one or two who

> don't seem to care or be affected by other strategies. Does this help? How do I work with the little boy who actually thinks he's superman or the one who won't work unless he has *the* special silver crayon?

In the quotation, the respondent continually mentions the words "how to." These words suggest that the respondent needs procedural knowledge rather than declarative knowledge. This suggests that in-service teachers are as eager to practice skills as pre-service teachers are. In fact, in the 34 responses to this question, the words "how to" and "how do I" were mentioned 14 times. In addition, eight more respondents indicated that they would like assistance "dealing with" something. In spite of the fact that implications for the classroom and effective strategies were stressed in the present course, these teachers wanted even more practical information and practice in applying them. In fact, one respondent stated this explicitly when commenting on the group discussions: "As a teacher, I liked having real case scenarios like the final activity we did reading the 3 case studies."

Discussion

The present findings seem to indicate that the suggestions by Anderson et al. (1995), Shuell (1996), and VanZile-Tamsen (1997) for reforming courses for pre-service teachers may also apply to courses for practicing teachers. In the present study, many of the participants felt that the group discussions did help them learn a great deal. In addition, the focus of the course was on helping them figure out what they already knew and then figure out what they needed to learn within the framework of group activities. Although many of the students felt that this process was redundant over time and would prefer more applied activities, several students felt that the groups were important in helping them clarify their understanding of course content.

In spite of the fact that the majority of students found something worthwhile in the group discussions and they all listed topics that they felt impact teaching, the students still felt that the course should be more hands-on. This finding is consistent with findings from research involving pre-service teachers (VanZile-Tamsen, 1997). In spite of the fact that these practicing teachers are in a classroom all day, they indicated that they have difficulty applying course concepts to their own teaching without assistance from the instructor and practice. Making an effort to find implications for the classroom from articles and trying to generate strategies based on these readings was not enough. They wanted to practice strategies before applying them to their own classrooms. This finding suggests that instructors cannot assume that practicing teachers will be any more able to apply declarative knowledge to the classroom than pre-service teachers are. In fact, these in-service teachers seem to be indicating that they would appreciate more scaffolded, hands-on activities to help them hone their teaching skills.

Another interesting finding relates to the fact that the topics these practicing teachers felt should be included were similar to those identified by pre-service teachers (VanZile-Tamsen, 1997). Although many of the topics suggested are those taught at the undergraduate level, practicing teachers indicate that they need to know more. One explanation may relate to the fact that the information was first learned out of context in an undergraduate educational psychology lecture course removed from the every day experience of teaching (Anderson et al., 1995; Shuell, 1996). Another explanation may relate to the students' motivation/interest in the undergraduate course. Whatever the reasons for this finding, however, there are two important implications. First, graduate instructors cannot assume that graduate students have mastered basic concepts. Second, graduate instructors may need to structure class activities in such a way that students are assisted in applying these basic concepts to their own teaching through the use of authentic tasks and scaffolding.

A surprising aspect of the study related to the fact that, during whole-class discussions of reading materials, questions and comments that the instructor had not anticipated were often introduced. The result was an impromptu lecture or discussion on an issue of importance to the class which may have been missed had the instructor strictly followed lecture notes and kept students clearly focused on the specific topic at hand. This phenomenon is called enactivism by Davis and Sumara (1997). These authors suggest that enactivism is a step beyond constructivism. It occurs when students are encouraged to construct knowledge in social situations; through the process of social interaction with classmates and the instructor, new, unanticipated issues arise. In any teaching situation, instructors cannot determine what is learned. However, when group activities are used to encourage knowledge construction, this lack of instructor control is made very apparent. In addition, in a classroom that has been modeled after constructivist principles, the instructor may find it worthwhile to capitalize on the new issues and link them to the topic at hand, thus encouraging further knowledge construction.

Some caution should be used when attempting to generalize the findings of this study to other groups due to the nature of the present sample. First, the majority of these participants were White women. Research suggests that women tend to learn better in a collaborative setting and enjoy it more (Belenky, Clinchy, Goldberger, &

Tarule, 1986; Flynn, Savage, Penti, Brown, & Watke, 1991). Perceptions of the usefulness of group activities may have been very different had the sample included more men and been more ethnically diverse. In addition, all of these students were practicing teachers who entered the graduate program with strong beliefs about what constitutes teaching and learning (Fosnot, 1993a; Holt & Johnston, 1989). Those individuals who believe knowledge is contextual, who believe that learning can be facilitated by classmates, and who believe that the instructor is merely a facilitator of learning may respond more positively, and learn more, in a class like this than those who do not share these beliefs (King & Baxter Magolda, 1996).

In spite of limitations related to the sample, the findings of this study do indeed offer suggestions for the design of graduate educational psychology classes and for graduate education in general. Graduate students did find that group discussions helped to clarify their understanding of course concepts and promote meaningful learning. However, they also desire more authentic activities to assist them applying course concepts to their own teaching. The most challenging aspect for graduate instructors seems to be developing authentic activities for practicing teachers.

Shuell (1996) has suggested authentic activities for pre-service teachers; graduate instructors need to begin to develop and share authentic activities for practicing teachers. One way to begin may be by talking to teachers and asking them what types of activities will help them learn. Authentic learning activities can be developed that help teachers practice the skills they really need, thus making the graduate educational psychology course for practicing teachers more effective.

References

Anderson, L. M., Blumenfeld, P., Pintrich, P. R., Clark, C. M., Marx, R. W., & Peterson, P. (1995). Educational psychology for teachers: Reforming our courses, rethinking our roles. *Educational Psychologist, 30*(3), 143-157.

Baxter Magolda, M. B. (1998, November). *Learning-centered pedagogy: Bridging teaching and students' intellectual development.* Paper presented at the Board of Regents of the University System of Georgia Conference on Teaching and Learning, Atlanta, GA.

Belenky, M. F., Clinchy, B. M., Goldberger, N. R., & Tarule, J. M. (1986). *Women's ways of knowing: The development of self, voice, and mind.* New York: Basic Books, Inc.

Brown, J. S., Collins, & A., Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher, 18*(1), 32-42.

Covaleskie, J. F. (1992). Discipline and morality: Beyond rules and consequences. *The Educational Forum, 56*(2), 173-183.

Davis, B., & Sumara, D. J. (1997). Cognition, complexity, and teacher education. *Harvard Educational Review, 67*(1), 105-124.

Doyle, W., & Carter, K. (1996). Educational psychology and the education of teachers: A reaction. *Educational Psychologist, 31*(1), 23-28.

Flynn, E. A., Savage, G., Penti, M., Brown, C., & Watke, S. (1991). Gender and modes of collaboration in a chemical engineering design course. *Journal of Business and Technical Communication, 5*(4), 444-462.

Fosnot, C. T. (1993a). Learning to teach, teaching to learn: The Center for Constructivist Teaching-Teacher Preparation Project. *Teaching Education, 5*(2), 69-78.

Fosnot, C. T. (1993b). Rethinking science education: A defense of Piagetian constructivism. *Journal of Research in Science Teaching, 30*(9), 1189-1201.

Fosnot, C. T. (1996). Constructivism: A psychological theory of learning. In C. T. Fosnot (Ed.), *Constructivism. Theory, perspectives, and practice* (pp. 8-33). New York: Teachers College Press.

Gindis, B. (1995). The social/cultural implication of disability: Vygotsky's paradigm for special education. *Educational Psychologist, 30*(2), 77-81.

Holt, L., & Johnston, M. (1989). Graduate education and teachers' understandings: A collaborative case study of change. *Teaching and Teacher Education, 5*(2), 81-92.

Jost, K. L., Havard, B. C., & Smith, A. J. (1997, February). *A study of problem-based learning in a graduate education classroom.* Proceedings of Selected Research and Development Presentations at the 1997 National Convention of the Association for Educational Communications and Technology, Albuquerque, NM. (ERIC Document Reproduction Service No. ED 409 840).

King, P. M., & Baxter Magolda, M. B. (1996). A developmental perspective on learning. *Journal of College Student Development, 37*(2), 163-173.

Knowles, M. (1990). *The adult learner: A neglected species* (4th ed.). Houston, TX: Gulf Publishing Company.

LeFrancois, G. R. (1997). *Psychology for teachers* (9th ed.). Boston: Wadsworth Publishing Company.

Mayer, R. E. (1996). Learners as information processors: Legacies and limitations of educational psychology's second metaphor. *Educational Psychologist, 31*(3/4), 151-161.

McMillan, J. H., & Schumacher, S. (1997). *Research in education: A conceptual introduction* (4th ed.). New York: Longman.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis*. Thousand Oaks, CA: Sage Publications.

Moran, A. (1991). What can learning styles research learn from cognitive psychology? *Educational Psychology, 11*(3/4), 239-245.

Nelson, K. A. (1990). Gender communication through small groups. *English Journal, 79*(2), 58-61.

O'Banion, J. (1997). *A learning college for the 21st century*. Phoenix, AZ: The Oryx Press.

Ormrod, J. E. (1997). *Educational psychology: Developing learners* (2nd ed.). Upper Saddle River, NJ: Merrill.

Rogoff, B. (1990). *Apprenticeship in thinking*. New York: Oxford University Press.

Shuell, T. J. (1996). The role of educational psychology in the preparation of teachers. *Educational Psychologist, 31*(1), 5-14.

Tishman, S., Perkins, D. N., & Jay, E. (1995). *The thinking classroom*. Boston: Allyn & Bacon.

VanZile-Tamsen, C. (1997). Enhancing educational psychology for pre-service teachers. *GATEways to Teacher Education, 10*, 14-24.

von Glasersfeld, E. (1996). Introduction: Aspects of constructivism. In C. T. Fosnot (Ed.), *Constructivism. Theory, perspectives, and practice* (pp. 8-33). New York: Teachers College Press.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 2, 45-54

Breaking the Rote Memorization Mindset of Preservice Teachers Standards-Based Instruction: An Integrated Preservice Teacher Education Model

Regina Halpin
Mississippi State University

After the dissemination of the National Council of Teachers of Mathematics (NCTM) Curriculum and Evaluation Standards, revisions in teacher education programs have been focused on implementing interdisciplinary programs. The quantitative and qualitative data for this study were collected during two semesters of an interdisciplinary, field-based block of methods courses taken by preservice elementary teachers prior to student teaching. The purpose was to investigate the effect of a constructivist model on increasing teachers' self-efficacy in mathematics. Descriptions of the interdisciplinary model and lesson format are provided. A paired t test was conducted on pre- and post- data (N=73) and significant differences were found. Content analyses were conducted to categorize the preservice teachers' responses supporting their beliefs. Finally, recommended implications for teacher education programs are discussed.

The literature has shown that teachers' beliefs about what it means for students to apply mathematical concepts influence their approach of how mathematics should be taught (Cooney, 1985; Emenaker, 1996; Kloosterman & Stage, 1992; Schofield, 1981; Thompson, 1984). Ashton (1985) defined teacher efficacy as the belief teachers have in their ability to have a positive effect on the students' learning. Research has shown that teacher efficacy has a major influence on the learning environment provided for students (Ashton & Webb, 1986; Raymond, Santos, & Masingila, 1991; Thompson, 1984). A high sense of efficacy has contributed to higher student achievement (Ashton & Webb, 1986), classroom environments which are more responsive to students (Woolfolk, Rosoff, & Hoy, 1990), and more effective small group work (Gibson & Dembo, 1984; Wilson, 1996). The teachers became more like facilitators rather than dictators in the classroom (Ashton & Webb, 1986; NCTM, 1991) and were more willing to try new forms of innovative teaching techniques (Guskey, 1988). The teachers with low self-efficacy beliefs ignored incorrect responses from students (Gibson & Dembo, 1984) and managed the classroom using more rigid and controlling methods (Karp, 1991; Woolfolk et al., 1990). According to the direction for change in learning during the past 10 years, educators are expected to meet the goals and expectations of a society concerned with the mathematical literacy of children as supported in the following quote from the National Council of Teachers of Mathematics Curriculum and Evaluation Standards (NCTM, 1989):

> [T]oday's society expects schools to insure that all students have an opportunity to become mathematically literate, are capable of extending their learning, have an equal opportunity to learn, and become informed citizens capable of understanding issues in a technological society. As society changes, so must its schools. (p. 5)

Therefore, in response to this public demand for reform in the teaching of mathematics (Owen & Lamb, 1996), it is important for preservice teacher education instructors to understand teachers' beliefs toward mathematics in order to modify the preservice instruction as recommended by the National Council of Teachers of Mathematics Curriculum and Evaluation Standards for Teaching Mathematics (1989) and the Professional Standards for Teaching Mathematics (1991) for Grades K-12.

The educational goals for students supported by NCTM reflect the importance of teaching students in Grades K-12 to value mathematics, become confident in their ability to do mathematics, become mathematical problem solvers, and learn to communicate and reason mathematically (NCTM, 1989, p. 5). Unfortunately, these goals for students learning mathematics are opposite of how most preservice teachers were taught in grade school. Most preservice teachers in today's teacher education programs were taught mathematics in a classroom where the driving force was rote memorization and the teacher's main task was to give specific explanations on how always to get the right answer using

Regina Halpin is an Associate Professor in the Department of Curriculum and Instruction at Mississippi State University. Correspondence relating to this article should be directed to Regina Halpin, Department of Curriculum and Instruction, Box 9705, Mississippi State, MS 39762 or by e-mail at rhalpin@ra.msstate.edu.

a specific formula (Owen & Lamb, 1996; Raymond & Santos, 1995). Unless a change occurs in teacher education programs, today's preservice teachers will become tomorrow's mathematics teachers teaching as they were taught. The cycle must be broken.

In the few years the NCTM standards have been in place, many new teaching philosophies have been developed and some, such as constructivism, have been reinstated in an effort to reform teacher education programs across the nation. The constructivist theory has become a major component of the reform in mathematics education (Owen & Lamb, 1996). Even though educators in preservice education programs may vary on a definition for constructivism, the commonality has been that students must learn to apply mathematics through exploration and discovery and take ownership for their learning. The knowledge must be constructed actively by the student using manipulatives physically and mentally by communicating their constructed ideas. Schifter and Fosnot (1993) supported the theory that mathematical understanding involves interpretive and constructive activity. According to Clements and Battista (1990), a teacher of the constructivist mold will not believe that the students will become problem solvers by passively receiving information. Instead, the students must be involved actively. Therefore, it is pertinent to evaluate the effectiveness of education programs that allow preservice teachers to experience constructing their own ideas and teaching mathematics based on a constructivist philosophy.

Purpose of Study

This study was timely because research supports that a constructivist-based classroom provides students an opportunity to explore and discover the meaning of mathematics. Even though the NCTM Standards (1991, 1989) have opened doors to new ideas in teacher education, change takes time. Therefore, the current debate does not concern the importance of the approach to teaching mathematics. Instead, the current issue to be investigated in teacher education is finding an effective method to prepare preservice teachers to teach mathematics and develop the mindset they need to transfer the constructivist theory into the classroom. Preservice education programs must provide a cutting-edge model for new teachers who will be entering school systems that may or may not be currently implementing instruction based on a constructivist point of view. Thus, it has become important to evaluate revised teacher education programs to determine whether new approaches are effective in producing elementary teachers who understand the constructivist theory and are confident in their ability to apply the theory in the teaching of mathematics and becoming national certified teachers (NCTM, 1991; Owen & Lamb, 1996).

This study was designed to assess how preservice elementary teachers' beliefs in their ability to teach elementary mathematics were influenced by participating in an interdisciplinary, constructivist-based teacher education model based on the NCTM Curriculum and Evaluation Standards (1989) and the Professional Teaching Standards (1991). If taken individually, the components of the model described in this study were not new. However, it was the combination of the components that was being evaluated. It was recognized that students' confidence in their ability to teach mathematics could increase even in a traditional methods course. Therefore, this study focused on evaluating the effectiveness of an interdisciplinary methods course designed to develop students' belief in their ability to teach mathematics as supported by the NCTM Professional Standards (1991), "In order to change our perspective about mathematics teaching and learning, we need direction on how mathematics can be taught and learned to enhance the development of mathematical power" (p. 2). The program under investigation was an interdisciplinary, constructivist approach for training preservice elementary teachers to meet this challenge. This senior experience was required for all elementary majors prior to student teaching. The emphasis for this study was on the elementary mathematics methods portion of the interdisciplinary experience referred to hereafter as the Senior Block and whether there would be a difference between the preservice teachers' pre- and post mean ratings of their belief in their ability to teach mathematics. Additional research questions guiding this study were:

1. What percentage of the elementary preservice teachers will consider mathematics their favorite subject to teach prior to and after the semester?
2. Which of the NCTM Goals for Students (NCTM, 1989, p. 5) do elementary preservice teachers feel to be most important as their first goal to meet as a new teacher involved in an interdisciplinary curriculum?

Description of the Interdisciplinary Senior Block

Three years ago, the elementary teacher education program reviewed in this study was revised to include an interdisciplinary block of methods instruction complemented by a field-based component. Traditionally, the four method courses were taught separately and the instructional model encouraged the attitude that the four subjects were isolated topics in the classroom. Therefore, the primary revision of the teacher education program was to design a 12 hour credit block of mathematics,

science, language arts, and social studies method courses taught simultaneously as an integrated curriculum during the same semester. The goal was to make major shifts in the teacher education program "to move from current practice to mathematics teaching for the empowerment of [the] students" (NCTM, 1991, p. 3). The following major shifts that are woven throughout the NCTM Professional Teaching Standards (1991) were the foundational support for the teacher education program evaluated in this study:

- Toward classrooms as mathematical communities and away from classrooms as simply a collection of individuals;
- Toward logic and mathematical evidence as verification and away from the teacher as the sole authority for right answers;
- Toward mathematical reasoning and away from merely memorizing procedures;
- Toward conjecturing, inventing, and problem-solving and away from an emphasis on mechanistic answer-finding;
- Toward connecting mathematics, its ideas, and applications and away from treating mathematics as a body of isolated concepts and procedures.

The Senior Block was a 5-day, 4-hours per day program. Four university faculty members were responsible for providing an interdisciplinary model of instruction and supervising the preservice teachers' field-based experiences in the schools. Sixty percent of the semester was devoted to the integrated methods, classroom management, and evaluation instruction. The remaining 40% was devoted to a field-based experience under the supervision of a certified elementary teacher and a university methods instructor.

The university faculty worked together to provide interdisciplinary instruction to the preservice teachers to prepare them for the field-based experience. For example, the science and mathematics faculty worked together to model interdisciplinary teaching by providing instruction and experiences on conducting experimental and observational experiments, collecting the data, and graphing the data using spreadsheets. The mathematics and language arts instructors modeled in the classroom how children's literature could be used to supplement the teaching of many mathematics concepts. Social studies topics such as economics and geography provided an ideal opportunity to demonstrate how mathematics could be used in all aspects of life. Simultaneously, instruction on classroom management and evaluation was incorporated into each of the four discipline areas.

The preservice teachers selected and created their own groups among all of their classmates. Therefore, everyone was a member of a group of three to four during the entire semester for the field-based component. During the weekly lesson planning, the university faculty assisted the groups of preservice teachers with writing behavioral objectives and developing interdisciplinary lessons as modeled during the methods instruction. The lesson planning was an important feature of the Senior Block, because the preservice teachers learned from experience how to teach lessons from an interdisciplinary approach in a classroom of children. Therefore, if a preservice teacher preferred to teach language arts or science, an interdisciplinary approach could prevent future occurrences of mathematics being taught as an isolated topic focused on rote memorization. For consistency and as a learning guide, the lesson plan format shown in Figure 1 was provided for the preservice teachers.

After the preservice teachers received the university faculty's suggestions and ultimately approval for a lesson, the groups returned to their assigned classrooms to work in small groups with the students under the supervision of the cooperating classroom teacher. Each group taught their planned lessons using a combination of small and large group instruction and worked with the students on an individual basis in a self-contained classroom. Therefore, the interdisciplinary methods instruction provided by the university faculty was immediately applied by the preservice teachers in the classroom. The culminating event for the preservice teachers was the planning and teaching of a one-week thematic unit on a topic of their choice that had not been taught that year by their supervising teacher.

Description of the Integrated Mathematics Methods Component

The focus of the mathematics method instruction in the Senior Block was a constructivist philosophy. The preservice teachers were involved actively in using manipulatives to discover the meaning of mathematics. During the first day of instruction, the four NCTM themes, communication, connections, problem solving, and reasoning (NCTM, 1989, p. 11) were introduced to the preservice teachers through hands-on activities. Secondly, the pre-service teachers became familiar with using manipulatives such as linking cubes, pattern blocks, Cuisenaire rods, and base ten blocks by participating in several activities designed to be used in the elementary classroom to teach mathematics. Finally, the preservice teachers divided into small groups to discuss how the four NCTM themes had been addressed in each activity. As a result, the preservice teachers were not taught about the NCTM Curriculum and Evaluation Standards but instead,

were provided a model of how to implement each standard based on the constructivist philosophy. For the remainder of the semester, the preservice teachers were involved actively in using manipulatives to construct their own understanding of mathematics concepts that had, in the past, been only a collection of rules. Furthermore, the preservice teachers had many opportunities to implement the NCTM Standards through field-based teaching experiences and assigned projects. The purpose of the field-based component and assigned projects was to challenge the preservice teachers to develop their own understanding of how to teach based on the instruction modeled during their teacher education program.

Grade/Subject:
Lesson Title:
Interdisciplinary Concepts:
    Language Arts:
    Math:
    Science:
    Social Studies:
Classroom Teacher Signature:
University Supervisor Signature:

PURPOSE OF LESSON:

LESSON OBJECTIVES:
1.
2.
3.

MATERIALS NEEDED: (# of students: _____)

PRE-INSTRUCTIONAL TECHNIQUES: (*ESTIMATE TIME NEEDED IN PARENTHESES*)
    Statement Of Review: ( )
    Motivating Technique: ( )
    Other: ( )
INSTRUCTIONAL PROCEDURES AND STRATEGIES: ( )
Sequence your instructional procedures to include *Directed Teaching, Guided Practice, Independent Practice, Corrective Activities, Extension Activities*

EVALUATION: ( )
    Summative Evaluation:
    Formative Evaluation:

QUESTIONS TEACHER WILL ASK STUDENTS:

QUESTIONS TEACHER ANTICIPATES FROM STUDENTS:

DISCUSSION OF APPLICATION OF NCTM STANDARDS:

BIBLIOGRAPHY:

Figure 1. Lesson Plan Format

Three examples of projects the students completed during the two semesters of this study included interdisciplinary mathematics lessons, article critiques, and a teacher-made manipulative designed to address a learning problem the preservice teacher had observed in the mathematics classroom or thought might occur in the classroom. All lesson plans consisted of the same components described earlier in addition to one or two paragraphs describing specifically how the lesson was exemplary for providing a learning environment that enforced problem solving, reasoning, making connections, and communicating mathematically. After teaching the lesson, the preservice teachers reflected on how well the lesson was executed as planned and wrote a summary of how to improve the lesson. The reflections included a summary of the students' questions and participation, accuracy of estimated times, and unexpected interruptions.

Periodically during the semester, articles published in *Teaching Children Mathematics*, a NCTM journal, were assigned. After reading each article, the preservice teachers reflected on the teaching techniques described in the article including how the NCTM Curriculum and Evaluation Standards (1989) were applied and their ideas of how they could expand on the teaching techniques presented as supported by the NCTM Professional Teaching Standards (1991). They expressed their opinions in a one-page article critique and a whole group discussion among the preservice teachers was initiated during the following class period. The discussion focused on their ideas of how the teaching methods presented in the article could be modified to accommodate different learning styles at various grade levels.

When the preservice teachers enter the workforce, the mathematics manipulatives used during their teacher training program may not be available in the classroom. Therefore, it was important for teachers to think creatively and thoroughly as to how they would improvise in a school with limited resources and still be able to provide every student with a problem solving learning environment focused on exploration and discovery. Each preservice teacher chose a mathematics concept and developed a teacher-made manipulative to be used with small groups of students. The project included a one page discussion of how the manipulative could be used for promoting problem solving in the mathematics classroom. The NCTM Curriculum and Evaluation Standards (1989) were used as a guide for determining what should be taught at the different grade levels and what should be expected of children in various grades. At the end of the semester, each preservice teacher provided copies of their lesson plans and project descriptions to exchange with their peers so everyone was able to create their own mathematics methods resource notebook.

