
Volume 17, No. 2—Spring 2004

MID-WESTERN EDUCATIONAL RESEARCHER

• Official Publication of the Mid-Western Educational Research Association •



The Keith Building, Cleveland, OH

On the Cover

The Keith Building, in downtown Cleveland, is an historic landmark that was born in the Age of Vaudeville, erected in 1922. Connected to the Playhouse Square complex and the Palace Theater, it has hosted the likes of Bob Hope, Jack Benny, Bill Robinson, Harry Houdini and more. Because of space needs at Cleveland State University, the Keith Building is now hosting several floors of University operations, including *The Reading First—Ohio Center for Professional Development and Technical Assistance in Effective Reading Instruction*, a collaborative effort of Cleveland State University, John Carroll University, and the University of Akron. The Center is being funded by a six-year federal grant to provide professional development and technical assistance to districts and schools that have been awarded Reading First grants to improve reading achievement in grades K-3. Working with district coordinators, school principals, literacy specialists and teachers within the awarded districts, the Center's mission is to provide professional development to teachers and principals in sound teaching practices based upon the scientifically-based reading research (SBRR). Teachers will be trained to use SBRR principles and strategies in daily 120-minute literacy blocks; principals will be provided workshops to help them to support SBRR practices in their schools and technical assistance to facilitate planning and development of an effective core reading program and complementary support services for children at risk of reading failure.

In its first year, the Center will work with 12 districts and 65 schools state-wide; this represents over 20,000 children and over 1,000 teachers. The goal is to help each school to achieve 100% of their children reading at or above grade level by the end of the third grade. In addition, e-learning modules developed for teachers in Reading First schools will be disseminated state-wide to all teachers of K-3 students through the popular State Institutes for Reading Instruction (SIRI).

Information for Contributors to the Mid-Western Educational Researcher

The *Mid-Western Educational Researcher* accepts research-based manuscripts that would appeal to a wide range of readers.

All materials submitted for publication must conform to the language, style, and format of the *Publication Manual of the American Psychological Association*, 5th ed., 2001

(available from Order Department, American Psychological Association, P.O. Box 2710, Hyattsville, MD 20784).

Four copies of the manuscript should be submitted typed double-spaced (including quotations and references) on 8 1/2 x 11 paper. Only words to be italicized should be underlined. Abbreviations and acronyms should be spelled out when first mentioned. Pages should be numbered consecutively, beginning with the page after the title page. Manuscripts should be less than 20 pages long. An abstract of less than 100 words should accompany the manuscript.

The manuscript will receive blind review from at least two professionals with expertise in the area of the manuscript.

The author's name, affiliation, mailing address, telephone number, e-mail address (if available), should appear on the title page only. Efforts will be made to keep the review process to less than four months.

The editors reserve the right to make minor changes in order to produce a concise and clear article.

The authors will be consulted if any major changes are necessary.

Manuscripts should be sent with a cover letter to:

James A. Salzman, *MWER* Co-Editor
Cleveland State University, Rhodes Tower Rm. 1343, 2121 Euclid Avenue, Cleveland, OH 44114

The *Mid-Western Educational Researcher* (ISSN 1056-3997) is published quarterly by the MidWestern Educational Research Association through The Ohio State University. The Summer issue serves as the annual meeting program. Non-profit postage paid at Columbus, Ohio, with permission of the College of Education, Daryl Siedentop, Interim Dean.
POSTMASTER: Send address change to Jean W. Pierce, Dept. EPCSE, Northern Illinois University, DeKalb, IL 60115.