Method

Participants

Eighty-eight preservice elementary teachers were enrolled in the Senior Block during two semesters, but because 15 were not present when one of the two questionnaires was administered, the sample size was 73. The participant pool consisted of 67 females and 6 males ranging in age from 20 to 45 with a mean of 23. The racial breakdown of the sample was 4 African Americans, 67 Caucasians, and 2 Hispanics. The same instructor taught both semesters, and the same teaching structure was implemented.

Questionnaire

This study of the teacher education model was conducted during two consecutive semesters. The questionnaire given in Figure 2 was administered at the start and finish of each semester. The questions were designed to gather responses from the preservice teachers on their current knowledge of the NCTM Standards, the subject they felt the most confidence in teaching, and a rating on a scale of 1-5 of their perceived ability to teach mathematics. The end of the semester questionnaire included an additional question: Which of the Goals for Students as stated in the NCTM Curriculum and Evaluation Standards (1989, p.5) do you feel to be the most important challenge for a first year teacher in an integrated curriculum, and how will you meet this goal? To assess whether the preservice teachers chose to integrate mathematics with other subjects, a descriptive analysis was conducted on the topics of their one-week thematic units and how many chose to include mathematical concepts.

Using the test-retest procedure with equivalent sample sizes, the reliability of the questionnaire was .82. The validity of the questionnaire was addressed by asking two colleagues to review and edit the questions prior to the study. As a result, minor changes were made in the wording of the questions, but both colleagues were consistent in their interpretation of the purpose of the questionnaire. The questionnaire was also administered to a sample of five teachers and follow-up interviews were conducted to determine how long it took each teacher to answer the questions, if all of the questions were understood, and whether the teachers' interpretations of the purpose of the instrument were consistent. All five teachers were consistent in determining the purpose of the questionnaire and did not suggest any changes to the wording of the questions.

YOUR NAME _____

LAST MATH CLASS TAKEN AND WHEN _____

What is the National Council of Teachers of Mathematics (NCTM)?

Have you used the NCTM Curriculum and Evaluation Standards and Professional Teaching Standards? If so, how?

Of all of the subjects you will teach in the elementary classroom, which one do you feel most comfortable teaching at this point in your career? Why?

ON A SCALE OF 1 (LOW ABILITY) TO 5 (HIGH ABILITY)

How would you rate your ability to teach elementary mathematics based on the constructivist theory (hands-on approach) for teaching?

For the Post-questionnaire:

Which of the "Goals for Students" as stated in the NCTM Standards do you feel to be the most important challenge for a first year teacher in an integrated curriculum?

How will you meet the goal you have chosen?

Post-question sent via electronic mail:

How would you teach math in order to meet the goal of "building students' confidence in their ability to do mathematics?"

Figure 2. Pre-Post Questionnaire

Results and Discussion

A combination of quantitative and qualitative analyses were conducted using the data obtained from the questionnaire, informal interviews, and the students' thematic units. At the beginning of the semesters, all of the 73 preservice teachers responded that they did not know about NCTM or the NCTM Standards (1989, 1991). Furthermore, only 5 of the 73 responded that mathematics was the subject they had the most confidence in teaching. Science was reported most often, because there were always experiments for the students to conduct. Their second choice was language arts, because it involved children's literature. Through informal interviews, most of the preservice teachers indicated they had a bad experience in an elementary school mathematics class and did not elect to take higher mathematics courses in high school. Basically, mathematics was hard and involved working with numbers; no hands-on experiments like in science. These responses from the teachers indicated that the preservice teachers' mindset was based on the traditional teaching of mathematics where the constructivist theory was not applied. A paired sample t test was used to determine if there was a significant difference in the preservice elementary teachers' reported pre- and post-belief ratings in their ability to teach mathematics before and after participating in the Senior Block. On a scale of 1-5 (1 being lowest), the average rating of the preservice teachers' belief in their ability to teach mathematics before the Senior Block was 2.86 with a standard deviation of 1.11, as compared to a post average of 4.01 and .69 standard deviation. The results indicated there was a significant difference between the preservice teachers' pre and post mean ratings of their belief in their ability to teach mathematics with a t value of 8.98 ($p < .001$).

After completing the Senior Block, all the pre-service teachers could knowledgeably respond to the question concerning which NCTM student goal was most important for a new teacher to strive for first and why they supported their decision. The most encouraging results

were that the number of preservice teachers responding they now had the most confidence teaching mathematics based on a constructivist theory (hands-on) increased from 5 to 41 (56% of the sample). To further support these results, the preservice teachers' thematic units were examined. Of the 24 thematic units, none of the preservice teachers chose a mathematical concept as their topic. However, of the 24 thematic units, 22 (92%) had a mathematical concept (e.g. fractions, adding, subtracting, rounding, proportions) integrated into at least three of the five days' lessons. Because of the amount of time the preservice teachers were in the classroom, these lessons were taught in addition to the math lesson for the day. Therefore, the preservice teachers chose to further enhance the students' mathematical skills by integrating the concepts into other subject areas.

Descriptive analyses were used to summarize the percentage of preservice teachers choosing the student goal they felt to be the most important challenge as a new teacher in the mathematics classroom. These results are provided in Table 1. Because the purpose of this study was to investigate how preservice teachers' belief in their teaching ability was affected by the implementation of the NCTM Standards, it was encouraging that 53% of the preservice teachers responded it was most important for students to become confident in their ability to do mathematics. These results, along with the informal interviews, indicated that what the teachers believed to be true for their students was based on their own mindset after being shown how to teach mathematics based on the NCTM Standards. The preservice teachers' responses reflected a renewed and more confident attitude toward their ability to teach mathematics, because they had experienced their own understanding of the mathematical methods and concepts as opposed to being shown only how to teach the concepts and encouraged to memorize the procedures. Through informal interviews, there were numerous times the preservice teachers would make the statement that for the first time, they understood the mathematical concepts they were expected to teach. Another common statement reflected their recognition that they had not been taught mathematics this way during elementary school, but their mindset had definitely changed after completing the Senior Block.

A content analysis was used to describe the 39 preservice teachers' reasons why they felt developing a student's self-confidence in mathematics was the most important challenge. The results are provided in Table 2. It seems reasonable that the other four goals were selected less often considering 19 of the 39 preservice teachers choosing confidence in mathematics supported their choice by stating that the first step toward meeting the remaining four goals was for students to have confidence in their ability to do mathematics.

Table 1
Preservice Teachers' Selected Student Goal

Student Goals	Frequency (%)
Value mathematics	11 (15%)
Become confident in ability	39 (53%)
Become problem solvers	13 (18%)
Communicate mathematically	0 (0%)
Reason mathematically	10 (14%)
Total	73 (100%)

Table 2
Content Analysis of Why Preservice Teachers Think Student Confidence in Mathematics is Important

Categories	Frequency (%)
First step toward meeting all student goals	19 (49%)
Mathematics will be more fun and interesting	5 (13%)
Mathematics is important in all subjects	11 (28%)
Students will explore and discover using manipulatives	4 (10%)

When the preservice teachers were asked how they would accomplish their selected goal for their students, all of the responses could be grouped into one of the following categories:

1. use manipulatives and fewer worksheets,
2. provide more nonroutine problems and fewer problems encouraging rote memorization, or
3. alternative assessment techniques (e.g. portfolios, rubrics) de-emphasizing assigned grades based on tests.

At the end of the semester, an electronic mail message was sent to all of the preservice teachers asking how they would teach mathematics in an effort to meet the goal of building students' confidence in their ability to do mathematics. The following are explanations from three of the preservice teachers:

Preservice teacher #1:

> . . . I will set up a classroom that is full of manipulatives so children are able to explore. I would allow children to work in groups so they can learn from their peers. I do not want my

students spending all of their time doing math fact worksheets. I believe if teachers would get away from using worksheets all of the time and marking answers only right or wrong, students would become more confident in math.

Preservice teacher #2:

It is important for today's teachers to provide an atmosphere in which confidence can develop in every child. The child will discover patterns with manipulatives, make their own sense of things, and invent new procedures if they have confidence in their math ability or the feeling 'I can do math.' I will give math problems that have too much or not enough information [nonroutine problems].

Preservice teacher #3:

I believe the best way to boost a student's self confidence in the classroom is to allow them to work at their own pace with less pressure from the teacher. I like the idea of using rubrics and portfolios as assessment techniques to evaluate the students' work. I remember when I was in school and we had timed tests. I never did well on the timed tests because I felt too much pressure from the teacher. I am going to avoid quizzes and tests that might make my students stressed out or cause them to feel incompetent.

The focus of all of the preservice teachers' responses indicated clearly that their beliefs were supported by the NCTM Standards. Their responses supported the quantitative results that there was an increase in their self-efficacy, and they were prepared to transfer their standard-based beliefs into their own teaching based on how they were trained. To provide more insight into how the preservice teachers felt their self-efficacy was developed, their responses to how they would teach mathematics in order to build students' confidence were analyzed and categorized as follows (more than one response was given in some cases):

1. Doing mathematics is learning mathematics (77%).
2. Assessment of students includes writing and explaining; it was more than receiving grades for right and wrong answers (63%).
3. Using mathematics in other subjects makes teaching the concepts more fun than just memorizing numbers and formulas (57%).
4. Students should lead the classroom discussion instead of the teacher doing all of the talking (45%).
5. Exploring and discovering is more meaningful ("I finally understood some mathematics after teaching it this way.") (44%).

These results were supportive of an increase in the preservice teachers' self-efficacy for teaching mathematics as outlined in the NCTM Professional Teaching Standards (1991). Furthermore, the preservice teachers understood through their own experiences, the importance of providing students with a learning environment that enhances their ability to develop mathematical power.

Conclusions and Recommendations

The results of this study suggest preservice teachers' lack of belief in their ability to teach mathematics is a pre-existing condition. It is expected that preservice teachers would enter a methods course unprepared to teach the content and exit the course better prepared. Unfortunately, as supported by these results, preservice teachers enter teacher education lacking more than the skills to teach mathematics. Specifically, the preservice teachers participating in this study had more negative beliefs toward their ability to teach mathematics than toward any of the other subjects. However, an interdisciplinary, constructivist teacher education model can contribute positively to an increase in preservice teachers' self-efficacy. The research has supported the importance of providing preservice teachers the opportunity to learn for themselves how to teach and communicate mathematics with meaning and not just numbers. The underlying challenge for teacher education programs is now to combine different components into an effective teaching model for mathematics education in conjunction with NCTM's recommendations for the K-12 curriculum.

The findings of this study support the effectiveness of a model containing several components combined in an effort to positively influence preservice teachers' belief in their ability to teach mathematics. Based on these findings, four recommendations are made:

1. Initial preservice instruction should include modeling of the NCTM Curriculum and Evaluation Standards through applications using various manipulatives.
2. Preservice teachers' projects and assignments should include reflective thinking on their previous and present experiences relating to teaching methods and their philosophy of teaching mathematics.
3. Instructors should team-teach to model the importance of an interdisciplinary approach to teaching mathematics.
4. Preservice teachers need time in the classroom as a supplement during the methods courses to serve as a means of closure for the entire experience.

The NCTM Curriculum and Evaluation Standards (1989) and Professional Teaching Standards (1991) should be understood by the preservice teachers through application of various teaching methods and the use of different manipulatives. If the rote memorization mindset is to be broken, we must teach preservice teachers as we would like them to teach in their own classroom. Therefore, preservice teachers must be involved actively in their own learning of the content, as they will expect their students to be involved in their own learning process. To complement the methods instruction, projects and assignments should require preservice teachers to reflect on their teaching strategies and their philosophy of teaching mathematics from an approach other than a lecture-only approach. As observed in this study, it is not enough for the preservice teachers to simply study about applying a collection of teaching methods without thinking through the entire process. For example, even though preservice teachers are not accustomed to writing in the mathematics classroom and it is difficult for them at first, they are experiencing first-hand the meaning of communication as supported by the NCTM standards.

At this point in their career, most preservice teachers have not clearly defined their philosophy on the most effective approach to teaching mathematics. One could speculate that preservice teachers' belief in their ability to teach mathematics would be reflected in their philosophy of how mathematics should be taught based on their own experiences. Therefore, it is important for the teaching of elementary mathematics to be modeled in an interdisciplinary environment as supported by a constructivist approach; not a lecturing approach on the innovative teaching strategies we want preservice teachers to use. Many preservice teachers enter the teacher education program lacking the confidence for teaching mathematics. By incorporating the mathematics methods with other disciplines in which the preservice teachers feel more comfortable, they begin to understand how mathematics can be taught effectively with meaning. As a result, the preservice teacher no longer feels the need to be the focal point of the classroom; the students became the "doers" of mathematics because they are exploring and discovering the meaning of mathematics as it related to their daily lives.

Finally, the applications should be followed by actual classroom experiences until the preservice teachers' beliefs are firmly in place. That is why it is important to provide closure for the entire experience; in this case, the field-based experience is the closure component to the program. Most preservice teachers are nervous, inexperienced, and lack confidence when they enter the classroom for the first time. The sooner they are responsible for planning lessons, handling classroom management situations, and teaching mathematics, the better prepared they will become.

After participating in an interdisciplinary, constructivist program as described in this study, the preservice teachers have experience in the classroom applying the new knowledge they have acquired under the supervision of school and university faculty without being fully responsible for the entire classroom. By allowing the preservice teacher to work in a small group, everyone is provided additional support during the learning process. Consequently, the preservice teachers' belief in their ability to teach mathematics will hopefully increase due to the success being experienced in the classroom. One would hope that by the time the preservice teachers enter student teaching, they will have developed confidence in developing their own philosophy for teaching mathematics based on how they were taught during their teacher education program. The results of this study indicated a change in the preservice teachers' self efficacy did occur in an interdisciplinary environment and the combination of the components that were being evaluated proved to be effective. However, the recommendations provided should be applied to similar programs, and future research should be conducted to determine if the preservice teachers transfer a positive attitude toward teaching mathematics regardless of the structure of the methods instruction, and whether gender differences are a factor. More importantly, will the preservice teachers transfer a new teaching belief into student teaching and their first year of teaching?

References

Ashton, P. T. (1985). Motivation and the teacher's sense of efficacy. In C. Ames & R. Ames (Eds.), *Research on Motivation in Education, 2*, 141-171. Orlando, FL: Academic Press.

Ashton, P. T., & Webb, R. B. (1986). *Making a difference: Teachers' sense of efficacy and student achievement.* New York: Longman.

Clements, D. H., & Battista, M. T. (1990). Constructivist learning and teaching. *Arithmetic Teacher, 38*(1), 34-35.

Cooney, T. J. (1985). A beginning teacher's view of problem solving. *Journal of Research in Mathematics Education, 16*, 324-336.

Emenaker, C. (1996). A problem-solving based mathematics course and elementary teachers' belief. *School Science and Mathematics, 96*(2), 75-84.

Gibson, S., & Dembo, M. (1984). Teacher efficacy: A construct validation. *Journal of Educational Psychology, 76*, 569-582.

Guskey, T. R. (1988). Teacher efficacy, self-concept, and attitudes toward implementation of instructional innovation. *Teaching and Teacher Education, 4*, 63-69.

Karp, K. S. (1991). Elementary school teachers' attitude toward mathematics: The impact on students' autonomous learning skills. *School Science and Mathematics, 91*(6), 265-270.

Kloosterman, P., & Stage, F. K. (1992). Measuring beliefs about mathematical problem solving. *School Science and Mathematics, 92*(3), 109-115.

National Council of Teachers of Mathematics Curriculum and Evaluation Standards for School Mathematics (1989). NCTM: Reston, Va.

National Council of Teachers of Mathematics Professional Standards for Teaching Mathematics (1991). NCTM: Reston, Va.

Owen, L. B., & Lamb, C. E. (1996). *Bringing NCTM standards to life: Best practices from elementary educators.* Eye on Education. Princeton, N.J.

Raymond, A., & Santos, V. (1995). Preservice elementary teachers and self-reflection: How innovation in mathematics teacher preparation challenges mathematics beliefs. *Journal of Teacher Education, 40*(1), 58-70.

Raymond, A., Santos, V., & Masingila, J. (1991). The influence of innovative instructional process on mathematical belief systems. Paper presented at the annual meeting of the American Educational Research Association, Chicago, Illinois.

Schofield, H. (1981). Teacher effects on cognitive and affective pupil outcomes in elementary school mathematics. *Journal of Educational Psychology, 73*, 462-471.

Schifter, D., & Fosnot, C. (1993). *Reconstructing mathematics education.* New York: Teachers College Press.

Thompson, A. G. (1984). The relationship of teachers' conceptions of mathematics and mathematics teaching to instructional practice. *Educational Studies in Mathematics, 15*, 105-127.

Wilson, J. D. (1996). An evaluation of the field experiences of the innovative model for the preparation of elementary teachers for science, mathematics, and technology. *Journal of Teacher Education, 47*(1), 53-59.

Woolfolk, A. E., Rosoff, B., & Hoy, W. K. (1990). Teachers' sense of efficacy and their beliefs about managing students. *Teaching and Teacher Education, 6*, 137-148.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 2, 55-62

Evaluation of Teacher Preparation Graduates by NCATE Accredited Institutions: Techniques Used and Barriers

J. Jackson Barnette
The University of Iowa

Keith Gorham
Scottsboro (AL) City Schools

A survey, designed to assess the use and perceptions of effectiveness of techniques used to evaluate graduates of NCATE accredited teacher preparation programs, was mailed to a random sample of NCATE accredited institutions. The return rate was 74%. The five most used techniques were: questionnaires, teacher tests, principal/supervisor evaluations, exit comprehensive exams, and longitudinal studies. The five most highly rated methods were: principal/supervisor evaluation, observation by independent observer, questionnaire, exit comprehensive exam, and institutional review board at program completion. NCATE accredited institutions are using a variety of mostly traditional but some emerging techniques. It is clear that most respondents thought technical and methodological resources existed for conducting evaluations, but financial and time resources are lacking. Methods used to track program graduates were not highly rated relative to effectiveness. Those methods most closely under direct control of the preparation program were rated as most effective.

In 1987, the National Council for Accreditation of Teacher Education (NCATE) standards included the following criterion:

> The unit keeps abreast of emerging evaluation techniques and engages in regular and systematic evaluations, including follow-up studies, to determine the success and quality of graduates in the professional roles for which they were prepared. (The National Council for Accreditation of Teacher Education, 1987, p. 8)

In the 1995 NCATE standards (The National Council for Accreditation of Teacher Education, 1995), this criterion was separated into at least two standards:

> The unit engages in regular and systematic evaluations (including, but not limited to, information obtained through student assessment, and collection of data from students, recent graduates, and other members of the professional community) and uses these results to foster student achievement through the modification and improvement of the conceptual framework(s) and programs" (p. 16)

and "The unit ensures that a candidate's competency to begin his or her professional role in schools is assessed prior to completion of the program and/or recommendation for licensure" (p. 23).

Whether one considers the 1987 criterion or 1995 NCATE standards, evaluation of program graduates at completion and inservice is clearly a very important activity. However, historically, few methods have been used extensively other than follow-up surveys to graduates and their employers. It was clear that NCATE wanted to see more varied methods and sources of data than had been traditionally used for program graduate evaluation. As Kunkel (1988, p. 54) pointed out, "Continued use of program evaluation practices, techniques and findings are an important and an integral part of the redesign of the NCATE Standards of the accreditation of programs for the approval of teacher education programs."

Institutions of higher education must be aware of, and accountable for, the success and quality of graduates of their educational training programs (Ayers, 1989a). The extent of the success and quality of graduates can provide much insight into the effectiveness of an institution's educational training program. In judging the value of educational training programs, the most compelling information results from determining what the graduates are able to accomplish (Peterson, 1989).

Effective educational program planning and implementation presume the presence of effective program evaluation (Lindsay, 1985). He stated that if teacher education faculty believe in the existence of detailed knowledge and instructional strategies necessary for

Dr. Jack Barnette, Associate Dean and Associate Professor, College of Public Health, The University of Iowa, Iowa City, IA 52242, (319)335-8905, (319)335-6635 (FAX), email: jack-barnette@uiowa.edu. Dr. Keith Gorham, Principal, Scottsboro Junior High School, Scottsboro, AL 35768, (256)259-1204

competent performance of teaching, then they have a responsibility to gather evidence on the extent to which the knowledge and strategies are retained and implemented by graduates of educational training programs.

Demands for teacher education program accountability from both the parent institution and NCATE provide significant motivation for maintaining relationships with graduates to determine their success and quality. Lindsay (1985) suggested that systematic evaluation of graduates and the resulting data could be used for revision and improvement, as well as confirmation and justification, of programs in teacher education. Kunkel (1988, pp. 50-51) stated, "Quality evaluation of professional education units requires attention to three items. Effective evaluation of a unit must be holistic, deal with soft and hard data, and be continually future oriented."

In theory, the success of a program of teacher education would be determined best by ascertaining the teaching ability of graduates (Ayers, 1988). Although the rationale for determining program effectiveness by ascertaining the success and quality of graduates through systematic evaluation is evident, conducting such evaluations can prove difficult (Lindsay, 1985).

The most dominant current evaluation technique is the follow-up survey (Boser, 1988; Holste & Matthews, 1993; Murphy, 1992). Clark and Nichols (1983) presented methods of increasing the precision of estimates in follow-up surveys. Evaluation techniques that NCATE institutions should employ to determine the success and quality of graduates could include follow-up studies. However, the 1987 NCATE criterion also stated that institutions should keep abreast of emerging evaluation techniques other than follow-up studies and institutions were expected to employ several evaluation techniques to determine training program effectiveness.

Bloom and Bacon (1995) reported the use of portfolio assessment to enable graduate students in behavioral disorders to document their creativity and skills. Some programs (Ayers, 1988; Warren 1990) have conducted longitudinal studies. McEneaney and Sheridan (1993) reported on the use of telephone interviews of recent program graduates. Nuccio (1990) reported on the use of a questionnaire to compare perceptions of three groups (current teachers, former teachers, and those persons who never taught) relative to perceptions of the effectiveness of their preparation. Ayers (1989b) described the use of the NTE and the PPST for predicting classroom performance. Peterson (1989) proposed the use of institutional review boards and Redfield (1988) discussed the use of pupil achievement in the evaluation of any educational endeavor.

Bond (1990) reported that, in a survey of 47 NCATE accredited institutions, questionnaires were used by 63% of the respondents, but many respondents thought data derived from questionnaires were more likely to be considered less useful than data from other sources such as interviews, visits, or telephone contacts.

The research reported here is based on results from a mailed survey focusing on answering four general research questions:

1. What techniques, singularly and in combination, have been used by NCATE accredited institutions for evaluation of graduates?
2. What are the perceptions of their use and effectiveness (ease of conducting, cost effectiveness, helpfulness in meeting NCATE standards, helpfulness in program improvement, and likelihood of being used in the future)?
3. What are the perceptions of barriers to conducting such evaluations?
4. What methods are used to track program graduates and perceptions of their effectiveness?

Methods

Based on the 1993 AACTE directory, a random sample of 145 NCATE accredited institutions was selected, which represented slightly more than 25% of the total number of NCATE accredited institutions. Survey forms were mailed to the deans/directors of education of these institutions. Of these, 107 were returned in time to be included in the analysis, a return rate of 74%. Approximately 20% of the NCATE accredited programs in the nation were represented. SAS® (SAS Institute) programs were used to analyze the data. These programs provided frequency counts and summary statistics by item and for a total score for each technique.

The survey was divided into four sections. The first section presented 13 different evaluation techniques where respondents were asked whether the techniques had been used and, if they had, how they were perceived. These techniques were selected after being found in the review of the literature. Each evaluation technique found in the literature review was presented as a questionnaire item. The technique of institutional review board was divided to include a board review at completion of the program and a review completed post program. Though not found in the literature, news/activity file was included in the survey. This inclusion was made after professional conversation regarding additional techniques not found in the literature. Following is the final set of techniques, with their definitions, used in the survey:

1. *Questionnaire*, a mailed survey completed by the graduate.
2. *Principal or supervisor evaluation*, a survey of the supervisor of the graduate to determine quality and success of the graduate in the job setting.

3. *Classroom students evaluation of graduate*, a survey of the students in the program graduate's classroom.

4. *Observation by an independent observer*, graduate evaluation conducted in the teaching setting by an independent observer, not a principal or supervisor.

5. *Institutional board of review of graduates, at completion*, institutional review of assembled dossiers of graduates, performed at program completion.

6. *Institutional board of review of graduates, post program*, institutional review of assembled dossiers of graduates, performed one to five years after program completion.

7. *Exit comprehensive exams*, written and/or oral examination of candidates immediately prior to program exit.

8. *Teacher tests*, standardized exams, such as the NTE, used to provide data for graduate evaluation.

9. *Telephone interviews*, calls placed to graduates to obtain data.

10. *Longitudinal studies*, continuing evaluations of graduates using the same variables conducted over several years.

11. *Pupil achievement*, measurement of the academic achievement of students in the graduate's teaching classroom.

12. *Professional organization activity*, evaluation of the extent of graduate's participation in professional organizations.

13. *News/activities file*, collection of newspaper or professional publication information relating to graduate's activities and accomplishments.

Respondents were asked to indicate if each presented technique had been used at their institution in the last three years. Those persons indicating "yes" were asked to respond to five additional questions indicating the success of the employed techniques at their institutions. The five additional questions under each item dealt with cost effectiveness, ease of implementation, usefulness for meeting the NCATE evaluation criterion, usefulness for program improvement, and likelihood of employing the technique over the next three years.

In responding to each item, respondents were allowed to select from five choices to indicate their agreement with each item: strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree.

The second section requested demographic information relating to the surveyed institution. Demographic information included whether the institution was private or public, whether the institution granted doctoral degrees, number of graduates each year from teacher education, years the institution had been NCATE-accredited, annual funds spent for program evaluation, and date of last NCATE accreditation. The third section presented 13 possible barriers to conducting evaluations and respondents were asked to indicate the extent to which they felt these were barriers in their institutions. The fourth section asked which methods were used to track program graduates and their perceived effectiveness.

Results

Characteristics of the respondent group

Table 1 presents the characteristics of the respondent group. Most (78.5%) were from public institutions whereas 21.5% were from private institutions. Slightly more than half (55.1%) were from doctoral-granting institutions. Slightly more than half (56.1%) graduated fewer than 300 students per year while 43.0% graduated 300 or more students per year. Most (71.0%) spend less than $5,000 per year to conduct evaluation of program graduates. These data indicate the variety of programs represented in the sample.

Table 1
Survey Respondent Characteristics, n= 107

Respondent represents	f	%
Public institution	84	78.5
Private institution	23	21.5
Doctoral granting	59	55.1
Not doctoral granting	48	44.9
Number of graduates per year:		
Less than 50	4	3.7
50 to 99	10	9.3
100 to 199	24	22.4
200 to 299	22	20.6
300 to 399	15	14.0
400 to 499	10	9.3
500 to 999	19	17.6
1000 or more	2	1.9
Funds spent per year		
Less than $1,000	22	20.6
$1,000 to 2,499	35	32.7
$2,500 to 4,999	19	17.8
$5,000 to 7,499	4	3.7
$7,500 to 9,999	2	1.9
$10,000 or more	5	4.7
Unknown	20	18.7

Methods Used for Program Graduate Evaluation

Table 2 presents the frequency of use of the 13 different techniques. Clearly, the four most frequently used techniques, used by at least half of the institutions, were: mailed questionnaires to graduates (97.2%); standardized teacher tests such as the NTE (71.0%); surveys completed by the principal or supervisor in the job setting (69.2%); and exit comprehensive exams, written or oral examinations of candidates immediately prior to program

exit (54.7%). Other fairly frequently used techniques were: longitudinal studies (43.0%); telephone interviews with the graduate (29.0%); news/activities files (24.3%); observation of the practicing graduate by an independent observer (17.9%); and classroom student evaluation of the graduate (17.1%).

Table 2
Frequency and Percent Use of Evaluation Techniques, n= 107

Rank	Evaluation technique	f	%
1	Questionnaire	104	97.2
2	Teacher tests	76	71.0
3	Principal/supervisor evaluation	72	69.2
4	Exit comprehensive exam	58	54.7
5	Longitudinal study	46	43.0
6	Telephone interview	31	29.0
7	News/activity file	26	24.3
8	Observation by independent observer	19	17.9
9	Classroom student evaluation	18	17.1
10	Inst. review board, at completion	11	10.4
11	Professional organization activity	6	5.6
12	Pupil achievement	4	3.7
13	Inst. review board, post program	2	1.9

Table 3 presents reported frequency of use of multiple techniques. Only the techniques with 20% or more use are included in this comparison. Examination of this table combined with the actual patterns of responses provides an indication of the various combinations of techniques used. In every case, no matter how many techniques are used, the questionnaire was the one used most often. Only one percent reported using only one technique, which was the questionnaire. Ten percent reported using two techniques, typically a combination of questionnaire and either teacher tests or principal/supervisor evaluation. Eighteen percent used three of the techniques, most often the questionnaire, teacher tests, and principal/supervisor evaluation. When four techniques were used, by 23% of the respondents, the set was most likely to be: questionnaire, teacher tests, principal/supervisor evaluation, and exit comprehensive exam. When five techniques were used, by 25% of the respondents, the combination was most likely to be the same as for four techniques with longitudinal studies added. With six or more techniques, used by 21% of the respondents, news/activity file or telephone interview was added to the five most likely used techniques.

Table 3
Frequency and Percent Use of Evaluation Techniques when One to Six or More Techniques are Used

	Number of techniques used											
	One used		Two used		Three used		Four used		Five used		Six+ used	
Evaluation	n = 1, 1%		n = 11,10%		n = 19,18%		n = 25,23%		n = 27,25%		n= 23, 21%	
Technique	f	%	f	%	f	%	f	%	f	%	f	%
Questionnaire	1	100	11	100	18	95	25	100	26	96	23	100
Teacher tests			4	36	9	47	19	76	22	81	22	21
Prin./super. Evaluation			3	27	10	53	14	56	24	89	21	20
Exit Comp. Exam			2	18	5	26	14	56	18	67	19	18
Longitudinal Study			2	18	6	32	9	36	13	48	16	15
Telephone Interview					3	16	7	28	10	37	11	10
News/act.File					2	11	3	12	9	33	12	11

Perceptions of the Evaluation Techniques

Table 4 presents results of respondent perceptions of the 13 evaluation techniques, ordered from high to low by total score (average of the items relating to ease of use, cost effectiveness, helpfulness in meeting the NCATE criterion, helpfulness in program improvement, and likelihood of use in the future). Only respondents who responded to all five items were included in this analysis. Five techniques had mean scores very close to 4.0 on the five-point scale. These were: principal/supervisor evaluation (4.06), observation by an independent observer (4.03), questionnaire (4.00), exit comprehensive exam (3.98), and institutional review board at program completion (3.98).

A cluster of four techniques had mean scores at about 3.75. These were: classroom student evaluation (3.76), longitudinal study (3.75), telephone interview (3.74), and teacher tests (3.73). Of the ten most used, the techniques of institutional review board post program completion (3.50), professional association activity (3.48), pupil achievement (3.30), and news/activity file (3.03) received the lowest overall scores.

Comparing the rank orders of the techniques across the five variables provides some interesting results. Observation using an independent observer was rated high relative to meeting the NCATE evaluation standard and helpfulness in improving programs but was rated lower than other techniques relative to ease of use and cost-effectiveness. Teacher tests and exit exams were ranked high on ease of use and cost-effectiveness but were relatively low on helpfulness in meeting the NCATE standard or improving the program. Institutional review boards at program completion were ranked high in terms of helpfulness both in meeting the NCATE evaluation standard and in improving programs, while they were ranked relatively low in terms of ease of use and cost-effectiveness. Longitudinal studies were ranked high relative to helping with program improvement but low relative to ease of use and cost-effectiveness.

Respondents were asked how likely it would be that the technique would be used within the next three years. These were responses from those who had reported using the technique. The four techniques most likely to be used in the future were: institutional review board at program completion, questionnaires, principal/supervisor evaluation, and exit comprehensive exam. Comparing these with the techniques used (see Table 2), it seems institutional review boards at program completion are likely to be used more in the future and teacher tests used less than in the past.

Table 4
Means and Ranks of Perceptions of Evaluation Techniques, Ordered High to Low by Total Score

Evaluation Technique	Easily Conducted		Cost Effective		Helpful NCATE Std		Helpful Impr.Prog.		Prob. Use in Future		Total Score	
	M	Rank	M	Rank	M	Rank	M	Rank	M	Rank	M	Rank
Prin./Superv. Eval.	3.76	3	3.90	1	4.27	2.5	4.04	4	4.35	3.5	4.06	1
Obs. by Indep. Obsrvr.	3.67	5	3.67	4	4.39	1	4.22	2	4.22	5	4.03	2
Questionnaire	3.72	4	3.73	3	4.23	4	3.96	6	4.36	2	4.00	3
Exit Compr. Exam	3.82	2	3.86	2	4.04	6.5	3.84	8	4.35	3.5	3.98	4.5
Inst. Rev. Bd., Compl.	3.18	9	3.45	6	4.27	2.5	4.45	1	4.55	1	3.98	4.5
Clsrm. Student Eval.	3.44	6	3.44	7	4.06	5	3.94	7	3.94	11	3.76	6
Longitudinal Study	3.13	10	3.42	8	4.04	6.5	4.07	3	4.04	8	3.75	7
Telephone Interview	3.35	7	3.32	10	3.84	10	4.03	5	4.13	7	3.74	8
Teacher Tests	3.91	1	3.53	5	3.88	9	3.28	10	4.14	6	3.73	9
Inst. Rev. Bd., Post	3.00	11.5	3.00	13	4.00	8	3.50	9	4.00	9.5	3.50	10
Prof. Org. Activity	3.00	11.5	3.40	9	3.80	11	3.20	12	4.00	9.5	3.48	11
Pupil Achievement	3.25	8	3.25	12	3.25	12	3.25	11	3.50	13	3.30	12
News/Activity File	2.79	13	3.29	11	3.08	13	2.42	13	3.54	12	3.03	13

Barriers to Conducting Evaluation of Program Graduates

Respondents were asked to indicate their perceptions of the extent to which certain factors were barriers to conducting these evaluations. They were asked to respond to 13 items on a scale of zero (not at all a barrier) to four (always a barrier). Table 5 presents the results for these items. Clearly, five of these were rated higher as barriers: lack of faculty/staff time (2.77), cost (2.61), lack of responsiveness of graduates to requests for information (2.40), difficulty in locating program graduates (2.30), and lack of responsiveness of supervisors and/or administrators to requests for information (2.22). Lack of technical expertise (1.19) and lack of guidelines (1.35) were not viewed as barriers by the respondent group. Thus, getting time and financial resources for conducting evaluation seemed to be more problematic than institutional personnel capability or existence of methodologies.

Table 5
Barriers to Conducting Program Graduate Evaluation, Ordered High to Low

Rank	Barrier	*n*	*M*	*SD*
1	Lack of faculty/staff time	104	2.77	1.26
2	Cost	99	2.61	1.17
3	Lack of responsiveness of graduates to requests for information	104	2.40	1.12
4	Difficulty in locating graduates	105	2.30	1.22
5	Lack of responsiveness of supervisors and/or administrators to requests for information	98	2.22	1.02
6	Benefits gained not perceived as worth the effort required to conduct	71	1.93	1.21
7	Lack of priority given to evaluation	104	1.90	1.10
8	Difficulty getting faculty to agree on criteria for evaluating graduates	104	1.76	1.03
9	Lack of faculty leadership to conduct such evaluation	103	1.75	1.23
10	Difficulty getting standardized scores on teacher tests	97	1.47	1.23
11	Lack of good examples for conducting such evaluations	104	1.44	1.23
12	Lack of guidelines for conducting such evaluations	104	1.35	1.12
13	Lack of faculty technical expertise	103	1.19	1.15

Scale: 0 = Not at all to 4 = Always a barrier

Method of Tracking Program Graduates

Three of the highest five barriers related to problems in locating graduates and lack of responsiveness for requests for information from program graduates and supervisors or administrators. Respondents were asked whether they used six methods to track graduates and if so, their perceptions of the effectiveness of the methods. Table 6 presents the results for these items. The most frequently used method was through the college/university alumni association (90.4%), followed by education alumni association (50.5%), teacher placement service (49.5), and contacting parents using "permanent address" files (45.6%). Respondents who reported using the methods were asked to rate effectiveness of the methods on a scale of 0 (not at all) to 4 (very effective). The ratings on the methods ranged from 2.21 to 2.59, not much variation. The two highest were: teacher placement service (2.59) and education alumni association (2.54). Respondents were asked to list other methods they use to track graduates using an open-ended item. Other methods indicated were: through education faculty/departmental contacts, from SDE pay rosters, direct mailings to all schools/supervisors in state, beginning teacher conference, registrar's files, student locator cards, maintaining contact with select groups through guaranteed payment for interviews and survey responses, first year mentoring program, advisor/mentor contacts, and through an educational research center.

Table 6
Methods Used to Track Program Graduates and Perceived Effectiveness

Method	Used method f	Used method %	Effectiveness *n*	Effectiveness *M*	Effectiveness *SD*
Education alumni association	52	50.5	50	2.54	0.95
College/university alumni association	94	90.4	88	2.38	1.00
Monitoring local school system rosters	36	35.3	33	2.21	0.86
Monitoring state teacher association rosters	21	20.4	19	2.42	1.12
Contacting parents of graduates or using "permanent address" files	47	45.6	44	2.27	0.92
Through teacher placement service	50	49.5	44	2.59	0.92

Effectiveness scale: 0 = Not at all to 4 = Very effective

Summary and Discussion

Conducting evaluation of program graduates is essential to evaluating teacher preparation effectiveness. Until recently almost all such evaluations were based on questionnaires to graduates and their supervisors and, to a limited extent, longitudinal studies and standardized teacher tests, such as the NTE. It seems the NCATE focus on increasing the scope of program graduate evaluation has stimulated broader thinking about different techniques. These techniques vary in their ease of use, their cost-effectiveness, and helpfulness in meeting the graduate evaluation standards and improving teacher preparation program(s). The five most used techniques were: questionnaire, teacher tests, principal/supervisor evaluation, exit comprehensive exams, and longitudinal studies. Overall the five most highly rated methods were: principal/supervisor evaluation, observation by independent observer, questionnaire, exit comprehensive exam, and institutional review board at program completion.

Many, if not most, NCATE accredited institutions are using a variety of mostly traditional, with some emerging, techniques. At any rate, most of the institutions use multiple methods or techniques to collect data and institutions believe these techniques are generally useful. Institutional review boards at program completion seem to be an emerging technique. The barriers to conducting such evaluations are more associated with resources such as lack of faculty/staff time and financial support. Additional barriers are related to difficulty in getting graduates and their supervisors to respond to requests for information. It is clear that most respondents felt the technical and methodological resources existed for conducting these evaluations, but time and fiscal resources are lacking.

Tracking program graduates is critical for conducting these evaluations. This tracking or follow-up is a problem for almost every program. Methods used by the respondent group tended to not be rated very high in effectiveness. Those methods most closely tied to the preparation program were rated as most effective. They were: through the teacher placement service and through the education alumni association. Relying on sources less under the control of the unit were viewed as not being as effective as those under the control of the unit.

Results of this research indicate there has been some change in recent program graduate evaluation techniques. The follow-up survey has not been replaced but is being supplemented with a variety of other techniques viewed as being more helpful and worth the additional costs. There are, however, continued concerns that time and fiscal resources will be provided by institutions to conduct this important work. There is a need to conduct further study in this area.

Of high potential is the increased use of technology to track and communicate with program graduates and collect data from field-based settings. Although not a part of this study, another important aspect of this issue which needs to be researched is the methods of integrating results from such multiple methods, such as in using various triangulation methods.

This research has the typical limitations associated with survey research. Generalizibility of these results is limited by two factors, representativeness of the sample and time. The sample return rate was reputable, and respondents clearly attended to the items. Of more concern is the time factor. Educational research and evaluation practices are relatively slow to change so these results are likely to be applicable today. At least evidence provided by this research indicates NCATE-accredited institutions are taking steps to be responsive to the more current standards and improve on previous practices.

Teacher preparation administrators need to facilitate and provide needed resources for those planning and conducting accreditation evaluations. Educational researchers/evaluators need to continue to develop new methods and find better ways of implementing extant methods considered to be most useful yet costly and time-demanding. School administrators need to be engaged in the planning for evaluation and provision of information. The importance of participating fully in evaluation information provision, including keeping the granting institution informed as to professional status and location and responding to information requests, must be conveyed to program graduates.

References

Ayers, J. B. (1988). Teacher education follow-up evaluation: How to do it. In W. J. Gephart & J. B. Ayers (Eds.), *Teacher Education Evaluation* (pp. 85-111). Boston, MA: Kluwer Academic Publishers.

Ayers, J. B. (1989a, February). *A fifteen year look at beginning teachers*. Paper presented at the annual meeting of Eastern Educational Research Association, Savannah, GA.

Ayers, J. B. (1989b, February). *The NTE, PPST, and classroom performance*. Paper presented at the annual meeting of the Association of Teacher Educators, St. Louis. MO.

Bloom, L., & Bacon, E. (1995). Using portfolios for individual learning and assessment. *Teacher Education and Special Education, 18*(1), 1-9.

Bond, F. T. (1990, February). *Methods and instruments used to conduct follow-up studies of teacher education graduates*. Paper presented at the annual meeting of

the Eastern Educational Research Association, Clearwater, FL.

Boser, J. A. (1988). Teacher education graduate surveys: variables related to response rate. *Journal of Educational Research, 81*(6), 369-373.

Clark, S. B., & Nichols, J. O. (1983, May). *Increasing the precision of estimates in follow-up surveys: A case study.* Paper presented at the annual forum of the Association for Institutional Research, Toronto, Ontario.

Holste, D., & Matthews, D. (1993). Survey of 1991 teacher education graduates conducted in May 1992. Urbana-Champaign: University of Illinois, Council on Teacher Education.

Kunkel, R. C. (1988). The professional education unit. In W. J. Gephart & J. B. Ayers (Eds.), *Teacher Education Evaluation* (pp. 42-57). Boston, MA: Kluwer Academic Publishers.

Lindsay, M. (1985). Procedures for follow up studies of teacher education graduates. *Journal of Teacher Education, 36*(2), 29-33.

McEneaney, J. E., & Sheridan, E. M. (1993). *Evaluating the effectiveness of an undergraduate teacher education program.* (ERIC Document Reproduction Service No. ED 358 062)

Murphy, P. D. (1992, February). *Follow-up studies in program evaluation.* Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, San Antonio, TX.

National Council for Accreditation of Teacher Education (1987). *Standards for the accreditation of professional education units*. Washington, DC: Author.

National Council for Accreditation of Teacher Education (1995). *Standards, procedures, and policies for the accreditation of professional education units.* Washington, DC: Author.

Nuccio, E. J. (1990, April). *Perceptions of the knowledge base by three groups of teacher education graduates.* Paper presented at the annual meeting of the American Educational Research Association, Boston, MA.

Peterson, K. (1989). Teacher education evaluation through an institutional review board of graduates. *Journal of Research and Development in Education, 22*(4), 14-20.

Redfield, D. L. (1988). Pupil achievement: The missing link in the evaluation chain. In W. J. Gephart & J. B. Ayers (Eds.), *Teacher Education Evaluation* (pp. 112-131). Boston, MA: Kluwer Academic Publishers.

SAS®, SAS Institute, Cary, North Carolina.

Warren, R. D. (1990, October). *A longitudinal examination of teaching potential, career paths, and teacher retention of teacher education graduates.* Paper presented at the annual meeting of the Midwestern Educational Research Association, Chicago, IL.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 2, 63-74

Person-Fit and the Rasch Model: How Seriously Model Fit Is Affected by Appropriateness Measurements in the Rasch Model

Necati Engec
State of Louisiana, Department of Education

In this study, the researcher investigated the impact of aberrant response behaviors such as guessing, sleeping, and cheating on ability parameter estimates in the Rasch model. How seriously are invariance properties of ability and item parameter estimates affected by appropriateness measurements in the Rasch model? A 20-item multiple choice math test was administered to 5200 examinees. Results indicated that (1) the invariance properties of Rasch model ability and item parameter estimates increase when person-fit statistics are close to zero or larger than zero, and (2) the standard errors associated with ability and item parameter estimates also increase, as person-fit statistics increase. Grouping students in terms of aberrant response behaviors can violate the normality assumption in each group, making KR-20 computations questionable. Test developers should be careful about heterogenous ability distributions, while excluding aberrant response behaviors to get more accurate and invariant parameter estimates.

The general purpose of a person-fit statistic or appropriateness measurement is to detect aberrant response behaviors of students. These are the students who answer an item by guessing, cheating, sleeping, plodding, or something else, not using their knowledge of the subject. This response pattern is called "aberrant response behavior" (Wright & Stone, 1979, pp.170-180), and statistical techniques to detect this kind of examinee are called person-fit statistics or appropriateness measurements. While a very low, negative person-fit statistic indicates that students answered the item by cheating, guessing, or something else, a high person-fit statistic indicates that students answered the item in terms of knowledge in that subject. A clear example is given by Hulin, Drasgow, and Parsons (1983, pp. 122-127). Generally, a person-fit statistic will be a negative number that is relatively large in magnitude for any response vector with many correct responses to hard items and incorrect responses to easy items. In the early person-fit literature, several types of aberrant response behavior and numerous statistical approaches existed to assess person-fit. Aberrant response behaviors such as (1) sleeping, (2) guessing, (3) cheating, (4) plodding behavior, (5) alignment errors, (6) extremely creative examinees' behavior, and (7) deficiency of abilities are summarized by Meijer (1996) in a special issue of *Applied Measurement in Education*, and statistical approaches to assess person-fit are investigated by the authors in that issue. An illustration of seven item score patterns typical of the seven types of aberrant behaviors is given in Table 1. These patterns are based on a fictitious 12 four-choice item test. The π values were chosen so that there are four easy items (items 1 through 4), four items of medium difficulty (items 5 through 8), and four difficult items (items 9 through 12) (Meijer, 1996). A similar example can be found in Wright and Stone (1979).

Table 1
Aberrant Item Score Patterns on a Fictitious 12-Item Test

	Item												
Person	1	2	3	4	5	6	7	8	9	10	11	12	Behavior
1	0	0	0	1	1	1	1	1	1	1	1	1	Sleeping
2	1	1	1	1	0	0	1	0	0	0	0	1	Guessing
3	1	1	0	1	0	1	0	0	0	1	1	1	Cheating
4	1	1	1	1	1	0	1	1	0	0	0	0	Alignment errors
5	1	1	1	1	1	1	0	0	0	0	0	0	Plodding
6	0	0	0	0	1	1	1	1	0	1	1	1	Extremely creative
7	0	0	1	0	1	1	1	0	1	1	1	0	Deficiency of abilities
π	.90	.85	.83	.82	.57	.55	.50	.49	.30	.25	.21	.15	

Source: Meijer, R.R., (1996). Person-fit research: An introduction. *Applied Measurement in Education*, *9*(1), 3-8.

Wright and Stone (1979, pp.181-188) suggested correction measures for aberrant responses: (1) when the majority of unexpected responses are "incorrect" and $t > 3$ then delete all the "too easy" items and compute the new ability estimates after the deletion, and (2) when the majority of unexpected responses are "correct" and $t > 3$ then delete all the "too hard" items and compute new ability estimates after deletion. The goal of this study was not to investigate the aberrant response behaviors or

Necati Engec, Ph.D., is a research analyst in Planning and Analysis Section, Office of Management and Finance, Louisiana Department of Education. Correspondence and requests for reprints should be addressed to Necati Engec, State of Louisiana, Department of Education, P.O. Box 94064, Baton Rouge, Louisiana 70804-9064. E-mail: nengec@hotmail.com. Phone: (225)342-0668. Fax: (225)342-7185.

statistical approaches to assess person-fit. Instead, the purpose of this study was to investigate the impacts of aberrant response behaviors on the item and ability parameter estimates in the Rasch Model. In other words, by excluding these students (not the items) from the data set, can one get more accurate item and ability parameter estimates? How is model fit affected by aberrant response behaviors? The researcher looked at the relationship between person-fit and invariance properties of the Rasch model that is the basic indicator of the accurate parameter estimates and model-fit. Ultimately, the interest was to see how seriously model fit is affected by appropriateness measurements. Under the model fit, how seriously person-fit statistics affected several things were investigated: (1) invariance of ability parameter estimates, (2) standard errors associated with IRT maximum likelihood ability estimates, (3) invariance of item parameter estimates, (4) standard errors associated with item difficulty estimates, (5) Pearson Chi-square lack of fit statistics, and (6) test information functions. In this study, these concerns were investigated under ten different groups with the same sample size and test length but different person-fit statistics. Most of the students in group 1 were the students who answered items on the test by cheating, guessing, or sleeping, since the students in that group have a low, negative person-fit statistic. On the other hand, the students in group 10 were the students with very high, positive person-fit statistics. The mean of the person-fit statistics in group 1 is the smallest, while that in group 10 is the largest. The most direct factor that affects the accuracy of parameter estimates is the size of the data set. In general, increasing the number of students and/or items results in a reduction in the standard error values of the parameter estimates. That is why in each group, sample size and test length were held constant to eliminate their effect in the parameter estimate.

First, model fit and its aspects are explained, including invariance of ability parameter estimates, standard errors associated with IRT maximum likelihood ability estimates, invariance of item parameter estimates, standard errors associated with item difficulty estimates, Pearson Chi-square lack-of-fit statistics, and test information function. Second, person-fit statistics developed by Levine and Rubin (1979) and Drasgow (1982) and their interpretation are studied. Last, the impacts of the person-fit statistics on the model fit are investigated and discussed.

Assessment of Model-Data Fit

One of the basic indicators of model data fit in item response theory is the invariance of item and ability parameter estimates. That implies that item parameter estimates are not group dependent. When the group in which item parameters are calculated is changed, item parameter estimates do not change. In the same way, ability parameter estimates are not test dependent. That means that when the test by which ability parameters are calculated is changed, ability parameters do not change. Basically, the property of invariance of item and ability parameters implies that the parameters that characterize an item do not depend on the ability distribution of examinees, and the parameters that characterize examinees do not depend on the set of test items. The invariance properties of Rasch model parameters can be assessed using several straightforward methods discussed below.

Invariance of Ability Parameters

Ability estimates from different samples of test items (for example, even- and odd- numbered items) are compared. Invariance is established when the estimates do not differ in excess of the measurement errors associated with the estimates (Wright, 1968). The interest here was to determine if the ability estimates differed based on whether the item numbers were even or odd when the person-fit statistics differed with the same sample sizes and test lengths.

Invariance of Item Parameter Estimate

Comparison of the Rasch model item parameter estimates (b values) obtained from two or more subgroups (such as male and female) is of interest. When the estimates are invariant, the plot should be linear with the amount of scatter reflecting errors due only to the sample size. Baseline plots can be obtained by using randomly equivalent sample sizes (Shepard, Camilli, and Williams, 1984). The interest here was to see whether item parameter estimates based on the subgroups of male and female differed based on the level of person-fit statistics with constant sample sizes and test lengths.

In the application of the Rasch model, the invariance of model parameters is gained only at the expense of making strong assumptions about the nature of the data. It is widely recognized that these assumptions are unlikely to be fully met in practice. But the purpose here was not to check the assumptions of the Rasch model or the invariance of model parameter estimates. The concern was to investigate how seriously the invariance of parameter estimates change in a varied range of person-fit statistics.

Standard Errors Associated with IRT Parameter Estimates

An asymptotic estimate of the standard error associated with each item difficulty and ability estimate provides an indication of the amount of error in the estimates. The most direct factor that affects the accuracy of parameter estimates is the size of the data set. In general, increasing the number of students and/or items results in a reduction in the standard error values of the parameter estimates. In this study, by holding sample size and test length constant in each group, the impact of the person-fit

on the standard error associated with each item and ability parameter estimate was investigated.

Pearson Chi-square Lack-of-fit Statistics

Pearson Chi-square lack-of-fit statistics (i.e., model fit) were also investigated under the different person-fit statistics. High and statistically significant Chi-square values may suggest that these items are not adequately described by the Rasch model. The Pearson Chi-square lack-of-fit statistic in each group which has different person-fit statistics was investigated.

Person-fit Statistic and Test Information Function

In this study, estimated reliability values based on the conditional standard errors of measurement provided by the IRT model were investigated under the different person-fit groups. These estimated reliability coefficients can be interpreted as an estimate of the conventional alpha (KR-20) internal consistency reliability. It assumes that ability follows a standard-normal distribution. Also, expected and average information were investigated under different person-fit statistics. The expected information is the average amount of information that a test would provide for a sample of examinees drawn from a population with a normal distribution of ability. The average information is the average height of the plotted information curve. The average information is thus the amount of information that a test would provide for a sample of students with abilities that are rectangularly distributed.

Person-fit statistics. Although numerous statistical approaches exist to determine the varied range of person-fit statistics, this research was focused on the l_z statistic as developed by Drasgow, Levine, and Williams (1985). Unlike l_o and l_g developed by Levine and Rubin (1979), the l_z score represents the standardized likelihood of an individual's pattern of item response given a set of estimated IRT model parameters.

The simplest of the l_o indexes studied by Levine and Rubin (1979), and Levine and Drasgow (1982) is as follows:

$$l_o = \text{Log max Prob}(U|Q) = \log \text{Prob}(U|Q)$$

where U is a vector of item responses, prob(U|Q) is the likelihood function and Q is the maximum likelihood estimate of Q. For any response vector with many correct responses to hard items and incorrect responses to easy items, l_o will be a negative number that is relatively large in magnitude. It is clear that l_o succeeds in detecting the aberrant response pattern of students. But there is no explicit provision in l_o for omitted responses. For tests in which students omit varying numbers of items, l_o may be misleading. This is because an examinee who answers more items than another examinee will likely have a response vector with a smaller l_o value and appear more aberrant.

To improve the l_o index, Drasgow (1982) computed l_o for all items answered, and then calculated the geometric mean likelihood as follows:

$$l_g = \exp(\text{Lo}/n),$$

where n is the number of items attempted by the examinee. Clearly, a small value of l_g indicates an atypical response vector. Both l_o and l_g can be misleading if students at different ability levels are compared. Drasgow, Levine, and Williams (1982) standardized l_o as follows:

$$l_z = \frac{l_o - E(l_o)}{[\text{Var(Lo)}]^{1/2}}$$

where $E(l_o) = \sum_I [\text{Pi1 log Pi1} + \text{Pi0 log Pi0}]$, and

$$\text{Var}(l_o) = \sum_I \text{Pi1 Pi0}[\log(\text{Pi1/Pi0})]^2.$$

Use of l_z allows comparisons among index scores for examinees answering different numbers of items. In addition, l_z eliminates the dependence upon Q, so that index scores for students at different ability levels can be compared.

Literature Review

There is an abundance of research related to appropriateness measurement and person-fit statistics. However, no single research study investigating how seriously model fit is affected by the appropriateness measurement using the deletion of aberrant response behaviors could be found.

Wright and Stone (1979) summarized the normal, sleeping or fumbling, guessing, and plodding response patterns; investigated misfitting records to see how the diagnosis of irregular response patterns might be accomplished; and suggested correcting measures. To summarize, the statistical aspects of their correction strategies are: (1) when the majority of unexpected responses are "incorrect" and $t > 3$, then delete all the "too easy" items and compute the new ability estimates after the deletion, and (2) when the majority of unexpected responses are "correct" and $t > 3$, then delete all the "too hard" items and compute new ability estimates after deletion.

In the Monte Carlo investigation of several person- and item-fit statistics for the item response model, Roger and Hattie (1987) reported the behavior of several person- and item-fit statistics commonly used to test and obtain fit

to the one-parameter item response model. Using simulated data for 500 persons and 15 items, the sensitivity of the total-*t*, mean square residual, and between-*t* statistics to guessing and multi-dimensionality were examined. Additionally, 25 fitting persons and a misfitting item were generated to test the power of the three fit statistics to detect deviations in a subset of observations. They found that neither the total-*t* nor mean square residual were able to detect deviation from any of the models fitted. Use of these statistics appears to be unwarranted. The between-*t* was a useful indicator of guessing and heterogeneity in discrimination parameters, but did not detect multidimensionality. These results show that the use of person- and item-fit statistics to test and obtain overall fit to the one-parameter model can lead to acceptance of the model even when it is grossly inappropriate. These results also show that assessments of the model fit based on this strategy are inadequate. Alternative methods must be sought.

Reise (1990) investigated the comparison of item- and person-fit methods of assessing model data fit in IRT. The Chi-square item-fit index and likelihood-based person-fit index were compared. It was concluded that under simulated conditions of data misfit, the Chi-square statistic detected misfit at a higher rate than the likelihood-based statistics, indicating that the Chi-square statistic was slightly more sensitive to response pattern aberrance. On the other hand, McKinley and Mills (1985) studied the performance of several Chi-square fit indices in Monte Carlo research. They found it difficult to choose between the competing methods. No particular Chi-square fit index was clearly superior to the others. However, in their research, the Chi-square yielded the smallest number of erroneous acceptances of fit under conditions of simulated model-data misfit.

An extensive review of goodness-of-fit statistics is provided by Troub and Lam (1985). Some of the Pearson Chi-square statistics have been proposed by Martin-Lof (cited in Gustafsson 1980, p.213) and Wright and Pancahapakesan (1969) for the Rasch model, and Bock (1972) for the two-parameter model. As a Chi-square statistic, Q1 has been added by Van Den Wollenberg (1982) for the Rasch model, and another by Yen (1981) for either the one-, two-, or three-parameter model.

Other indicators of misfit are likelihood ratio statistics. Best known of these is Andersen's (1973) conditional likelihood-ratio statistics for the Rasch model. Waller (1981) proposed another such statistic for the one-, two-, and three-parameter models. McKinley and Mills (1987) conducted a study to evaluate four goodness-of-fit procedures using data simulation techniques.

The review, critique, and validation of appropriateness measurement were studied by Levine and Drasgow (1982), and optimal appropriateness indices were introduced by Levine and Drasgow (1984) to provide the highest rates of detection of aberrant response patterns that can be obtained from item responses. Another study done by Drasgow, Levine, and McLaughlin (1985) was on methods to detect inappropriate test scores with optimal and practical appropriateness. As seen in the literature review, there is no single research showing how model fit is affected by appropriateness measurements. The goal of this study was not to investigate aberrant response behaviors or statistical approaches to assess person-fit. The basic purpose of this study was to look at the relationship between person-fit and invariance properties of the Rasch model as a basic indicator of accurate parameter estimates and model fit. By excluding the aberrant responses such as sleeping, guessing, and cheating in item response theory models, can one get more accurate item and ability parameter estimates? Of interest was how seriously model fit is affected by appropriateness measurements. These concerns are investigated in 10 different groups with the same sample size and test length, but with different person-fit statistics. The mean of the person-fit statistics in group 1 was the smallest, while the mean was largest in group 10. That means that in group 1, there were many students who answered the items on the test by cheating, guessing, sleeping, or plodding, while there were only a few aberrant response behaviors in group 10. In each group, sample size and test length were held constant to eliminate the effect of parameter estimate.

Grouping of Examinees Based on the Person-fit Statistics

The purpose of this study was to investigate the impacts of aberrant response behaviors such as cheating, guessing, or sleeping on item and ability parameter estimates in the Rasch Model. In other words, by excluding these students from the data set, can more accurate item and ability parameter estimates be obtained? The data for this study were dichotomous and came from the California Achievement Test (CAT/5), which is a norm-referenced test (NRT) administered to 5,231 students (2,627 females and 2,604 males) in grades 4 and 6 as part of the LEAP, Louisiana Educational Assessment Program. Like the criterion-referenced tests (CRTs) and graduate exit examinations (GEE), norm-referenced tests administered to Louisiana students in grades 4 and 6 are a part of the LEAP. The California Achievement Test (CAT/5) (Louisiana Progress Profiles State report, 1995-1996) measures students' mastery in the following content areas: reading, language, mathematics, word analysis, spelling, study skills, science, and social studies.

In this study, actual data rather than simulated data were used. Ten different groups were created, each of them based on different person-fit statistics. First, person-fit statistics were calculated for all examinees; then they were grouped according to person-fit statistics maintaining the same sample size and test length in each group.

Parameters, such as sample size, length of the tests, and number of subjects, were the same for each group. Age, grade, SES and gender are not factors affecting the accuracy of parameter estimates. The most direct factor that affects the accuracy of parameter estimates is the size of the data set. In general, increasing the number of students and/or items results in a reduction in the standard error values of the parameter estimates. In this study, in each group, by holding sample size and test length constant, the impact of the person-fit on the standard error associated with each item and ability parameter estimate was investigated. Table 2 shows the grouping of students according to their person-fit statistics, sample size and summary statistics.

Table 2
Person-Fit Statistics Based on Different Subgroup for Math Test

Group No	N	Mean(Lz)	Sd(Lz)	Max	Min
Group 1	500	-50.84	36.11	-377.75	-43.58
Group 2	500	-41.24	1.21	-43.57	-39.35
Group 3	500	-25.73	17.17	-39.35	-2.05
Group 4	500	-1.41	.30	-2.04	-.99
Group 5	500	-.66	.17	-.99	-.41
Group 6	500	-.21	.12	-.40	-.01
Group 7	500	.15	.07	.01	.27
Group 8	500	.43	.10	.27	.62
Group 9	500	.77	.07	.63	.89
Group 10	500	1.27	.17	1.01	1.61
Control Group	**500**	**-13.64**	**30.19**	**-377.75**	**1.61**

As seen in Table 2, the sample size (N=approximately 500) and test length (20 item math test) in each group were held the same to isolate the sample size and test length effects on the parameter estimate. The mean of the l_z index for each group is different to see the effect of person-fit statistics (l_z) on the item and ability parameter estimate. Person-fit, l_z, is maximum in group 10, minimum in group 1, and almost zero in group 7. Different standard deviations among the groups were not the concern of this study and did not influence the parameter estimates. In other words, unlike in group 10, there were a lot of aberrant responses in group 1. That is, the number of students who answered the test items by cheating, guessing, sleeping, or plodding increased from group 1 to group 10. The control group included randomly selected students from the whole group (N=5231) to determine the regular parameter estimate in the Rasch model. The control group reflected randomly selected students for regular parameter estimates and was not totally separate from the 10 subsamples. For the regular parameter estimates, that group could have been used without controlling aberrant responses. Unlike the control group, the aberrant responses were controlled in group 1 to group 10. The total of Ns in Table 2 is 5,000. The last 231 students out of 5,231 were eliminated from the data set in terms of the ascending rank of the person-fit statistics. For each group, the mean, standard deviation, maximum and minimum of the person-fit statistics which reflect aberrant response behaviors are given in Table 2. In the study, item and ability parameters were estimated for each group to see the impact of the aberrant response behaviors on the item and ability parameter estimates in the Rasch model.

To estimate model parameters and to determine whether the invariance of parameters is significantly different among the subgroups based on different l_z indices, the Rasch model was used:

$$Pi(Q) = \frac{1}{1 + \exp^{-(Q - bi)}}, \qquad i = 1, 2, 3, \ldots, n$$

where

Pi(Q) is the probability that a randomly chosen examinee with ability Q answered item i correctly,
bi is the item I difficulty parameter,
n is the number of items in the test, and
exp is the transcendental number whose value is 2.718.

Person-fit and Invariance of Ability Parameters Estimations

The first activity was to investigate the invariance of the ability parameters across different samples of items. To see how seriously model fit (accuracy of parameter estimates) was affected by aberrant responses (appropriateness measurement), the Rasch model ability estimates were compared with different samples of test items (even- and odd-numbered items) for every subgroup based on different l_z indices, in order to see the relationship between the invariance of ability parameter estimates of the Rasch model and person-fit statistics. The interest here was to show whether the ability estimates based on even- and odd-numbered items differed or not in different groups based on varied levels of person-fit statistics with the same sample size and test length. Invariance of ability parameter estimates based on students' performance on the odd-numbered items and on the even-numbered items for each group indicated the accuracy of estimates. Simply, in this part of the study for each group, students' ability levels were calculated twice using odd- (items 1,3,5,7,9) and even- (items 2,4,6,8,10) numbered items, and calculating the Pearson correlation coefficient between them.

Table 3 provides the correlations between ability estimates obtained from odd- and even-numbered items for every subgroup. Invariance is established when the estimates do not differ in excess of the measurement errors associated with the estimates (Wright, 1968).

Table 3
Correlations Between Ability Estimates Obtained From Even- and Odd-Numbered Items for Every Subgroups

Group No	Correlation	*P*
Group 1	0.43	<0.0001
Group 2	-0.23	<0.0001
Gr oup 3	0.25	<0.0001
Group 4	0.22	<0.0001
Group 5	0.31	<0.0001
Group 6	0.48	<0.0001
Group 7	0.71	<0.0001
Group 8	0.64	<0.0001
Group 9	0.62	<0.0001
Group 10	0.68	<0.0001
Control Group	**0.53**	**<0.0001**

As seen in Table 3, the correlation coefficient for the control group is .528, and the maximum correlation coefficient is r = .708 where person-fit indices are the closest to zero (mean = .15). The correlation coefficients are very low in the groups that have negative l_z statistics while correlation coefficients are very high, even bigger than the control group, in the groups that have positive l_z statistics. Groups 1, 2, 3, 4, 5, and 6 have negative person-fit statistics and low correlation coefficients. On the other hand, groups 7, 8, 9, and 10 have positive l_z statistics and high correlation coefficients, even larger than the control group. These correlation coefficients clearly show that the invariance properties of the Rasch model ability estimates increase when person-fit statistics are bigger than zero. That means that if person-fit statistics are bigger than zero, one can get much more accurate ability parameter estimates. Further, that might mean that students who have negative l_z statistics do not fit the model very well. In current practice, if the Pearson Chi square lack-of-fit statistic for any item in the test is bigger than the degrees of freedom, that item is eliminated from the test since it does not fit the model. In the same way, Table 3 might suggest we eliminate students who have l_z indices that are smaller than zero to more accurately estimate parameters. Basically, it appears that ability parameter estimates are affected negatively by the students who answer the items in the test by guessing, cheating, sleeping, and so on.

Standard Errors Associated with IRT Maximum-Likelihood Ability Estimates

To see the relationship between person-fit statistics and IRT maximum-likelihood ability estimates, standard errors of measurement associated with these ability parameter estimates were investigated. These measurement errors provided an indication of the amount of error in the estimates. The interest here was the effect of aberrant responses on the standard errors of measurement associated with ability parameter estimates. Table 4 provides, for each possible correct score and group, standard error of measurement values associated with corresponding ability values. Note that it is not possible to provide ability estimates for zero or a perfect score. That is why a number of correct scores of zero and 20 and, for some groups, 18 are missing. There was no standard error for these. Also Table 5 shows their summary statistics. Note that N in Table 5 is the number of possible correct answers. For instance, for group 1, some students answered just one item on the test correctly, some 17, so there are 17 different score levels.

Table 4
Standard Errors Associated with IRT Maximum-Likelihood Ability Estimations for Different Groups Based on Person Fit Statistics

Score	GRP1	GRP2	GRP3	GRP4	GRP5	GRP6	GRP7	GRP8	GRP9	GRP10	Cont.Grp
0	.	.	.	.	.	.	.	.	.	.	.
1	1.09	1.09	1.06	1.06	1.07	1.10	1.13	1.39	2.12	1.24	1.12
2	0.84	0.84	0.80	0.79	0.81	0.84	0.89	1.02	1.09	1.09	0.87
3	0.74	0.74	0.70	0.68	0.71	0.74	0.79	0.85	0.87	1.13	0.76
4	0.69	0.68	0.64	0.62	0.65	0.68	0.74	0.79	0.81	1.30	0.71
5	0.67	0.66	0.62	0.59	0.62	0.65	0.72	0.76	0.79	1.18	0.68
6	0.65	0.64	0.60	0.56	0.60	0.64	0.70	0.76	0.81	1.14	0.66
7	0.64	0.63	0.60	0.55	0.59	0.64	0.70	0.76	0.85	1.19	0.65
8	0.63	0.63	0.59	0.54	0.59	0.63	0.69	0.77	0.87	1.06	0.64
9	0.62	0.63	0.59	0.53	0.58	0.64	0.68	0.76	0.84	1.01	0.63
10	0.62	0.63	0.59	0.53	0.58	0.64	0.67	0.73	0.79	1.08	0.62
11	0.62	0.62	0.59	0.54	0.59	0.64	0.66	0.70	0.76	1.14	0.61
12	0.62	0.63	0.59	0.55	0.59	0.65	0.66	0.68	0.74	1.01	0.60
13	0.63	0.64	0.60	0.57	0.61	0.66	0.66	0.68	0.73	0.93	0.61
14	0.66	0.66	0.61	0.61	0.64	0.68	0.68	0.69	0.76	0.96	0.62
15	0.71	0.71	0.64	0.66	0.69	0.73	0.72	0.73	0.84	1.17	0.64
16	0.80	0.81	0.69	0.77	0.79	0.82	0.81	0.82	1.09	.	0.70
17	1.06	1.07	0.79	1.04	1.06	1.08	1.07	1.07	.	.	0.80
18	.	.	1.06	.	.	.	.	.	.	.	1.06
19	.	.	.	.	.	.	.	.	.	.	.
20	.	.	.	.	.	.	.	.	.	.	.

Table 5
Summary Statistics of Standard Errors

Variable	N	Mean	Std Dev	Minimum	Maximum
GROUP1	17	0.727	0.147	0.623	1.091
GROUP 2	17	0.728	0.148	0.629	1.098
GROUP 3	18	0.692	0.149	0.597	1.065
GROUP 4	17	0.663	0.168	0.536	1.064
GROUP 5	17	0.697	0.156	0.588	1.077
GROUP 6	17	0.736	0.147	0.639	1.101
GROUP 7	17	0.767	0.140	0.662	1.133
GROUP 8	17	0.825	0.183	0.681	1.397
GROUP 9	16	0.926	0.335	0.737	2.120
GROUP 10	15	1.114	0.102	0.936	1.303
Control Grp	**18**	**0.724**	**0.152**	**0.607**	**1.125**

For item 1, item 19 (except group 3 and the control group), and item 20, there was no standard error since there was no variation in these items. The purpose of this study was not to detect good or bad items. The concern here was the means of the standard errors of ability estimates. That is why these items were not eliminated from the test. As seen in Tables 4 and 5, standard errors associated with maximum likelihood ability parameter estimate varied in the groups based on different person-fit statistics. The maximum mean standard error, as seen in Table 5, is 1.11 in group 10, which had the largest person-fit statistic. Groups 8 and 9, that have the largest person-fit statistics after group 10, also have the largest standard errors, 0.825 and 0.926 respectively. Standard errors associated with maximum likelihood ability parameter estimates were very low in the groups with negative l_z statistics, while standard errors associated with maximum likelihood ability parameter estimates were very high in the groups that have positive l_z statistics, even higher than the control group. Groups 1, 2, 3, 4, 5, and 6 have negative person-fit statistics and small standard errors associated with maximum likelihood ability parameter estimates. On the other hand, groups 7, 8, 9, and 10 have positive l_z statistics and higher standard errors associated with maximum likelihood ability parameter estimates, even larger than the control group. Table 5 clearly shows that standard errors associated with maximum likelihood ability parameter estimates increased when person-fit statistics were larger than zero. As a summary of Tables 3 and 5, it can be concluded that ability parameter estimates were affected positively by the students who answered the items in the test by cheating, guessing, sleeping, and so on, but standard errors associated with each ability estimate increased. The reason that standard errors associated with ability parameter estimates increased might be ability distributions. Grouping students according to aberrant response behaviors creates homogeneous ability distributions. Satisfactory item and ability estimation requires heterogenous ability distributions.

Person-fit and invariance of item parameter estimates

The next activity was to investigate the invariance of the item parameter estimates for each group. Rasch model difficulty estimates obtained in two subgroups (male and female) were compared for every Lz-subgroup based on different l_z indices, to see the relationship between the invariance of difficulty parameter estimations of the Rasch model and person-fit statistics. The interest was to determine if the difficulty estimates based on gender differed under varying levels of person-fit statistics with constant sample sizes and test lengths. In this part of the study, item parameter estimates for each group were obtained twice using male and female students, and then the Pearson correlation coefficient between item parameter estimates was calculated for each group. When the estimates are invariant, the correlation coefficient should be high. Invariance is established when the estimates do not differ in excess of measurement errors associated with the estimates.

Table 6 provides comparisons between item difficulty estimates obtained from male and female groups for every Lz-subgroup. The correlations between two item difficulty estimations based on male and female students and their p-values are given in the last two columns in Table 6.

Table 6
Correlations between Difficulty Estimates Obtained from Male and Female Examinees For Every Lz-Subgroups

Group No	Correlation	P
Group 1	0.92	<0.0001
Group 2	0.86	<0.0001
Group 3	0.85	<0.0001
Group 4	0.68	<0.0001
Group 5	0.84	<0.0001
Group 6	0.87	<0.0001
Group 7	0.91	<0.0001
Group 8	0.94	<0.0001
Group 9	0.92	<0.0001
Group 10	0.96	<0.0001
Control Group	**0.93**	**<0.0001**

As seen in Table 6, the correlation coefficient for the control group is .93, and the maximum correlation coefficient for the groups is .96 in the same group where the person-fit index is the highest (mean=1.27). This is even higher than that of the control group. The correlation coefficients were low in the groups with negative l_z statistics, while the correlation coefficients were high, even higher than the control group, in the groups with positive l_z statistics. Groups 1, 2, 3, 4, 5, and 6 had negative person-fit statistics and low correlation

coefficients. On the other hand, groups 7, 8, 9, and 10 had positive l_z statistics and high correlation coefficients, even higher than that of the control group. These correlation coefficients clearly show that the invariance properties of the Rasch model item difficulty estimates increase when person-fit statistics are bigger than zero. That suggests that if person-fit statistics are bigger than zero, one can get a much more accurate item difficulty parameter estimate. Further, that might mean that students who have negative l_z statistics do not fit the model very well. In the current use of IRT models, if the Pearson Chi square lack-of-fit statistic for any item in the test is bigger than the degree of freedom, that item is eliminated from the test since it does not fit the model. In the same way, Tables 3 and 6 suggest we eliminate the students who have l_z indexes that are greater than zero to estimate ability more accurately along with the difficulty parameter. In other words, by eliminating students who answered the item on the test by cheating, guessing, sleeping, and so on, more accurate item and ability parameter estimates can be obtained, since parameter estimates are affected by aberrant responses.

Standard Errors Associated with Each Item Difficulty Estimate

An asymptotic estimate of the standard errors associated with each item difficulty parameter estimate was determined for every group as an index of the influence of person-fit statistics. Since the most direct factor that affects the size of the standard error values is size of the data set, the sample size and test length for every group were held constant (around 500). Standard errors associated with each item difficulty estimate and summary statistic are given in Tables 7 and 8. Note that the first column in Table 7 is the item number, and N in Table 8 is the total number of items in the corresponding group. In table 7, since zero and perfect scores are eliminated by the Rasch model, item 20 for all groups, and items 19 and 12 and 8 for some groups were eliminated from the test. That is why N in Table 8 is varied across the groups. There is no Pearson Chi-square lack of item-fit statistic for these items in Table 9.

Table 7, standard errors associated with every item difficulty, and the summary statistics in Table 8 clearly show that when person-fit statistics increase, standard errors associated for every item difficulty estimation also increase. It can easily be seen that the largest mean errors were 0.315, 0.195, and 0.165 in groups 8, 9, and 10, where person-fit statistics are the largest, even though the mean for the control group was 0.133. In other words, standard errors associated with maximum likelihood difficulty parameter estimate were very low in the groups with negative l_z statistics, while standard errors associated with maximum likelihood difficulty parameter estimate were very high, even higher than the control group, in the groups with positive l_z statistics. Groups 1, 2, 3, 4, 5, and 6 had negative person-fit statistics and small standard errors associated with maximum likelihood difficulty parameter estimates. On the other hand, groups 7, 8, 9, and 10 had positive l_z statistics and larger standard errors associated with maximum likelihood difficulty parameter estimates which were even larger than the the control group. Tables 7 and 8 clearly demonstrate that standard errors associated with maximum likelihood difficulty parameter estimates increase when person-fit statistics are greater than zero. In summary, it can be concluded that ability parameter estimates are affected positively by students who answer the test items by cheating, guessing, sleeping, and so on, but standard errors associated with each item parameter estimate increase. As mentioned in the previous section regarding ability parameter estimates, the reason that standard errors associated with item parameter estimates increase might be related to their ability distributions. Grouping students according to aberrant response behaviors creates homogeneous ability distributions. Satisfactory item and ability estimation requires heterogenous ability distributions.

Pearson Chi-square Lack-of-fit Statistics

To estimate the relationship between item-lack-of-fit statistics and person-fit statistics, Pearson Chi-square lack-of-fit (model-fit) statistics and their summary statistics are given in Tables 9 and 10 respectively, for each group based on varied person-fit statistics.

High and statistically significant Chi-square values suggest that these items are not adequately described by the Rasch model. At the end of Table 9, degrees of freedom, critical values, and the number of miss-fit items with significant Chi-square values (miss-fit) are also given for each group based on different person-fit statistics. The interest here was in determining the impact of aberrant response behaviors on the Pearson lack-of-fit statistics (Miss-Fit Items).

In Table 9, degrees of freedom, critical values, and the number of miss-fit items in the group are given for all groups. For instance, in the control group, there are 19 items, 15 degrees of freedom, and 2 miss-fit items. Also summary statistics related to Pearson Chi-square lack-of-fit are given in Table 10. As seen in Tables 9 and 10, the mean of Chi-square statistics in the control group is almost the smallest one, 15.47; and Group 1, which has the smallest l_z statistic, has the smallest mean of Pearson Chi-square lack-of-fit. On the other hand, group 3 has the largest Pearson Chi-square lack-of-fit, and almost every item is miss-fit in that group. It seems that the Chi-square statistic is not influenced by person-fit statistics in a certain way. Basically, it does not seem there is any correlation between aberrant responses and miss-fit items.

Table 7
Standard Errors Associated with Each Item Parameter Estimation for Every Group Based on Different Person-Fit Statistics

Item	GRP1	GRP2	GRP3	GRP4	GRP5	GRP6	GRP7	GRP8	GRP9	GRP10	CONT.GR
1	0.146	0.107	0.097	0.099	0.100	0.105	0.115	0.119	0.111	0.140	0.108
2	0.215	0.133	0.119	0.099	0.106	0.118	0.122	0.130	0.144	0.206	0.124
3	0.105	0.102	0.112	0.110	0.113	0.125	0.131	0.169	0.147	0.171	0.121
4	0.101	0.147	0.148	0.132	0.150	0.225	0.197	0.224	0.253	0.428	0.163
5	0.108	0.101	0.106	0.108	0.113	0.123	0.145	0.148	0.147	0.160	0.121
6	0.139	0.103	0.095	0.099	0.100	0.105	0.116	0.121	0.111	0.137	0.107
7	0.189	0.125	0.110	0.098	0.104	0.106	0.122	0.119	0.137	0.156	0.115
8	0.241	0.218	0.131	0.115	0.126	0.144	0.147	0.262	.	.	0.187
9	0.101	0.133	0.159	0.134	0.153	0.210	0.192	0.240	0.223	0.608	0.160
10	0.107	0.103	0.112	0.108	0.114	0.126	0.136	0.169	0.144	0.172	0.121
11	0.101	0.140	0.160	0.128	0.145	0.204	0.182	0.212	0.303	0.490	0.157
12	0.139	.	0.100	.	.	.	.	.	.	.	0.117
13	0.185	0.102	0.095	0.100	0.101	0.107	0.116	0.124	0.111	0.137	0.108
14	0.102	0.122	0.109	0.098	0.103	0.107	0.125	0.119	0.133	0.153	0.117
15	0.208	0.148	0.154	0.138	0.161	0.207	0.192	0.228	0.220	0.995	0.163
16	0.248	0.141	0.118	0.103	0.104	0.115	0.121	0.129	0.140	0.269	0.125
17	0.099	0.168	0.128	0.106	0.120	0.112	0.161	0.123	0.527	.	0.142
18	0.102	0.114	0.143	0.118	0.137	0.133	0.233	0.157	0.220	0.224	0.137
19	.	0.127	0.150	0.119	0.140	0.174	0.214	0.187	0.253	0.608	0.145
20	.	.	.	.	.	.	.	.	.	.	.

Table 8
Summary Statistics For Standard Errors Associated
For Every Item Difficulty Estimation

Groups	N	Mean	Std Dev	Minimum	Maximum
GROUP 1	18	0.146	0.053	0.099	0.248
GROUP 2	18	0.129	0.029	0.101	0.218
GROUP 3	19	0.123	0.022	0.095	0.160
GROUP 4	18	0.111	0.013	0.098	0.138
GROUP 5	18	0.121	0.020	0.100	0.161
GROUP 6	18	0.141	0.042	0.105	0.225
GROUP 7	18	0.153	0.038	0.115	0.233
GROUP 8	18	0.165	0.048	0.119	0.262
GROUP 9	17	0.195	0.103	0.111	0.527
GROUP 10	16	0.315	0.246	0.137	0.995
CONTROL GRP	**19**	**0.133**	**0.023**	**0.107**	**0.187**

Table 9
Pearson Chi-Square Lack-Of-Fit (Miss-Fit Items) According to Groups Based on Different Person-Fit Statistics

Item	GRP1	GRP2	GRP3	GRP4	GRP5	GRP6	GRP7	GRP8	GRP9	GRP10	CONT.
1	8.41	13.24	19.80	13.03	74.90	55.86	28.90	36.38	42.02	55.25	14.78
2	16.30	18.70	24.62	16.33	16.95	47.30	14.69	52.96	23.23	34.89	12.62
3	2.40	19.25	35.07	28.25	26.58	14.93	23.27	16.76	9.183	8.25	16.73
4	15.61	7.79	117.87	7.90	25.24	18.26	19.43	13.50	14.19	6.09	19.42
5	10.75	21.70	21.53	31.63	13.28	29.67	25.93	28.68	61.10	12.76	22.56
6	10.24	17.83	27.13	28.77	23.85	34.65	23.10	17.64	26.06	10.82	11.98
7	3.79	18.01	21.33	9.72	9.67	16.79	19.73	27.33	63.46	29.87	6.93
8	3.44	18.04	46.14	11.94	29.83	32.12	10.10	76.12	.	.	7.22
9	13.24	43.86	133.92	6.51	14.37	11.32	12.51	11.24	17.56	3.10	13.53
10	4.43	9.32	89.32	29.30	13.33	14.89	22.51	7.70	18.04	10.63	31.96
11	8.59	11.39	48.38	10.06	16.27	9.74	16.10	22.46	15.01	5.59	7.21
12	7.39	.	134.08	.	.	.	.	.	.	.	11.16
13	16.69	8.55	25.36	16.96	94.62	22.55	30.58	57.99	38.42	45.75	19.22
14	11.71	21.67	19.16	46.92	16.73	21.89	9.98	21.60	29.26	32.98	11.49
15	8.57	22.10	98.79	13.36	12.67	10.57	19.57	23.14	17.56	2.60	29.67
16	3.37	24.32	16.70	6.43	9.07	8.49	24.97	46.98	17.55	17.02	12.74
17	2.85	15.42	36.11	15.18	13.98	51.08	100.69	83.69	202.56	.	14.52
18	38.91	12.26	156.97	18.12	15.31	40.69	107.40	100.42	112.23	29.11	15.48
19	.	10.84	81.80	17.67	28.68	12.17	11.17	13.30	18.62	5.18	14.69
20	.	.	.	.	.	.	.	.	.	.	.
Df	7	8	10	14	15	14	14	15	14	11	15
CV	14.07	15.51	18.31	23.68	25.00	23.68	23.68	25.00	23.68	19.68	25.00
Misfit	4	10	18	5	5	7	6	9	8	6	2

Table 10
Summary Statistics of Pearson Chi-Square Lack-of-Fit

Variable	N	Mean	Std Dev	Minimum	Maximum
GROUP 1	18	10.376	8.546	2.407	38.910
GROUP 2	18	17.466	8.315	7.795	43.866
GROUP 3	19	60.745	46.974	16.707	156.978
GROUP 4	18	18.230	10.749	6.430	46.923
GROUP 5	18	25.300	22.788	9.077	94.625
GROUP 6	18	25.170	15.219	8.495	55.864
GROUP 7	18	28.928	28.043	9.985	107.403
GROUP 8	18	36.554	27.383	7.706	100.421
GROUP 9	17	42.712	48.606	9.183	202.565
GROUP 10	16	19.372	16.433	2.601	55.252
CONTROL GRP	19	15.472	6.800	6.937	31.967

Person-fit statistics and test information function

This part of the study provides estimates of the relationship between person-fit statistics and the test information function. In other words, the interest was to look at the impact of aberrant responses on the test information function. Test information function is sometimes referred to as the test information function curve, and test response function is sometimes referred to as the test characteristic curve for every group based on different person-fit statistics. They are not given here graphically. Instead, two types of averages, expected and average information obtained from the test information plot and estimated reliability values based on the conditional standard error of measurement provided by the IRT model, are given according to different groups based on person-fit statistics in Table 11. The expected information is the average amount of information that a test would provide for a sample of examinees drawn from a population with a normal ability distribution. The average information is the average height of the plotted information curve. The average information is thus the amount of information that a test provides for a sample of examinees with abilities that are rectangularly distributed. Estimated reliability can be interpreted as an estimate of the conventional alpha (KR-20) internal-consistency reliability.

Table 11
Test Information Functions for Every Group Based on Different Person-Fit Statistics

Group No	Expected Information	Average Information	Estimated Reliability
Group 1	2.452	2.247	0.709
Group 2	2.452	2.250	0.710
Group 3	2.740	2.469	0.732
Group 4	3.144	2.530	0.754
Group 5	2.748	2.391	0.732
Group 6	2.392	2.207	0.705
Group 7	2.106	2.041	0.677
Group 8	1.772	1.827	0.638
Group 9	1.410	1.527	0.584
Group 10	0.805	0.818	0.443
Control Group	0.915	1.034	0.473

The expected information, which is the average amount of information that a test would provide for a sample of examinees drawn from a population with a normal ability distribution, is at a minimum in groups 7,

8, 9 and 10, which have the biggest and positive person-fit statistics. On the other hand, the expected information is at maximum in groups 1, 2, 3, 4, 5, and 6, which have the smallest and negative person-fit statistics. It seems that as person-fit statistics increase, the expected information decreases. In the same way, the average information, which is the average height of the plotted information curve, increases as person-fit statistics also decrease. For instance, average information is at a maximum in group 10, which has the largest person-fit statistic. Estimated reliability, which can be interpreted as an estimate of the conventional alpha (KR-20) internal-consistency reliability, also decreases when person-fit statistics increase. For example, in group 1, which has the smallest negative person-fit statistics, estimated reliability is at a maximum. But the best result in terms of the expected-average information and estimated reliability is obtained in group 1 as seen in Table 11. The results show the reason internal consistency reliability decreases is that grouping students according to aberrant response behaviors creates homogeneous ability distributions. Satisfactory item and ability estimation with high internal consistency reliability requires heterogenous ability distributions. But it can be said that the person-fit statistics influence the expected-average information and estimated reliability negatively. Grouping examinees in terms of person-fit statistics may create a skewed distribution of total scores that make reliability computations questionable for each group. But the control group is normally distributed, and when compared with group 1, it is clear that expected-average information and estimated reliability are affected negatively by person-fit statistics.

Conclusion and Practical Implications of Study

The goal of this study was to investigate the impact of the aberrant response behaviors such as cheating, guessing, and sleeping on the item and ability parameter estimates in the Rasch model. More specifically, the purpose of this study was to investigate how seriously person-fit statistics affect: (1) invariance of ability parameter estimates, (2) standard errors associated with IRT maximum likelihood ability estimates, (3) invariance of item parameter estimates, (4) standard errors associated with each item difficulty estimates, (5) Pearson Chi-square lack-of-fit statistics, and (6) test information function. Results of this study indicated that: (1) The invariance properties of the Rasch model ability estimates increase when person-fit statistics are bigger than zero. That means that if person-fit statistics are greater than zero, we can get much more accurate ability parameter estimates. (2) It seems that as person-fit statistics increase, standard errors associated with ability parameter estimates also increase. (3) The invariance properties of the Rasch model item parameter estimates increase as person-fit statistics increase positively. This suggests that if person-fit statistics are greater, we can get much more accurate item parameter estimates. (4) When person-fit statistics increase, standard errors associated with every item difficulty estimation also increase. (5) The Chi-square statistic is not influenced by person-fit statistics, and (6) it seems that as person-fit statistics increase the expected information decreases. In the same way, the average information, that is the average height of the plotted information curve, increases as person-fit statistics decrease. Estimated reliability, which can be interpreted as an estimate of the conventional alpha (KR-20) internal-consistency reliability, also decreases when person-fit statistics increase.

As a practical implication of this study, it can be said that students who have negative l_z statistics do not fit the model very well. In the current use of the Rasch Model, if the Pearson Chi-square lack-of-fit statistic for any item in the test is greater than the degrees of freedom, that item is eliminated from the test since it does not fit the model. This study tells us to eliminate students who have l_z indices that are smaller than zero to make more accurate parameter estimates. But, in that case, standard errors associated with maximum likelihood item difficulty and ability parameter estimates increase and expected information and estimated reliability decrease in a certain way.

Generally speaking, this study shows test developers that invariance of ability and item parameter estimates are affected positively by the aberrant response behaviors in the Rasch model. In other words, by excluding students who answer the test items by cheating, guessing, and so on, more accurate parameter estimates can be obtained even though standard errors associated with parameter estimates increase and test information function decreases. This study shows that by excluding miss-fit items from the test and excluding students who have aberrant response behavior, item- and ability-parameter invariance can be obtained in the Rasch model. The study also shows that satisfactory item- and ability-parameter estimation requires heterogeneous ability distributions. The results show that internal consistency reliability decreases because grouping students according to aberrant response behaviors creates homogeneous ability distributions. Satisfactory item and ability estimation with high internal consistency reliability requires heterogenous ability distributions. But the control group is normally distributed, and when compared with group 1, it can be said that expected-average information and estimated reliability are affected negatively by person-fit statistics.

As a general conclusion, it can be recommended to test developers that excluding aberrant response behaviors with the normal and heterogenous ability distribution results in more accurate and invariant parameter estimates

in the Rasch model. Since excluding aberrant responses can make reliability computations questionable, test developers should be careful about heterogenous ability distribution while excluding aberrant response behaviors to get more accurate and invariance parameter estimates.

References

Andersen, E. B. (1973). A goodness of fit test for the Rasch model, *Psychometrika, 38*, 123-140.

Bock, R. D. (1972). Estimating item and parameters and latent ability when responses are scored in two or more nominal categories, *Psychometrika, 37*, 29-51.

Drasgow, F., Levine, M. V., & McLaughlin, M. L. (1987). Detecting inappropriate test scores with optimal and practical appropriateness indices. *Applied Psychological Measurement, 11*(1), 59-79.

Drasgow, F., Levine, M. V., & Williams, E. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology, 38*, 67-86.

Drasgow, F. (1982). Choice of test model for appropriateness measurement, *Applied Psychological Measurement, 6*, 297-308,

Gustafsson, J. E. (1980). Testing and obtaining fit of data to the Rasch Model. *British Journal of Mathematical and Statistical Psychology, 33*, 205-233.

Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). *Item Response Theory: Application to Psychological Measurement.* Homewood, IL: Dow Jones-Irwin.

Levine, M. V., & Drasgow, F. (1982) Appropriateness measurement: Review, critique, and validating studies. *British Journal of Mathematical and Statistical Psychology, 35*, 42-56.

Levine, M. V., & Drasgow, F. (1984). *Performance envelopes and optimal appropriateness measurement* (report No: 84-5). Champaign IL: University of Illinois, Department of Educational Psychology, Model Based Measurement Laboratory. (ERIC Document Reproduction Service No: ED 263 126)

Levine, M. V., & Rubin, D. F. (1979) Measuring the appropriateness of multiple choice test scores. *Journal of Educational Statistics, 4*, 269-290.

McKinley, R. L., & Mills, C. N. (1985). A comparison of several goodness-of-fit statistics, *Applied Psychological Measurement, 3*, 371-375.

Meijer, R. R. (1996). Person-fit research: An introduction. *Applied Measurement in Education, 9*, 1, 3-8.

Reise, S. P. (1990). A comparison of item and person fit methods of assessing model data fit in IRT, *Applied Psychological Measurement, 14*(2), 127-137.

Roger, H. J., & Hattie, J. A. (1987). A Monte Carlo investigation of several person and item fit statistics for item response models. *Applied Psychological Measurement, 11*(1), 45-57.

Shepard, L. A., Camilli, G. & Williams, D. M. (1984). Accounting for statistical artifacts in item bias research, *Journal of Educational Statistics, 9*, 93-128.

Troub, R. E., & Lam, Y. R. (1985). Latent structure and item sampling models for testing, *Applied Psychological Measurement, 36*, 19-48.

Van den Wollenberg, A. L. (1982). Two new test statistics for the Rasch model, *Psychometrika, 47*, 123-140.

Waller, M. I. (1981). A procedure for comparing logistic latent trait models, *Journal of Educational Measurement, 18*,119-25.

Wright, B. D. (1968). *Sample free test calibration and person measurement. Proceedings of the 1967 invitational conference on testing problems.* Princeton, NJ: Educational Testing Service.

Wright, B. D., & Panchhapakesan, N. (1969). A procedure for sample free item analysis, *Educational Psychological Measurement, 29*, 23-48.

Wright, B. D., & Stone, M. H. (1979). *Best Test Design.* Chicago: Mesa Press.

Yen, W. M. (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, 5,: 245-62.

RESEARCH IN THE SCHOOLS
1999, Vol. 6, No. 2, 75-92

An Introductory Primer on the Appropriate Use of Exploratory and Confirmatory Factor Analysis

Kevin M. Kieffer
Texas A & M University

Factor analysis has historically been used for myriad purposes in the social and behavioral sciences, but an especially important application of this technique has been to evaluate construct validity. Since in the present milieu exploratory factor analysis (EFA) and confirmatory factor analysis (CFA) are both commonly employed in analyzing data, a firm command of the theoretical underpinnings of both techniques is essential. Despite the fact that both techniques belong to the same general linear model, there are important differences between the two analytic traditions. The present paper provides an introduction to both techniques, presents heuristic data to compare the two analytic methods, and discusses analytic traditions in EFA and CFA analyses by examining recently published exemplars of both practices.

The utilization of factor analytic techniques in the social sciences has been indelibly intertwined with developing theories and evaluating the construct validity of measures. As stated by Gorsuch (1983, p. 350), "A prime use of factor analysis has been in the development of both the theoretical constructs for an area and the operational representatives for the theoretical constructs." Since factor analysis has been deemed the "reigning queen of the correlational methods" (Cattell, 1978, p. 4), it is not surprising that Pedhazur and Schmelkin (1991, p. 66) noted that "of the various approaches to studying the internal structure of a set of variables or indicators, probably the most useful is some variant of factor analysis."

The majority of researchers utilizing factor analytic techniques have employed what are contemporaneously termed "exploratory" factor analytic techniques (EFA). In this application of factor analysis, researchers are primarily concerned with the development of theories about a phenomenon of interest. More recently, however, a growing number of researchers have been employing "confirmatory" factor analytic techniques (CFA) that directly permit the testing of extant theories or the evaluation of instrument structure based on theoretical expectations. A hybrid of these two techniques utilizes exploratory factor extraction in combination with confirmatory factor rotation (cf. Thompson, 1992); this hybrid, however, is not considered in the present article.

I gratefully acknowledge the assistance of Professors Bruce Thompson and James E. McLean for their helpful and insightful comments on an earlier version of this article. Kevin M. Kieffer is currently a doctoral candidate in the counseling psychology program at Texas A&M University. Correspondence regarding this article should be sent to Kevin M. Kieffer, College of Education, Department of Educational Psychology, College Station, TX 77843-4225 or e-mailed to kmkieffer@earthlink.net.

The purpose of the present article was to briefly explicate both analytic techniques using a small heuristic data set that permits readers to grasp the conceptual underpinnings of both methods. The two techniques are directly compared using the Holzinger and Swineford (1939) data set that has been utilized extensively in illustrating factor analytic principles (cf. Gorsuch, 1983). The data set consists of the scores of 301 junior high school students on 24 different psychological inventories. For the purpose of the present heuristic example, however, only 13 of the original 24 tests were utilized. In addition to the comparison of EFA and CFA analytic techniques using the Holzinger and Swineford data set, various published exemplars of both approaches from the psychology literature were examined as well.

Exploratory Factor Analysis

Historical Overview

The genesis of the analytic technique commonly referred to as exploratory factor analysis (as well as its confirmatory variant) can be traced back to the seminal work of Pearson (1901) and Spearman (1904). Pearson (1901) first conceptualized a general method of examining and extracting latent variables underlying data structures. Spearman, through his work on personality theory, provided the conceptual and theoretical rationale for both exploratory and confirmatory factor analysis. Despite the fact that the conceptual bases for these methods have been available for many decades, it was not until the widespread availability of both the computer and modern statistical software that these analytic techniques were employed with any regularity.

A renewed interest in factor analytic techniques was evidenced, however, in the middle of this century following a schism within the American Psychological

Association (APA). Many practitioners abandoned the flagship organization of psychologists during the early 1940s due to an overemphasis on research interests and a paucity of practice-related pursuits (Thompson & Daniel, 1996). As a precondition for returning to the APA, these practitioners seeking professional status demanded that standards for conduct--including test standards--be articulated by the APA.

Prior to the first articulation of official testing standards in the 1950s, psychologists and educators did not yet recognize the current three types of validity. Indeed, a type of validity termed "factorial validity" was recognized by some prominent scholars at the time:

> The factorial validity of a test is given by its loadings in meaningful, common reference factors. This is the kind of validity that is really meant when the question is asked "Does this test measure what it is supposed to measure?" (Guilford, 1946, p. 428)

Following the test standards developed by APA, many researchers and test constructors began utilizing factor analytic techniques to demonstrate the validity of scores generated by their instruments. The true power of factor analytic techniques, however, was not realized as a practical matter until more recently when the computer and modern statistical software packages became available.

Theoretical Underpinnings

Exploratory factor analysis (EFA) is an analytic technique in which the primary concern is to reduce a larger set of variables into a smaller and more manageable set based on the consistency of the data. As indicated by Gorsuch (1983, p. 2),

> The purpose of factor analysis is . . . to summarize the interrelationships among variables in a concise but accurate manner as an aid in conceptualization. This is often achieved by including the maximum amount of information from the original variables in as few derived variables, or factors, as possible to keep the solution understandable.

Gorsuch (p. 90) further stated, "Reducing the number of variables to a more reasonable subset is often a prime goal of factor analysis. A reduced variable set is sought that will contain as much of the information in the initial set of variables as possible."

EFA has historically been utilized for two general purposes in the social sciences (Pedhazur & Schmelkin, 1991). The first general purpose has been to better understand the internal structure of an instrument or a data set when no previous information on the data structure is available. This is an example of utilizing EFA as a tool in the generation of theories about phenomena of interest, and is an example of an appropriate application of this analytic technique. The second general application of EFA has been to re-examine patterns in data sets when researchers have questioned the tenability of the emergent factors in previous research. Many researchers have utilized EFA to determine if the data structure in a current study resembles the factor structure of previous research on the same phenomenon of interest or when utilizing the same instrument. In this second application, many researchers have erroneously employed EFA to evaluate whether a given theory adequately fits a data set by examining if the same general factors reemerge in subsequent studies. This purpose is often more appropriately explored through confirmatory factor analysis, as model-to-data fit can be directly evaluated in the confirmatory case. The more common and appropriate of these two applications of EFA has been in exploring the factor structure of a set of indicators (e.g., variables, test items, individuals, occasions) when no previous research is available.

EFA is not conceptually different from other statistical techniques, as all analytic techniques are correlational (Cohen, 1968; Thompson, 1997a) and are part of one general linear model (GLM) subsumed by canonical correlation analysis (Knapp, 1978). Like other analyses, the goal of conducting an EFA is to explain the maximum amount of shared variance with the fewest explanatory concepts. The conceptual unity of factor analysis with canonical correlation analysis (and thus all other analyses) was illustrated by Hetzel (1996, p. 178):

> [S]ince all parametric methods are actually special cases of canonical correlation analysis . . . and since canonical correlation analysis itself invokes a principal components analysis [as factor analysis does] . . . [a]ll parametric methods actually invoke principal components mathematics.

Thus, EFA is not conceptually dissimilar from other statistical analyses as all analyses attempt to explain shared variance through principal components mathematics. The manner in which EFA permits the examination of shared variance among a set of indicators, however, is somewhat different across analyses. Thus, to understand the conceptual and practical distinctions between exploratory and confirmatory factor analysis, it is critical to understand the terminology and mechanics of both techniques.

Logic and Mechanics of EFA

Performing an EFA can be conceptualized as a series of steps which require that certain decisions be addressed

at each individual stage. Consequently, there are many different ways in which to conduct an EFA, and each different approach may render distinct results when certain conditions are satisfied (cf. Gorsuch, 1983). The one consistent element in conducting an EFA, however, is that the results of the analysis are based solely on the mechanics and mathematics of the method and *do not* consider the a priori theoretical considerations of the researcher (Daniel, 1989).

Determination of Sample Size. When performing an EFA, it is necessary to collect data from many participants on many variables of interest. But how many participants and variables are enough? As noted earlier, the purpose of factor analysis is to reduce the complex nature of large data sets to smaller, more manageable dimensions. Thus, it is likely that a given researcher will want to simplify the structure of many variables (i.e., test items, psychological tests) involved in an analysis. Because there is no set number of variables to reduce, EFA can theoretically be used on any number of variables that a researcher desires to explore. As for sample size, authors have often debated the number of participants needed to conduct an analysis. As with almost any analytic technique, the more participants that are surveyed, the more stable the results of the analysis (because the sample size approaches the size of the population). Tabachnick and Fidell (1983, p. 379) developed a hierarchy ranging from poor to excellent in which a sample size of 200 or below was labeled "poor" to conduct an EFA, 300 was labeled "good," and 1,000 participants was "excellent." Stevens (1996, p. 372) suggested that the appropriate sample size to employ is a function of the number of participants per variable (ranging from five to 20 participants per variable) and the component saturation (in which components that have four or more coefficients greater the |.60| are reliable regardless of sample size). Based on these previous recommendations, therefore, a minimum of 300 participants or at least five participants per variable is generally necessary to conduct the analysis with some degree of confidence in the reliability of the variable scores and resultant factors.

Matrices of Association. After determining an appropriate sample size, one of the first decisions to be made in performing an EFA is to determine the manner in which the data will be represented in the analysis. Since all statistical analyses are correlational, the focus of every statistical analysis is on the relationship among a set of variables or other entities (e.g., people) that may be factored. However, many matrices of association (e.g., correlation matrices, variance-covariance matrices) can be utilized for this purpose. The most common matrices of association utilized in EFA are correlation matrices (in which values of 1.0 are on the main diagonal and bivariate correlation coefficients between the variables are off-diagonal), perhaps partly because this is the default in most statistical software packages.

Factor Extraction. After the researcher has chosen which matrix of association will be utilized in the analysis, the researcher must then determine which extraction method to employ in conducting the analysis. Factor extraction refers to removing the common variance that is shared among a set of variables. There are currently several different techniques available for the extraction of common variance, and the results generated by the analysis can differ based on the particular method of extraction utilized.

Of the techniques available, principal components analysis and principal factors analysis are the two most widely used extraction methods in EFA. Although some researchers have argued that the difference between these extraction methods is negligible (cf. Thompson, 1992), other researchers have contended that the difference is substantial enough to warrant careful consideration of the extraction method utilized (cf. Gorsuch, 1983). As noted by Thompson and Daniel (1996),

> Analysts differ quite heatedly over the utility of principal components as compared to common or principal factor analysis. . . . The difference between the two approaches involves the entries used on the diagonal of the matrix of association that is analyzed. When a correlation matrix is analyzed, principal components analysis uses ones on the diagonal whereas common factor analysis uses estimates of reliability, usually estimated through an iterative process. (p. 201)

Although the relative advantages and disadvantages of the two extraction methods have been discussed elsewhere (cf. Gorsuch; Stevens, 1996; Tinsley & Tinsley, 1987), factor extraction methods demand brief consideration in the present article.

Principal components analysis (PCA) uses the total variance of each variable in examining the shared variance between variables (Hetzel, 1996). As noted previously, this is accomplished by placing values of 1.0 on the main diagonal of the correlation matrix (as each variable is expected to correlate perfectly with itself) and leaving the bivariate correlation coefficients on the off-diagonals. One limitation in utilizing PCA is that as the number of factored variables decreases or as the factored variables become less reliable "some of the factors represent correlated error variance and as such would be unlikely to be replicated in subsequent studies" (Hetzel, p. 186). Conversely, the differences associated with utilizing PCA versus other extraction methods becomes negligible (a) as

the number of factored variables increases and (b) as the factored variables become more reliable (Thompson & Daniel, 1996).

Snook and Gorsuch (1989, p. 149) explained this first influence (i.e., the number of factored entities), noting that "As the number of variables decreases, the ratio of diagonal to off-diagonal elements also decreases, and therefore the value of the communality has an increasing effect on the analysis." For example, with five factored variables the five diagonal entries in the correlation matrix represent 20% (5 / 25) of the 25 entries in the matrix, but with 50 variables the diagonal entries represent only 2% (50 / 2,500) of the 2,500 matrix entries. Thus, Gorsuch (1983) suggested that with 30 or more variables the differences between solutions from these two methods (PCA versus PFA) are likely to be small, and each will lead to similar interpretations. Of course, researchers rarely factor data involving only a small number of variables, because small data sets are already manageable and understandable. Thus, the differences between these two extraction methods are often negligible in actuarial research situations.

In principal factors analysis (PFA), an estimate of the reliability of the scores on each variable is placed on the main diagonal of the correlation matrix rather than a value of unity (i.e., 1.0). The reason for restructuring the correlation matrix is due to the assumption that only the common or reliable variance indigenous to a variable will correlate with other variables in the matrix (Gorsuch, 1983). Thus, because placing a value of unity on the main diagonal would introduce measurement error variance into the factor extraction procedure, proponents of PFA have contended that the most appropriate value to place on the main diagonal of the correlation matrix is some index of the common or reliable variance of variable scores. Several indices of common or reliable score variance can be utilized on the main diagonal of the correlation matrix, and the interested reader is referred to Gorsuch, Stevens (1996), or Tinsley and Tinsley (1987) for a more thorough explanation of this topic.

However, to understand the differences associated with placing a value of unity (as in PCA) versus placing an estimate of common or reliable variance on the main diagonal (as in PFA), it is important to discriminate between the concepts of total variance, common variance, unique variance, and error variance. Weiss (1971, p. 85) noted that,

> The general factor-analytic model assumes that the total variance of a variable is composed of three components: common variance, unique variance, and error variance. For each variable included in the factor analysis of a correlation matrix, the mix of these three components may be different, but the sum of the components is the same--1.0.

Common variance, as defined by Weiss, is the reliable variance that a particular variable has in common with other variables (the so-called communality), unique variance is reliable variance indigenous to a specific variable, and error variance is unreliable variance (e.g., measurement error variance, sampling error variance) that is unlikely to occur in future samples.

As noted earlier, one main difference between PCA and PFA extraction procedures is the value that is placed on the diagonal in the correlation matrix analyzed in the procedure. Gorsuch (1983, p. 102) noted that,

> the extreme limits that the communalities [common variance] could take are, on the one hand, 0.0 if the variable has no correlation with any other variables in the matrix and, on the other hand, 1.0 if the variance is perfectly accounted for by the set of factors underlying the matrix.

PCA places a value of unity (the upper bound communality estimate) on the main diagonal which renders the possibly untenable assumption that each variable is perfectly reliable (because total variance is comprised of common, unique, and error variance, and this value is *set* equal to 1.0). PFA, however, places an *estimate* of common or reliable variance on the main diagonal (generated through an iterative process), and this index of common variance is contrived through a number of estimation procedures such as reliability estimates, squared multiple correlation, or highest bivariate correlation (Gorsuch; Weiss, 1971).

As noted by Mulaik (1966), however, the actual upper bound of the communality is the reliability coefficient of a given variable (because only reliable variance will correlate with other variables); consequently, placing a value of unity on the main diagonal of a correlation matrix will have the effect of introducing into the analysis measurement or sampling error that may not be replicated in future studies. This dynamic can influence the results of the factor extraction procedure, but, as noted previously, the effect of placing a value of unity versus a communality or reliability estimate on the main diagonal will decrease as the number of factored variables increases and as the factored variables become more reliable (Thompson & Daniel, 1996).

Factor and Coefficient Generation. After the researcher has decided which factor extraction method to employ, the analysis can be conducted. One advantage to employing a factor analysis is that each latent or synthetic variable (factor) extracted from the analysis is perfectly

uncorrelated with all of the other factors. This is often advantageous when the purpose of the EFA is theory generation as the interpretation of the extracted factors is thereby greatly simplified.

The extracted factors represent an attempt by the researcher to mathematically re-express the relationships between a set of variables with the fewest explanatory concepts possible. By doing this only a certain portion of the variance for any given variable will be reproduced by the factors (although it is theoretically possible for all of the variance associated with a variable to be reproduced by the factors, this rarely occurs in practice). The resultant mathematical manipulations required to extract the factors result in the formation of three matrices, any two of which determine the third: the factor pattern matrix, the factor structure matrix, and the factor correlation matrix. Regardless of the type of extraction method, typically the rows of the factor pattern and structure matrices are composed of the variables in the study, and the columns of the matrices are composed of the latent constructs or factors.

The factor pattern matrix is comprised of a series of weights (mathematically identical to ß weights in multiple regression analysis) that indicate the relative importance of a given variable to the extracted factors with the influence of the other variables removed (Stevens, 1996). The factor structure matrix, however, is simply the bivariate correlation of a measured/observed variable with scores on the extracted latent/synthetic factor. Factor structure coefficients are identical to structure coefficients in other analyses as they are simply the correlation between observed and latent variables (Thompson, 1997b; Thompson & Borrello, 1985). The extracted factors are always perfectly uncorrelated, which results in the equality of the factor pattern matrix and the factor structure matrix. Since the two matrices are equivalent, both matrices can be accounted for by employing the term, "factor pattern/structure matrix," to describe a single matrix.

After the factor pattern/structure matrix is contrived, it is possible to generate two variance-accounted-for statistics that help the researcher determine the amount of variable variance that is reproduced by the latent constructs. The first of these is the communality coefficient, h^2, which can be defined as the amount of variance on a variable that is reproduced by the factors. This value is calculated by summing the squared pattern/structure coefficients across the row for each variable. Since h^2 is a squared, variance-accounted-for statistic, it can range from 0 to 1.0.

Another variance-accounted-for statistic that is generated from the factor pattern/structure matrix is the eigenvalue. An eigenvalue represents the amount of variance in the original data set that is reproduced by a given factor. For the principal components case, eigenvalues can be computed by summing the squared factor pattern/structure coefficients down the columns of the matrix. Eigenvalues represent the amount of factor-reproduced variance, and their values can range from 0.0 to the total number of variables in the analysis. Eigenvalues can also serve as an effect size measure, as each eigenvalue can be divided by the number of total variables in the analysis, thus indicating the percentage of the total variance that a given factor represents or can reproduce. Due to the mathematics of EFA, factors with the largest eigenvalues are always extracted first, and each additional factor extracted will have a progressively smaller (or identical) eigenvalue than the first factor that was removed. In principal components analysis, the eigenvalues sum to the number of factored entities (e.g., variables), so if the first two eigenvalues for an EFA of 10 variables were 6 and 4, the remaining 8 eigenvalues would all be zero (and the pattern/structure coefficients on the last 8 factors would each also be zeros).

Factor Retention. After variance-accounted-for statistics and factor pattern/structure matrices have been computed, the researcher must decide the number of factors to retain in the analysis. Because different retention methods can often generate divergent results, two retention methods will be briefly discussed. However, it is almost always desirable to retain only noteworthy factors so that the solution is more parsimonious and manageable.

One method of determining the number of factors to retain is the eigenvalue-greater-than-1.0 rule (also called the K1 rule). This decision rule was initially developed by Kaiser (1958) based on the work of Guttman (1954), and is often the default option on statistical software packages (Hetzel, 1996). It is important to carefully examine all of the eigenvalues, however, as previous research has reported that in certain situations the eigenvalue-greater-than-1.0 rule can underestimate or overestimate the number of factors that should be retained (cf. Hetzel; Zwick & Velicer, 1986). For example, it would be foolish to always rigidly retain all factors with eigenvalues of 1.001 or to always rigidly ignore all factors with eigenvalues of 0.999. Further, the eigenvalue-greater-than-1.0 rule is often less appropriate to use in situations where PFA is employed due to the algebraic nature of the calculations involved (Zwick & Velicer). For a more thorough discussion of the latter issue, the interested reader is referred to Gorsuch (1983).

A second decision rule utilized to determine the number of factors to retain is the "scree" test developed by Cattell (1966). The scree test is a graphical technique in which the eigenvalues are listed along the X-axis and their magnitude is plotted on the Y-axis. The resultant chart is visually inspected to ascertain the point at which the slope

of the line connecting the eigenvalues becomes nearly zero (horizontal). All of the factors above the point at which the slope of the line becomes horizontal are retained in the analysis. This technique has been considered by some researchers as too subjective, but other researchers have contended that the scree test is one of the most accurate retention methods currently available (Zwick & Velicer, 1986).

Interpretation of Results. After the appropriate number of factors are retained in the analysis, it is necessary to interpret the results. It is invariably very difficult to interpret the initial factor/pattern structure matrix because many of the variables typically have noteworthy coefficient magnitudes on many of the retained factors (coefficients greater than |0.30|), especially on the first factor. Thompson (1984, pp. 31-34) demonstrates how the unrotated pattern/structure matrix actually misrepresents the true nature of the factors and how factor rotation resolves this misrepresentation. Interpretation of the factor analytic results is therefore almost always aided by the rotation of the factor solution, as it is possible to redistribute the common variance across the factors to achieve a more parsimonious solution. It is important to note, however, that factor rotation is not "cheating" and does not generate or discover more common variance; rather, factor rotation merely *redistributes* the variance that has been previously explained by the extracted factors.

After a factor solution is rotated, the first factor may not account for the largest portion of the variance and thus may not have the largest variance-accounted-for value. Since the variance has been redistributed throughout the factors, any of the factors could account for the largest proportion of the total variance. Additionally, after the rotation is conducted, eigenvalues are no longer termed as such; rather, after rotation, the variance-accounted-for statistic for the factors (columns of the factor pattern/structure matrix) is termed "trace." One of the most common mistakes that researchers frequently commit is believing that the eigenvalue for a given factor after extraction will equal the trace after the factor solution is rotated (Hetzel, 1996).

Two general types of factor rotation are available: orthogonal and oblique. Orthogonal rotation shifts the factors in the factor space but maintains 90 degree angles of all the factor axes to one another. This rotation strategy maintains the perfectly uncorrelated nature of the factors after the solution is rotated, and often aids in the interpretation process since uncorrelated factors may be easier to interpret. There are several orthogonal rotation strategies available, but one of the most popular orthogonal rotation techniques is the varimax method developed by Kaiser (1958). In this technique the factors are "cleaned up" so that ideally every observed variable has a noteworthy factor pattern/structure coefficient on only one of the factors. Varimax rotation produces factors that have noteworthy pattern/structure coefficients for a small number of variables and near-zero pattern/structure coefficients for the other variables.

There are several advantages to employing orthogonal rotation strategies. First, the factors remain perfectly uncorrelated with one another and therefore may be easier to interpret. Secondly, the factor pattern matrix and the factor structure matrix remain equivalent, thus only one matrix is estimated and interpreted. This means that the solution is more parsimonious (i.e., fewer parameters are estimated) and, in theory, more replicable.

Orthogonal rotation strategies do, however, have limitations. Orthogonal rotations often do not honor a given researcher's view of reality if the researcher believes that two or more of the extracted and retained factors are correlated. Secondly, orthogonal rotation of factor solutions may oversimplify the relationships between the variables and the factors and may not always accurately represent these relationships. Consequently, some researchers have challenged the utility of orthogonal rotation strategies. Thurstone (1947, p. 139) contended that the use of orthogonal rotation indicates "our ignorance of the nature of the underlying structure. . . . The reason for using uncorrelated [factors] can be understood, but it cannot be justified." Similarly, Cattell (1978, p. 128) argued (regarding researchers performing orthogonal rotation), ". . . in half of [the] cases it [orthogonal rotation] is done in ignorance of the issue rather than by deliberate intent." Consequently, even though orthogonal rotation eases the interpretability of the factor solution, it may not accurately portray the relationships between the variables and the emergent factors.

The second type of factor rotation is termed oblique rotation. This method of rotation provides for correlations among the latent constructs and is termed oblique because the angle between the factors becomes greater or less than the 90 degree angle that is utilized to perform an orthogonal rotation. One of the most popular oblique rotation strategies is promax (see Gorsuch, 1983 for a thorough explanation of oblique rotations).

In oblique rotation, the researcher is attempting to achieve the most parsimonious simple structure given that the factors are allowed to be correlated with one another. However, an oblique factor solution inherently tends to be less parsimonious. For example, if 5 factors for 100 factored entities (e.g., variables) are extracted and orthogonally rotated, only 500 factor pattern/structure coefficients are estimated (the 5 x 5 factor correlation matrix is not estimated, because it is constrained to have 1's on the diagonal and 0's everywhere else). If the same EFA factors are rotated obliquely, *1,010* coefficients (500 factor pattern coefficients, plus 500 factor structure coefficients, plus 10 factor correlation coefficients (the 10

non-redundant off-diagonal entries in the 5 x 5 factor correlation matrix)) are estimated. [It might be argued, however, that only *510* coefficients are estimated in this case, since with either the 10 unique factor correlation coefficients, and either the 500 pattern or the 500 structure coefficients, the remaining 500 pattern or structure coefficients are fully determined.]

Oblique rotation strategies can be useful to researchers for a variety of reasons noted previously. Oblique rotations may be difficult to interpret, especially if there is a high degree of correlation among the factors. Further, because the factor pattern and factor structure matrices are not equal, each has to be interpreted in conjunction with the other. However, a simpler structure for the pattern matrix resulting from an oblique rotation will always be the benefit of paying the price for this interpretation conundrum. But, as the degree of correlation between the factors decreases, both orthogonal and oblique solutions provide increasingly similar results. Given that oblique solutions are less parsimonious and therefore theoretically less replicable, an oblique rotation would therefore only be employed when the benefits of simpler, more interpretable structure outweigh the costs of less replicability (i.e., when the orthogonal factors are not readily interpretable, and the oblique factors are fairly highly correlated but more interpretable). Consequently, it is often important to consult both orthogonal and oblique rotation strategies to determine which rotation best fits the structure of the data (Gorsuch, 1983).

The *SAS/STAT User's Guide* (1990, p. 776) has indicated that, "You cannot say that any rotation is better than any other rotation from a statistical point of view; all rotations are equally good statistically. Therefore, the choice among different rotations must be based on non-statistical grounds. . . ." Thus, if an orthogonal and oblique solution both fit a given data set equally well, it is often preferable to choose the orthogonal rotation as the results will tend to be more replicable and the solution is often easier to interpret (due to the uncorrelated nature of the latent traits).

Heuristic Example Using EFA

Since the basic conceptual tenets of EFA have been described in the preceding sections of the present article, a heuristic example can be utilized to illustrate in concrete terms the manner in which an analysis would be conducted. For the purposes of the heuristic example in the present paper, only tests T1, T3, T4, T7, T8, T10, T13, T15, T16, T17, T19, T22, T23 from the original Holzinger and Swineford (1939) data set will be utilized. The means, standard deviations and variable labels for these raw data are presented in Table 1.

Table 1
Means and Standard Deviation of 13 Variables
From Holzinger and Swineford, (1939)

Variable	Mean	SD	Label
T1	29.615	7.005	Visual Perception Test
T3	14.229	2.830	Paper Form Board (Spatial)
T4	18.003	9.048	Thorndike Lozenges (Spatial)
T7	17.362	5.162	Sentence Completion Test
T8	26.126	5.675	Word Classification
T10	96.276	25.059	Speeded Addition Test
T13	193.468	36.329	Speeded Discrimination
T15	90.010	7.729	Memory of Target Numbers
T16	102.525	7.633	Memory of Target Shapes
T17	8.233	4.916	Memory of Number-Object Assn.
T19	14.037	4.077	Memory of Figure-Word Assn.
T22	26.239	9.197	Math Word Problem Reasoning
T23	18.136	9.140	Completion of Number Series

As described previously, the first step in completing a factor analysis is determining which matrix of association will be utilized in the analysis (e.g., the correlation matrix or the variance-covariance matrix). Using the correlation as opposed to the variance/covariance matrix can affect the results of the EFA in certain situations. The difference between the two matrices involves the metric in which the variables were measured (Marcoulides & Schumaker, 1996). Tabachnick and Fidell (1983) noted that,

> Many programs allow the researcher a choice between analysis of a correlation matrix and analysis of a variance-covariance matrix. If the correlation matrix is analyzed, a unit-free result is produced. That is, the solution reflects the relationships among the variables but not in the metric in which they were originally measured. If the metric of the scores was somewhat arbitrary to begin with, analysis of [the correlation matrix] is appropriate. (p. 19)

Conversely, the variance-covariance matrix contains information about the scale in which the variables were measured and is not metric-free as in the case of the correlation matrix. Thus, in the present example of EFA, the correlation matrix of the 13 variables was analyzed because the metric of the original scales was largely arbitrary in nature. This correlation matrix is presented in Table 2.

After the matrix of association is chosen, it is possible to determine the extraction method. Since principal components analysis (PCA) yields results similar to principal factors analysis as the number of factored entities increases and since PCA requires no further manipulation of the data array, PCA was chosen to extract the latent constructs.

Table 2
Correlation Matrix for Example Data

	T1	T3	T4	T7	T8	T10	T13	T15
T1	1.000							
T3	.365	1.000						
T4	.441	.305	1.000					
T7	.293	.173	.077	1.000				
T8	.331	.212	.171	.674	1.000			
T10	.067	.040	.072	.102	.134	1.000		
T13	.332	.227	.329	.227	.222	.341	1.000	
T15	.184	.036	.212	-.019	.052	.109	.072	1.000
T16	.365	.184	.305	.166	.292	.117	.278	.338
T17	.108	.026	.147	.092	.107	.331	.198	.305
T19	.185	.091	.159	.216	.297	.070	.196	.116
T22	.399	.188	.306	.470	.403	.069	.245	.075
T23	.481	.276	.397	.385	.424	.190	.332	.186

	T16	T17	T19	T22	T23
T16	1.000				
T17	.259	1.000			
T19	.277	.216	1.000		
T22	.278	.145	.322	1.000	
T23	.370	.202	.283	.535	1.000

The next step in completing the EFA is to determine how many factors to retain. The eigenvalue-greater-than-one rule was utilized to determine an appropriate number of factors to retain and resulted in the retention of four factors. To conserve space, only the retained factors are presented in Table 3.

Table 3
Unrotated Factor Pattern and Factor Structure Matrix

Variable	I	II	III	IV	h^2
T1	.691	-.028	-.381	.052	.625
T3	.438	-.102	-.482	.248	.496
T4	.562	.211	-.490	.037	.602
T7	.586	-.537	.351	.014	.755
T8	.648	-.430	.295	-.062	.696
T10	.297	.391	.452	.584	.786
T13	.568	.175	-.028	.514	.619
T15	.307	.588	-.018	-.440	.634
T16	.594	.294	-.061	-.299	.532
T17	.368	.555	.412	-.029	.614
T19	.474	-.002	.258	-.333	.403
T22	.676	-.279	.055	-.148	.560
T23	.757	-.063	-.047	-.025	.580
			(Sum =)		
Eigenvalues	3.997	1.534	1.278	1.092	7.902
% of Variance	30.7	11.8	9.8	8.4	60.8

In practice, it is often helpful to utilize several methods in determining the final set of factors to retain since each method tends to produce different results in some instances (cf. Zwick & Velicer, 1986). Notice that the first unrotated factor had an eigenvalue of 3.997 and accounted for 30.7% of the total variance (total variance is calculated as ((3.997 / 13) * 100). The eigenvalues of the subsequent factors decrease as the factor pattern/structure matrix is scanned from left to right. Correspondingly, the percentage of variance-accounted-for by each unrotated factor decreases in magnitude as the matrix is scanned from left to right.

The factor correlation matrix is presented in Table 4. The table indicates that each of the factors correlates perfectly with itself but does not correlate with any of the other factors (i.e., the factors are perfectly uncorrelated). This is the universal result of EFA factor extraction.

Table 4
Factor Correlation Matrix

	I	II	III	IV
Factor I	1.000			
Factor II	.000	1.000		
Factor III	.000	.000	1.000	
Factor IV	.000	.000	.000	1.000

It is important to note that two matrices are contrived as a result of the PCA: one is a factor pattern matrix and the other is a factor structure matrix (both matrices are identical to the matrix presented in Table 3). Because PCA was utilized to extract the factors, the extracted factors are perfectly uncorrelated and, consequently, the factor pattern and structure matrices are exactly equal. Thus, the factor pattern matrix and the factor structure matrix can be combined into one factor pattern/structure matrix because all of the values are identical and no information will be lost.

An examination of Table 3 will reveal that the factor saturation (i.e., which observed variables have noteworthy non-zero coefficients on which latent constructs) is so complex that it is difficult to interpret the factor pattern/structure matrix in its present form. That is, as reported in Table 3, 12 of the 13 variables have pattern/structure coefficients greater than |.30| for the first latent construct, and the thirteenth coefficient is nearly |.30| (i.e., .297 for variable T10). Thus, to more easily interpret the results, the four-factor solution was first rotated to the varimax criterion. The results of the varimax-rotated solution are presented in Table 5.

An examination of the results presented for the four-factor solution rotated to the varimax criterion reveals that ascertaining which variables are associated with which factors has been greatly facilitated by the rotation procedure. Factor I is most highly saturated with tests T7, T8, T19, T22, and T23. By examining the Table 1 information, it is possible to see that the five tests which were associated with Factor I seem to have a strong verbal component. Thus, it might be possible to name this factor the *verbal comprehension* factor, as all of the tests associated with this factor appear to measure verbal ability.

Table 5
Factor Pattern/Structure Matrix Rotated to the Varimax Criterion

Variable	I	II	III	IV	h^2
T1	.297	.713	.168	.029	.625
T3	.097	.688	-.110	.019	.496
T4	.027	.723	.280	.035	.602
T7	.851	.070	-.116	.108	.755
T8	.818	.133	.021	.090	.696
T10	.050	.003	.056	.884	.786
T13	.162	.502	.008	.584	.619
T15	-.077	.097	.786	.017	.634
T16	.242	.330	.602	.048	.532
T17	.093	-.083	.574	.519	.614
T19	.483	-.002	.411	.017	.403
T22	.660	.313	.162	-.014	.559
T23	.518	.478	.252	.139	.579
			(Sum =)		
Trace	2.610	2.230	1.704	1.410	7.902
% of Variance	19.7	17.2	13.1	10.8	60.8

Note. Coefficients greater than |.30| are underlined.

Factor II is most highly saturated with tests T1, T3, and T4, each of which seem to be assessing spatial ability. Consequently, this factor might be termed the *spatial ability* factor, as all the tests associated with it seem to be measuring perceptual ability. The third factor is most highly saturated with tests T15, T16, and T17. All of these tests appear to be measuring a *memory* component. Finally, Factor IV is most highly saturated with tests T10 and T13, which seem to be strongly associated with the *speeded execution* of tasks.

It is important to notice that the communality coefficients for the varimax rotated solution are identical to the communality coefficients in the unrotated four factor solution. The reason for this is that the variable variance reproduced by a given factor solution is redistributed in the rotated solution but no new variance is ever generated through a rotation procedure nor is any variance lost.

As stated previously, after the solution is rotated, the first extracted factor often does not account for the preponderance of the total variance. By examining the trace in Table 5, it is possible to view this dynamic in the present data set. Factor I accounted for 30.7% of the total variance in the unrotated factor solution and is still the strongest overall factor, but after rotation this factor accounts for only 19.7% of the total variance. However, the total variance-accounted-for the by the four factor solution before rotation (60.8%) is exactly equal to the total variance-accounted-for after rotation. As previously stated, no new variance is generated when the factors are rotated; rather the variance is merely distributed differently among the factors.

For the sake of comparing the parsimony of the results generated by orthogonal and oblique rotations, the unrotated factor pattern/structure matrix was also rotated obliquely to the direct oblimin criterion with delta equal to zero (see Gorsuch (1983) for a more detailed explanation of the effects of varying the value of delta). The results of the direct oblimin rotation are presented in Tables 6 and 7.

Table 6
Factor Pattern Matrix Rotated to the Direct Oblimin Criterion

Variable	I	II	III	IV	h^2
T1	.199	.105	-.683	-.015	.625
T3	.010	-.158	-.709	-.001	.496
T4	-.094	.244	-.727	.001	.602
T7	.887	-.212	.056	.061	.755
T8	.837	-.071	-.005	.037	.696
T10	-.011	-.025	.055	.896	.786
T13	.060	-.084	-.466	.571	.619
T15	-.144	.805	-.057	-.016	.634
T16	.169	.572	-.261	-.002	.532
T17	.040	.536	.168	.501	.614
T19	.483	.373	.107	-.031	.403
T22	.642	.089	-.213	-.071	.560
T23	.453	.171	-.393	.086	.580

Note. Communality coefficients are now computed differently. To compute the h^2 for variable T16, each pattern coefficient is multiplied by its corresponding structure coefficient and then summed across the rows. For T1, $h^2 = (.199)(.422) + (.105)(.245) + (-.683)(-.757) + (-.015)(.117) = .625$. The new communality coefficient is identical to the value obtained in the Table 3 analysis.

Table 7
Factor Structure Matrix Rotated to Direct Oblimin Criterion

Variable	I	II	III	IV	h^2
T1	.422	.245	-.757	.117	.625
T3	.193	-.046	-.687	.063	.495
T4	.171	.339	-.737	.112	.602
T7	.840	-.043	-.185	.167	.755
T8	.831	.095	-.250	.164	.696
T10	.115	.104	-.048	.884	.786
T13	.278	.089	-.541	.625	.619
T15	.023	.784	-.137	.092	.634
T16	.355	.644	-.401	.146	.532
T17	.173	.596	.011	.571	.614
T19	.517	.443	-.093	.093	.403
T22	.711	.233	-.411	.074	.560
T23	.618	.332	-.566	.235	.580
			(Sum =)		
Trace	2.577	1.657	2.215	1.459	7.902
% of Variance	19.8	12.7	17.0	11.2	60.8

Note. Trace are computed differently than eigenvalues. To compute the trace for Factor I, each pattern coefficient is multiplied by its corresponding structure coefficient and then summed down the rows.
For Factor I, Trace = (.199)(.422)+(.010)(.193)+(-.094)(.171)+(.887)(.840)+(.837)(.831)+(-.011)(.115)+(.060)(.278)+(-.144)(.023)+(.169)(.355)+(.040)(.173)+(.483)(.517)+(.642)(.711)+(.453)(.618) = 2.577. The trace still sum to 7.902, the sum of communality coefficients in the Table 3 analysis.

When interpreting the results of an oblique rotation, it is necessary to interpret two separate factor association matrices, because the factor pattern matrix is no longer identical to the factor structure matrix. However, similarly to multiple linear regression, it is *critically* important to interpret *both* pattern coefficients (standardized weights) and structure coefficients as each can provide only one piece of information regarding the overall structure (Thompson, 1997b).

By examining the variance-accounted-for by each factor (trace), the results of the direct oblimin rotation appear to closely resemble the results generated by the varimax rotation. The communality coefficients in the oblique case are exactly equal to the results attained in the unrotated and varimax rotated factor solutions, thus again illustrating only the redistribution of common variance. The sum of the communality coefficients is still equal to 60.8%, as in all of the prior analyses. The trace and communality coefficients are computed slightly differently, however, as it is necessary to multiply a given factor pattern coefficient by the corresponding factor structure coefficient and then to sum down the columns or across the rows to derive the various variance-accounted-for estimate (see the note on Table 6 and Table 7 for a more detailed explanation).

After examining both the factor pattern and structure coefficients, the saturation of the factors can be determined. Factor I is most highly saturated with tests T7, T8, T19, T22, and T23 which is identical to the results generated by the orthogonal rotation. Factor II is most highly saturated with T15, T16, and T17 (similarly to Factor III in the varimax rotation), Factor III is most highly saturated with tests T1, T3, and T4 (similarly to Factor II in the varimax rotation) and Factor IV is most highly saturated with tests T10 and T13 (identical to Factor IV in the varimax rotation). Thus, based on the present example, both the orthogonal and oblique rotations provide generally the same conclusions in terms of factor-variable saturation. In the oblique case, however, it is necessary to consult two different matrices of association whereas in the orthogonal rotation is it necessary to examine only the single factor pattern/structure matrix.

The primary difference between these two rotation strategies is the correlation between the factors. The factor correlation matrix for the oblique rotation is presented in Table 8. As is evidenced by the table, all of the factors are correlated with one another. However, most of the correlations in the present example are relatively small (with the possible exception of Factor I with Factor III, where $r^2 = -.301^2 = 9.0\%$), which is why the pattern and the structure matrices are so similar. Because both solutions are similar and the orthogonal rotation is interpretable and more parsimonious, most researchers would select this as the preferred solution.

Table 8
Factor Correlation Matrix after Rotation to the Direct Oblimin Criterion

	I	II	III	IV
Factor I	1.000			
Factor II	.190	1.000		
Factor III	-.301	-.156	1.000	
Factor IV	.164	.155	-.122	1.000

Summary of Exploratory Factor Analysis

The present heuristic example of EFA has demonstrated the usefulness of EFA as a tool in theory development and construct validation. EFA, however, does not invoke an inherent theoretical rationale for the determination of the number of factors to retain or for which rotation strategy to employ. Consequently, researchers must use subjective judgment in determining the extraction methods and rotation strategies that will be utilized in completing the analysis. EFA, therefore, is very useful in examining the structure of data for which there is either a paucity of research or for which no research has previously been conducted. However, because exploratory methods do not directly address the issue of theory validation, their application is this realm is extremely limited. Thus, the varied usefulness of EFA has predisposed some to characterize exploratory methods as neither "a royal road to truth as some apparently feel, nor necessarily an adjunct to shotgun empiricism, as others claim" (Nunnally, 1978, p. 371).

Confirmatory Factor Analysis

As mentioned previously, the conceptual base upon which both contemporaneous exploratory and confirmatory factor analysis rests was initially developed as a method of providing statistical confirmation for a psychological construct (Spearman, 1904). Thus, factor analysis has historically been associated with issues of construct validation and "is at the heart of the measurement of psychological constructs" (Nunnally, 1978, p. 112).

However, it was not until the development of the "true" confirmatory factor analytic techniques that directly permit the testing of hypotheses and theories in the latter half of the 20th century that the true power of factor analytic techniques has been realized. As stated by Thompson (1992, p. 83), "Without question, confirmatory factor analysis is more important to evaluating theory than is exploratory analysis." Similarly, Gorsuch (1983, p. 134) has stated, "confirmatory factor analysis is the more theoretically important--and should be the much more widely used--of the two major factor analytic approaches."

Confirmatory techniques have long been needed in social science research to provide support for the theories that have been generated through exploratory analyses. The seminal work of Jöreskog (1966, 1969) forever

redefined the landscape of factor analysis. Jöreskog developed a group of maximum likelihood estimation techniques that have since been loosely termed confirmatory factor analysis (CFA). As stated by Thompson (1992, p. 83), the CFA methods constitute,

> a truly confirmatory class of maximum likelihood methods; that is, the methods assume a sample and focus use of the sample data on best estimating population parameters. The methods are truly confirmatory because the methods allow the testing of any of the many possible factor analytic hypotheses as a model representing a complete *omnibus system* of hypotheses [emphasis in original].

Thus, the development of true confirmatory factor analytic methods has allowed researchers to *directly* test the fit between theories and data structure with a sample of data in hand rather than allowing a factor structure to emerge from a set of data without regard to theoretical expectations as in exploratory analysis.

Confirmatory Rotation After Exploratory Factor Extraction

True confirmatory factor analysis directly extracts factors based on a priori expectations. There is a hybrid method of factor analysis in which EFA is first utilized to extract factors and then a Procrustean rotation method is employed to "best fit" the extracted factors to a theoretically-expected structure. Thompson (1992) has developed some test statistic distributions to evaluate the fit of the actual to the expected factors. Given space limitations, this hybrid semi-confirmatory method is only mentioned here. The interested reader is referred to Thompson (1992) or Daniel (1989).

Confirmatory Factor Extraction

Confirmatory extraction is a "pure" form of CFA and does not require the rotation of factor results, because the estimates generated through the analysis are final estimates. The purpose of confirmatory extraction is to determine directly the fit of the observed variables with the theoretical assumptions of the researcher. In CFA, the factors are typically expected to have "simple structure," and it is this expectation rather than rotation that results in factor interpretability. Consequently, CFA results are not rotated.

It is necessary in CFA to utilize a statistical software package such as LISREL 7 (Jöreskog & Sörbom, 1989). Several competing models are input into the computer program, rather than only the single expected model, as more than one model can adequately represent a given data structure. When inputting the models, the researcher must decide which parameters of the model are "fixed" and which are "free." Free parameters refer to those specific parameters that will be estimated by the analysis. Conversely, a fixed parameter is a parameter that the researcher does not wish to estimate during the analysis. Fixed parameters are denoted as zeros in the target matrix selected. Factor correlation coefficients, factor coefficients (pattern/structure coefficients) and the measurement error variances are typically freed during a given analysis based on the a priori theoretical assumptions of the researcher.

The purpose in conducting the analysis is to determine which model or models best fit the data. Consequently, the analysis generates several statistics which help the researcher determine if the fit of the model with the data is supported. These "fit" statistics include the Bentler Comparative Fit Index (CFI; Bentler, 1990), the chi square--degrees of freedom ratio, the Goodness of Fit Index and the Adjusted Goodness of Fit Index (GFI, AGFI; Jöreskog & Sörbom, 1989), the Root Mean Square Residual (RMSR) and the Root Mean Square Error of Approximation. More complete explanations of these fit statistics are presented elsewhere (cf. Daniel, 1989; Gillaspy, 1997; Thompson, 1998), but each of the fit statistics is not an individual indicator of the model fit. Only when the fit statistics are used in conjunction with one another can the model-to-data fit be accurately assessed (Campbell, Gillaspy & Thompson, 1995).

If the model, given the estimates for the "freed" parameters, successfully reproduces the relationships among the measured variables, then the model is said to demonstrate a fit with the data. This does not mean that the results have been confirmed, however, as more than one model can fit the data (even models that were not hypothesized in the original analysis). Fit is assessed through the consultation of a variety of the aforementioned fit statistics. If no models sufficiently fit the data, it is possible to use the analysis to generate better fitting models. This is called a model specification search. For instance, LISREL provides diagnostic information as to the fit of the model if certain fixed (non-estimated) parameters would have been freed (termed modification indices which are estimates of improvements in χ^2 if these changes are made) or if fixed parameters would have been estimated/freed (when the ratio of a given parameter to its standard error is less than 2.0, a researcher can consider "fixing" parameters that were originally estimated). Although it is not always appropriate to change the model based on these statistics, in certain situations it might be appropriate to alter some of the fixed and freed parameters to generate a better model-to-data fit (see Byrne, 1989; 1994; Thompson, 1998).

Heuristic Example Using CFA

The same heuristic data set utilized in the example application of EFA (Holzinger & Swineford, 1939) was again employed to demonstrate the completion of a confirmatory extraction procedure. As noted previously, only tests T1, T3, T4, T7, T8, T10, T13, T15, T16, T17, T19, T22, and T23 of the 24 original tests were utilized in the analysis. However, note that this is a heuristic example only. In real research situations it is unacceptable to conduct a CFA on the same data first analyzed by EFA to devise the tested models. Further, it is important to note that sample size considerations are important in CFA as well, and authors have generally recommended that CFA analyses employ 500 to 1,000 participants in an effort to achieve result stability and factor reliability (Gorsuch, 1983; Stevens, 1996; Tabachnick & Fidell, 1983).

Model Specification

The first step in completing a CFA is typically to specify several competing models. Because several models can fit a given data set, the finding that a particular model best fits the data, given testing of several models, provides stronger support for the fitting model. Both of the tested models in the present example have fixed factor variance to 1.0 and have freed selected factor pattern and factor correlation coefficients. The covariance matrix among the 13 variables was utilized to complete the two analyses. In addition to the two posited models, each of the models was compared to a null model indicating that no factors would be identified by the analysis. An important point to note about model specification is that CFA does not ever *prove* that a model is true; rather, CFA only indicates which of the tested models best fits the data. It is possible that models not included in the analysis could fit the data equally as well or better than the models evaluated by the CFA.

Model #1. This model indicated that the original five factors delineated by the Holzinger and Swineford study would be identified in the analysis. The expectation was that each variable would have large factor pattern coefficients on only the one factor with which the variable was expected to be associated: (a) *Verbal* Factor: T7 and T8; (b) *Visual* Factor: T1, T3, and T4; (c) *Speed* Factor: T10 and T13; (d) *Memory* Factor T15, T16, T17, and T19; and (e) *Math* Factor: T22 and T23. Holzinger and Swineford (1939) indicated that the five expected factors were uncorrelated, but for the purpose of the present analysis, the factor correlation coefficients were estimated. Furthermore, in all of these analyses involving factors presumed to be measured by only two variables, the pair of pattern coefficients were constrained to be equal (see Marcoulides & Schumaker, 1996 for a more detailed discussion of this issue).

Model #2. This model indicated that the four factors retained in the EFA completed in the present article would be identified by the analysis. The expectation was that each variable would have large factor pattern coefficients on only the one factor with which the variable was expected to be associated: (a) *Verbal* Factor: T7, T8, T19, T22, and T23; (b) *Visual* Factor: T1, T3, and T4; (c) *Speed* Factor: T10 and T13; (d) *Memory* Factor T15, T16, and T17. Based on the results from the oblique rotation in the present EFA, it was expected that all of the factors would be correlated.

The next step in completing the CFA is to run the analysis and interpret the generated fit statistics. The current analysis was completed using LISREL 7. Selected fit statistics from the CFA analysis are presented in Table 9. It is important to note that the presented fit statistics are only a few of the several dozen fit statistics currently available (Fan, Thompson, & Wang, 1996). Fan, Thompson, and Wang indicated that because many of the fit statistics have been developed with different rationales, it is important to first appreciate the rationale and the differential interpretation of the statistics as well as to understand that these statistics may vary across research situations.

There are three general types of fit statistics currently available: covariance matrix reproduction indices, comparative model fit indices, and the parsimony weighted indices. The three types of fit statistics are explained more fully by Fan, Thompson, and Wang (1996). Since there is not currently one accepted fit statistic that accurately determines the model-to-data fit, it is necessary to examine several of these different types of indexes to determine if a given model fits the data (Thompson & Daniel, 1996). When conducting a CFA analysis, a given researcher hopes that a variety of fit statistics jointly corroborate the fit of a model.

The maximum likelihood estimated factor pattern and factor correlation matrix for model #1 are presented in Tables 10 and 11, respectively. The maximum likelihood estimates for Model #2 are presented in Tables 12 and 13, respectively. Both of the factor correlation matrices indicated a moderate to large degree of correlation between the factors for each model (ranging from .328 to .786). The chi-square statistic for Model #1 (χ^2 = 165.164, df = 58) was smaller in relation to the number of degrees of freedom utilized than in Model #2 (the ratio of chi-square to df should be 2 to 1 or less to indicate a good model-to-data fit), indicating that Model #1 better fit the data on this statistic. The GFI index indicated a better fit for Model #1 (GFI = 0.960) than for Model #2 (GFI = 0.899) as well. After sample size was taken into account, the AGFI still indicated that Model #1 better fit the data than Model #2.

Table 9
Selected Fit Statistics for Example CFA Analysis

	Model #1	Model #2
v	13	13
n	301	301
Null chi sq	1009.302	1009.302
Null df	78	78
Noncentrality	931.302	931.302
Model chi sq	165.164	240.421
Model df	58	60
Noncentrality	107.161	180.422[a]
NC / df	1.848	3.007[b]
GFI	0.960	0.899
AGFI	0.877	0.829
Parsimony Ratio	0.637	0.659[c]
GFI*Pars Ratio	0.612	0.593[d]
CFI	0.885	0.806[e]
Parsimony Ratio	0.744	0.769[f]
CFI*Pars Ratio	0.658	0.620[g]
RMSR	20.717	20.741
RMSEA	0.006	0.010[h]

[a]Noncentrality = χ^2 - df

[b]Noncentrality / df

[c]Parsimony Ratio = Model df / [(variables * (variables + 1)) / 2]

[d]GFI * Parsimony Ratio

[e]$$CFI = \frac{[(Null\ \chi^2 - Null\ df) - (Model\ \chi^2 - Model\ df)]}{(Null\ \chi^2 - Null\ df)}$$

[f]Parsimony Ratio = Model df / [variables * (variables - 1)) / 2]

[g]CFI * Parsimony Ratio

[h]RMSEA = [Model = χ^2 - Model df) / (Model df * (n-1))]$^{.5}$

Table 10
Maximum Likelihood Factor Parameter Estimates for Model #1

	Factor				
Variable	Verbal	Visual	Speed	Memory	Math
T1	0.000	0.757	0.000	0.000	0.000
T3	0.000	0.458	0.000	0.000	0.000
T4	0.000	0.608	0.000	0.000	0.000
T7	0.850	0.000	0.000	0.000	0.000
T8	0.790	0.000	0.000	0.000	0.000
T10	0.000	0.000	0.635	0.000	0.000
T13	0.000	0.000	0.480	0.000	0.000
T15	0.000	0.000	0.000	0.438	0.000
T16	0.000	0.000	0.000	0.677	0.000
T17	0.000	0.000	0.000	0.455	0.000
T19	0.000	0.000	0.000	0.434	0.000
T22	0.000	0.000	0.000	0.000	0.717
T23	0.000	0.000	0.000	0.000	0.749

Table 11
Phi (Factor Correlation) Matrix for Model #1

	Factor				
Factor	Verbal	Visual	Speed	Memory	Math
Verbal	1.000				
Visual	0.425	1.000			
Speed	0.328	0.450	1.000		
Memory	0.360	0.606	0.544	1.000	
Math	0.693	0.786	0.461	0.644	1.000

Table 12
Maximum Likelihood Factor Parameter Estimates for Model #2

	Factor			
Variable	Verbal	Visual	Speed	Math
T1	0.000	0.780	0.000	0.000
T3	0.000	0.465	0.000	0.000
T4	0.000	0.581	0.000	0.000
T7	0.679	0.000	0.000	0.000
T8	0.704	0.000	0.000	0.000
T10	0.000	0.000	0.637	0.000
T13	0.000	0.000	0.480	0.000
T15	0.000	0.000	0.000	0.463
T16	0.000	0.000	0.000	0.713
T17	0.000	0.000	0.000	0.436
T19	0.411	0.000	0.000	0.000
T22	0.686	0.000	0.000	0.000
T23	0.701	0.000	0.000	0.000

Table 13
Phi (Factor Correlation) Matrix for Model #2

	Factor			
Factor	Verbal	Visual	Speed	Math
Verbal	1.000			
Visual	0.667	1.000		
Speed	0.439	0.441	1.000	
Math	0.512	0.592	0.519	1.000

In terms of the parsimony of the two models, the parsimony ratios indicated that Model #2 was the less complex of two models. When the parsimony ratio was utilized to weight the GFI and the Comparative Fit Index (CFI), a better model-to-data fit was indicated by Model #1. Similarly, the Root Mean Square Residual (RMSR) and the Root Mean Square Error of Approximation (RMSEA) both indicated that Model #1 was a better fit to the data than Model #2. A good fit on the RMSR and RMSEA statistic is indicated as these two statistics approach zero. Here the RMSR is artifactually very large because the variance/covariance matrix involving scores on the battery of tests (not individual test items, which would tend to have smaller variances and covariances) was analyzed. The RMSEA statistic, however, indicated a good fit for both models.

After examining the fit statistics, it is important to examine the modification indices and the parameter-to-standard-error ratios to determine if a better model-to-data fit would be generated if some of the parameters were freed or fixed, respectively. It is important to note that utilizing the modification indexes and parameter-to-standard-error ratios to alter the expected model constitutes using the CFA in a non-confirmatory manner if the revised model is then fit to the same data. Based on the information generated by the modification indexes, a noteworthy reduction in the chi-square would be achieved in Model #1 if the T10 and T13 parameters were freed on the Visual factor (a decrease of 39.939 for both variables). Similarly for Model #2, a noteworthy reduction in the chi-square would be evidenced if the T10, T13, and T23 parameters were freed on the Visual factor (estimated reduction of 40.758 for T10 and T13 and a reduction of 30.331 for T23). Freeing these parameters, however, would decrease the parsimony of the two models and therefore affect the parsimony-weighted fit statistics.

An important consideration in completing a CFA is the effect that utilizing the correlation matrix versus the variance-covariance matrix will have on the analysis. As stated previously, the correlation matrix is a symmetric matrix with 1's on the main diagonal and correlation coefficients on the off-diagonals. The variance-covariance matrix has the variance of each variable on the main diagonal and covariances on the off-diagonals. Using the correlation matrix has the effect of standardizing the solution, as here both the factors and the variables would be constrained to unit variance (1.0). It can be argued, therefore, that using the correlation matrix can be likened to utilizing a variance-covariance matrix where the measured variables have been standardized to a variance of 1.0. If the fit statistics generated by utilizing the correlation matrix are equivalent to those generated by using the covariance matrix, then the correlation matrix solution can be interpreted. If the results are divergent, however, the variance-covariance matrix should be used in calculating the fit statistics, although a standardized or completely standardized solution would still be used to interpret the results (see Cudeck, 1989 or Byrne, 1989, 1994 for more detail).

Based on the results presented in the present article, Model #1 best fit the data presented in Holzinger and Swineford (1939). Model #1 posited that each variable was associated with only one of five expected factors and that the factors were correlated.

Summary of Confirmatory Factor Analysis

The present heuristic example of CFA has demonstrated the usefulness of CFA as a tool in providing support for the theoretical expectations of data structures. CFA allows researchers to directly test the theoretical expectations generated either by previous research or through intrinsic value assessment. As stated by Daniel (1989, p. 24), "Used alone or in tandem with exploratory methods, confirmatory methods can help the researcher avoid erroneous conclusions about factor structures which might emerge using exploratory methods alone."

Brief Comments on Analytic Traditions in EFA and CFA

Analytic traditions in EFA and CFA have differed to some degree. For instance, most EFA analyses have historically utilized orthogonal rotation strategies that allow the latent constructs to remain perfectly uncorrelated after the factor solution is rotated for the purpose of interpretation. One reason for the emergence of this tradition is the degree of difficulty in interpreting the results of the solution. When factors are removed from the underlying data structure, it is generally easier to interpret the results if the factors are perfectly uncorrelated.

Conversely, most researchers estimate the correlation among the factors when performing a confirmatory analysis. One reason the factors are allowed to be correlated in CFA is to evaluate the independence of the factors and to ensure that two latent constructs are not highly correlated and thus evaluating the same construct.

Several authors have noted that many published articles do not accurately and completely report the results of factor analyses. Tinsley and Tinsley (1987) indicated that most studies which employed factor analytic techniques in counseling psychology did not provide adequate information for independent post hoc analysis of the study results. Consequently, these authors found that interested and even skeptical researchers were simply forced to rely on the interpretation posited by the authors of the study rather than having the necessary statistics to independently evaluate the analysis themselves. In a recent review of select issues of the *Journal of Counseling Psychology* (*JCP*), Hetzel (1996) noted that the remarks posited by Tinsley and Tinsley nearly 10 years before had seemingly passed unnoticed by researchers publishing in the journal. In conducting his own analysis, Hetzel found that despite the ardent efforts of other authors to promote good reporting practice, none of the 13 articles reviewed reported all of the necessary components.

Because analytic traditions have tended to differ in exploratory and confirmatory factor analysis, the present author randomly reviewed the factor analytic practices in one psychology journal. *JCP* is a flagship journal of the American Psychological Association (the primary association of psychologists); consequently, this resource was deemed the best and most accurate indicator of the practices throughout the field of counseling psychology. Further, this is the journal used by Hetzel (1996) in his

review, thus enabling comparison across studies. Ten journal articles (five articles each that reported to be either exploratory or confirmatory factor analyses) were randomly selected from all of the articles utilizing CFA or EFA techniques over the 10 years, 1988-1997. The chosen articles are presented in Table 14 and the results of the analysis are presented in Table 15.

Table 14
Selected EFA and CFA Studies from the *Journal of Counseling Psychology*

Author	Year	Type of Analysis
Cole & Jordan	1989	CFA
Good et al.	1995	CFA
Hayes & Tinsley	1989	EFA
Larson et al.	1992	EFA
Lee & Robbins	1995	CFA
Rice, Cole & Lapsley	1990	EFA
Sodowsky, Taffe, Gutkin & Wise	1994	EFA
Tinsley, Bowman & York	1989	EFA
Tinsley, Roth & Lease	1989	CFA
Tracey, Glidden & Kokotovic	1988	CFA

Table 15
Analysis of Select Articles Based on Recommendations by Hetzel (1996) and Thompson and Daniel (1996)

	% of Studies Reporting	
Information	CFA	EFA
Coefficients Termed "Loadings"	100%	100%
Criteria for Factor Retention	—	100%
Eigenvalues	—	100%
Factors Correlated	60%	80%
Final Communality Coefficients	—	60%
Initial Communality Estimates Used	—	20%
Interpret Several Fit Statistics	20%	—
Item Means and Variances	—	20%
Matrix of Association		
Correlation Matrix	40%	20%
Variance-Covariance Matrix	20%	0%
Method of Factor Extraction	—	80%
Method of Factor Rotation	—	100%
Method of Sample Selection	80%	40%
Multiple Models Specified	80%	—
Number of Variables	80%	80%
Reliability Estimate Provided	40%	40%
Rotated Factor Pattern Matrix	—	60%
Rotated Factor Structure Matrix	—	60%
Sample Size	100%	100%
Sample Composition	100%	100%

Note. Blank line indicates that information does not pertain to the particular type of factor analysis.

In the present analysis of articles, 60% of the confirmatory studies and 80% of the exploratory studies allowed the factors to be correlated. This finding appears contrary to the analytic traditions previously discussed. A possible explanation of this result is that in several of the exploratory studies oblique rotation strategies were employed because previous research indicated that the factors should be correlated. However, because many of these researchers had theoretical expectations about the structure of the data prior to analysis, confirmatory analysis may have been more appropriate to utilize in these instances. Other authors in the exploratory case indicated that both orthogonal and oblique rotations were utilized initially, but because there were small correlations between the factors, the orthogonal results were interpreted. Similarly, several of the authors that reported utilizing confirmatory analyses contended that non-zero correlations among the factors were indicated in previous research. Based on the results of the present study, however, there is little support provided for the notion that exploratory analyses typically use orthogonal rotations.

Another tradition that is often evidenced in factor analytic studies is a propensity to utilize one matrix of association more often than another. For instance, most exploratory analyses typically utilize correlation matrices in computations whereas most confirmatory analyses use variance-covariance matrices. Several authors have indicated that variance-covariance matrices are often more appropriate to utilize with confirmatory techniques whereas the correlation matrix is often more appropriate in exploratory analyses (Bollen, 1989; Thompson & Daniel, 1996).

In the present review, it was difficult to evaluate this dynamic, as only 40% of the articles reported the matrix of association utilized in the analysis. Of those CFA studies reporting which matrix was used, 30% employed the correlation matrix in the analysis compared with only 10% that utilized the variance-covariance matrix. Similarly, 20% of the exploratory analyses reported using the correlation matrix in computing the factor analysis. While the trend in the exploratory analyses is consistent with analytic tradition, the trend for the confirmatory analyses is contrary to recommendations in the literature. Further, the present results appear to coincide with the Tinsley and Tinsley (1987) and Hetzel (1996) reviews, in that very few of the articles examined contained elements that have been deemed essential to proper external interpretation. However, one caveat to the interpretation of the present results is that the number of articles reviewed was very small (five each of EFA and CFA), which means that a change of one article translates to a 20% increase or decrease in frequency.

The conceptual point of this brief analysis of reporting practices is to help researchers understand the importance of accurately reporting essential elements of factor analytic studies. Because science is built upon the

accurate collection and analysis of data, it is important that results be conveyed to others in a format that is understandable and which permits the independent analysis of empirical results.

Factor Analysis in Educational Applications

Indeed, as demonstrated throughout the present article, both EFA and CFA are important and powerful analytic tools for developing theories or psychological instruments and when evaluating data structure based on expectations or previous research considerations. These analytic techniques are generally flexible and versatile and offer the informed researcher many analytic avenues to pursue when simply summarizing and simplifying data or when developing and testing new psychological constructs. As suggested by Hetzel (1996),

> factor analysis . . . can be employed in substantive investigations to minimize distortions induced by measurement error, and can provide direct assistance in psychological theory building. . . . [F]actor analysis has perhaps been most useful in testing theories of structure, developing theories about structure, and evaluating the construct validity of scores used in substantive inquiry. (p. 176)

These latter uses of factor analytic techniques are often most applicable to educational researchers, as it is often necessary to develop instruments to measure phenomena of interest or to assess empirically the accuracy of extant theories.

The present article described several situations in which both EFA and CFA are valuable to an educational researcher. The employed heuristic data set contained the responses of students on several different psychological tests. The results of these analyses could have been used to determine if certain subject or psychological areas were adequately represented in the sample test battery, or if the theoretical expectations of the researcher were able to withstand actuarial empirical scrutiny. Further, factor analytic techniques can be used to explore the items assembled on a test to determine if the items that should be grouped together (again either based on theory or previous research) are actually grouped together during the analysis based on the latent dimensions underlying the measure. Thus, the application of these techniques to educational areas is essentially endless; however, based on the considerations presented in the present article, some factor analytic techniques are more appropriate to use in some research situations than in others.

General Summary and Conclusions

The present article provided an introductory treatment of both exploratory and confirmatory factor analytic techniques. The pertinent aspects of each of the analyses have been illustrated and the conceptual explanations have been illuminated using concrete heuristic examples. In addition, exemplars of both techniques published in recent issues of a prominent psychology journal have been examined in regard to analytic traditions and practices of reporting indigenous to each technique.

Factor analysis remains a useful and viable analytic tool in social science research. Exploratory and confirmatory techniques allow researchers to develop and critically examine theories regarding the structure of data sets. Several critical decisions must be rendered throughout the analytic process, and these analyses require careful and thoughtful consideration on the part of the researcher. When utilized properly, however, factor analysis can be a powerful analytic tool for both the purposes of data exploration and theory confirmation.

References

Bentler, P. M. (1990). Comparative fit indices in structural models. *Psychological Bulletin, 107*, 238-246.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Byrne, B.M. (1989). *A primer of LISREL*. New York: Springer-Verlag.

Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows*. Thousand Oaks, CA: Sage.

Campbell, T. C., Gillaspy, J. A., & Thompson, B. (1995). *The factor structure of the Bem Sex-Role Inventory (BSRI): A confirmatory factor analysis*. Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas. (ERIC Document Reproduction Service No. ED 380 491)

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*, 245-276.

Cattell, R. B. (1978). *The scientific use of factor analysis in behavioral and life sciences*. New York: Plenum.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin, 70*, 426-433.

Cole, D. A., & Jordan, A. E. (1989). Assessment of cohesion and adaptability in component family dyads: A question of convergent and discriminant validity. *Journal of Counseling Psychology, 36*, 456-463.

Cudeck, R. (1989). The analysis of correlation matrices using covariance structure models. *Psychological Bulletin, 105*, 317-327.

Daniel, L. G. (1989, November). *Comparisons of exploratory and confirmatory factor analysis*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Little Rock. (ERIC Document Reproduction Service No. ED 314 447)

Fan, X., Thompson, B., & Wang, L. (1996, April). *The effects of sample size, estimation methods, and model specification on SEM fit indices*. Paper presented at the annual meeting of the American Educational Research Association, New York. (ERIC Document Reproduction Service No. ED 400 336)

Gillaspy, J. A. (1997). *The factor structure of the working alliance inventory: Confirmatory and exploratory factor analysis*. Unpublished doctoral dissertation, Texas A&M University, College Station, TX.

Good, G. E., Robertson, J. M., O'Neil, J. M., Fitzgerald, L. F., Stevens, M., DeBord, K. A., & Bartels, K. M. (1995). Male gender role conflict: Psychometric issues and relations to distress. *Journal of Counseling Psychology, 42*, 3-10.

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.

Guilford, J. P. (1946). New standards for test evaluation. *Educational and Psychological Measurement, 6*, 427-439.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika, 19*, 149-161.

Hayes, T. J., & Tinsley, H. E. A. (1989). Identification of the latent dimensions of instruments that measure perceptions of an expectations about counseling. *Journal of Counseling Psychology, 36*, 492-500.

Hetzel, R. D. (1996). A primer on factor analysis with comments on patterns of practice and reporting. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 4, pp. 175-206). Greenwich, CT: JAI Press.

Holzinger, K. J., & Swineford, F. (1939). *A study in factor analysis: The stability of a bi-factor solution* (No. 48). Chicago: University of Chicago.

Jöreskog, K. G. (1966). Testing a simple structure hypothesis on factor analysis. *Psychometrika, 31*, 165-178.

Jöreskog, K. G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika, 34*, 183-202.

Jöreskog, K. G. & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago: SPSS.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika, 23*, 187-200.

Knapp, T. R. (1978). Canonical correlation analysis: A general parametric significance testing system. *Psychological Bulletin, 85*, 410-416.

Larson, L. M., Suzuki, L. A., Gillespie, K. N., Potenza, M. A., Bechtel, M. A., & Toulouse, A. L. (1992). Development and validation of the counseling self-estimate inventory. *Journal of Counseling Psychology, 39*, 105-120.

Lee, R. M., & Robbins, S. B. (1995). Measuring belongingness: The social connectedness and the social assurance scales. *Journal of Counseling Psychology, 42*, 232-241.

Marcoulides, G. A., & Schumaker, R. E. (Eds). (1996). *Advanced structural equation modeling: Issues and techniques*. Mahwah, NJ: Lawrence Erlbaum.

Mulaik, S. A. (1966). Inferring the communality of a variable in a universe of variables. *Psychological Bulletin, 66*, 119-124.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 6*, 559-572.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design, and analysis: An integrated approach*. Hillsdale, NJ: Erlbaum.

Rice, K. G., Cole, D. A., & Lapsley, D. K. (1990). Separation-individuation, family cohesion, and adjustment to college: Measurement validation and test of a theoretical model. *Journal of Counseling Psychology, 37*, 195-202.

SAS Institute, Inc. (1990). *SAS/STAT user's guide: Volume 1, ANOVA-FREO* (4th Ed.). Cary, NC: Author.

Snook, S. C., & Gorsuch, R. L. (1989). Component analysis versus common factor analysis: A Monte Carlo study. *Psychological Bulletin, 106*, 148-154.

Sodowsky, G. R., Taffe, R. C., Gutkin, T. B., & Wise, S. L. (1994). Development of the multicultural counseling inventory: A self-report measure of multicultural competencies. *Journal of Counseling Psychology, 41*, 137-148.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

Stevens, J. (1996). *Applied multivariate statistics for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.

Tabachnick, B. G., & Fidell, L. S. (1983). *Using multivariate statistics*. New York: Harper & Row.

Thompson, B. (1984). *Canonical correlation analysis: Uses and interpretation*. Thousand Oaks, CA: Sage.

Thompson, B. (1992). A partial test distribution for cosines among factors across samples. In B. Thompson (Ed.), *Advances in social science methodology* (Vol. 2, pp. 81-97). Greenwich, CT: JAI Press.

Thompson, B. (1997a, August). *If statistical significance tests are broken/misused, what practices should supplement or replace them?* Invited paper presented at the annual meeting of the American Psychological Association, Chicago.

Thompson, B. (1997b). The importance of structure coefficients in structural equation modeling confirmatory factor analysis. *Educational and Psychological Measurement, 57*, 5-19.

Thompson, B. (1998, July). *The ten commandments of good Structural Equation Modeling behavior: A user-friendly, introductory primer on SEM.* Invited paper presented at the annual meeting of the U.S. Department of Education, Office of Special Education Programs (OSEP) Project Director's Conference, Washington, DC.

Thompson, B., & Borrello, G. M. (1985). The importance of structure coefficients in regression research. *Educational and Psychological Measurement, 45*, 203-209.

Thompson, B., & Daniel, L. G. (1996). Factor analytic evidence for the construct validity of scores: A historical overview and some guidelines. *Educational and Psychological Measurement, 56*, 197-208.

Thurstone, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.

Tinsley, H. E. A., Bowman, S. L., & York, D. C. (1989). Career decision scale, my vocational situation, vocational rating scale and decisional rating scale: Do they measure the same constructs? *Journal of Counseling Psychology, 36*, 115-120.

Tinsley, H. E. A., Roth,. J. A., & Lease, S. H. (1989). Dimensions of leadership and leadership style among group intervention specialists. *Journal of Counseling Psychology, 36*, 48-53.

Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of factor analysis in counseling psychology research. *Journal of Counseling Psychology, 34*, 414-424.

Tracey, T. J., Glidden, C. E., & Kokotovic, A. M. (1988). Factor structure of the Counselor Rating Form-Short. *Journal of Counseling Psychology, 35*, 330-335.

Weiss, D. J. (1971). Further considerations in the application of factor analysis. *Journal of Counseling Psychology, 18*, 85-92.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin, 99*, 432-442.

JOURNAL SUBSCRIPTION FORM

This form can be used to subscribe to *RESEARCH IN THE SCHOOLS* without becoming a member of the Mid-South Educational Research Association. It can be used by individuals and institutions.

◇◇◇

Please enter a subscription to *RESEARCH IN THE SCHOOLS* for:

Name: ____

Institution: ____

Address: ____

____

			COST
Individual Subscription ($25 per year)	Number of years	______	________
Institutional Subscription ($30 per year)	Number of years	______	________
Foreign Surcharge ($25 per year, applies to both individual and institutional subscriptions)	Number of years	______	________
Back issues for Volumes 1 - 5 ($30 per Volume)	Number of Volumes	______	________
TOTAL COST:			________

MAKE CHECKS PAYABLE TO MSERA
SEND FORM AND CHECK TO:

Dr. James E. McLean, Co-Editor
RESEARCH IN THE SCHOOLS
University of Alabama at Birmingham
School of Education, 233 Educ. Bldg.
901 13th Street, South
Birmingham, AL 35294-1250

MID-SOUTH EDUCATIONAL RESEARCH ASSOCIATION

The Mid-South Educational Research Association (MSERA) was founded in order to encourage quality educational research in the Mid-South and to promote the application of quality educational research in schools. Members of MSERA share interests in educational research, development, and evaluation. While most members are from institutions of higher education, many others represent state departments of education, public and private agencies, and public school systems. Graduate students comprise a significant portion of the membership. A majority of MSERA members are from the six states represented by the organization, but others are from many other states and several foreign countries. The MSERA is the largest regional educational research organization in the country.

The organization provides several services for its members. The annual meeting, held every November, offers many formal and informal opportunities for professional development through special training courses, sharing of research findings, and programmatic interests with colleagues. Members receive a subscription to *RESEARCH IN THE SCHOOLS* and the *Mid-South Educational Researcher*. The MSERA also provides recognition and cash rewards for its outstanding paper, an outstanding dissertation, and professional service.

MSERA Membership/Renewal Form

(Please print or type)

Name _____

Organization _____

Address _____

_____

Telephone Work: _____

Home: _____

Fax: _____

e-mail: _____

Amount Enclosed:	MSERA 1999 Membership ($25 professional, $15 student)	$_____
	MSER Foundation Contribution	$_____
	TOTAL	$_____

Make check out to MSERA and mail to:

Dr. Robert Calvery
Flippin School District
P. O. Box 239
Flippin, Arkansas 72634

RESEARCH IN THE SCHOOLS
Mid-South Educational Research Association
and the University of Alabama at Birmingham
901 South 13th Street, Room 233
Birmingham, AL 35294-1250

TM030698

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

NOTICE

Reproduction Basis

This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").