MID-WESTERN EDUCATIONAL RESEARCHER

Volume 17, Number 2 · Spring 2004

ISSN 1056-3997

Co-Editors

James A. Salzman
Cleveland State University
email: j.salzman@csuohio.edu

Jane Zaharias
Cleveland State University
email: j.zaharias@csuohio.edu

Editorial Advisory Board

Louise Fleming
Ashland University

Tom Ganser
University of Wisconsin
Whitewater

Richard Lipka
Pittsburg State University

Jackie Rickman
Western Illinois University

Ayres D'Costa
The Ohio State University

Isadore Newman
University of Akron

Anne D'Antonio Stinson
University of Wisconsin
Whitewater

Consulting Editors

Mary Bendixen-Noe
Managing Editor
The Ohio State University,
Newark

Gene A. Kramer
American Dental Association

Richard M. Smith

Layout and design

Judy E. Teska
JET Desktop Solutions

*Using SEM to Develop a Model for Educators
of Children Who are Exposed to the Trauma of Violence*
Rosalind Duplechain, State University of West Georgia

2

*Community Awareness Project: A Tool
for Preparing Culturally Sensitive Teachers*
Barbara A. Rucker, Cleveland State University

12

*Mid-Western Educational Research Association
Annual Meeting Announcements*

21

*Creating the Rosetta Stone: Deciphering the Language
of Accountability to Improve Student Performance*
Donna Snodgrass, Cleveland Municipal School District

25

Two Methods of Estimating a Study's Replicability
Isadore Newman, University of Akron
Keith McNeil, New Mexico State University
John Fraas, Ashland University

36

MWER Publication Address

James A. Salzman
Cleveland State University
Rhodes Tower 1343
2121 Euclid Avenue
Cleveland, OH 44114
Phone: (216) 687-5048
Fax: (216) 687-5379
email: j.salzman@csuohio.edu

MWER Subscription & MWER Membership Information

Jean W. Pierce
Dept. EPCSE
Northern Illinois University
DeKalb, IL 60115
Phone: (815) 753-8470
Fax: (815) 753-9250
email: jeanpierce@aol.com

Using SEM to Develop a Model for Educators of Children Who are Exposed to the Trauma of Violence

Rosalind Duplechain
State University of West Georgia

Abstract

This study used Structural Equation Modeling (SEM) to develop a teacher-friendly model for understanding the relationship between children's exposure to traumatic events, such as violence exposure, and their school achievement. With a sample of 146 urban elementary (grades 2-5) children, three variables (mental health behaviors, performance-related behaviors, and standardized achievement) were examined. The model suggested that children who are exposed to violence in Year I were more delinquent and ultimately underachieved in reading in Year II. This model should help teachers identify students exposed to violence and address their social and academic needs.

Imagine being expected to complete a complex task in a foreign language. In other words, imagine asking an American medical doctor to diagnose and treat a medical problem using textbooks that were written in Ukrainian. While the needed information is available in the textbooks, this information is inaccessible because it is in a language that is typically unfamiliar to the doctor. Thus, asking an American medical doctor to diagnose and to treat medical problems using a foreign language could greatly compromise the patient's medical health.

Yet this is exactly what the literature of trauma asks of teachers each day. Teachers, who turn to the literature on trauma, are being asked to identify and to assist children who have been exposed to traumatic events using the foreign language of psychology. This trauma literature tells teachers that children who are exposed to traumatic events, such as violence, experience a number of psychological problems like depression and anxiety. However, because teachers are trained to focus on achievement and behaviors related to achievement, providing them with psychological information is the same as providing American medical doctors with Ukrainian textbooks.

Not having teacher-friendly information for identifying and assisting traumatically exposed children is especially unfortunate given that many of today's children must cope with traumatic events. September 11, 2001, notwithstanding, today's children have witnessed, some first-hand and others vicariously, eight schools being attacked by fellow students (i.e., Santee, Heritage High, Columbine High, Thurston High, Westside Middle, Heath High, Pearl High, and Frontier Junior High). In addition, some students have lost close friends in these events. Moreover, Shakoor and Chalmers (1991) found that of the 1,035 students that they surveyed, ages 10-19, 75% of the boys and 70% of the girls witnessed someone being shot, stabbed, robbed or killed. Uehara, Chalmers, Jenkins, and Shakoor (1996) surveyed 1,011 students, 10-19 year-olds, and found that 55% had

witnessed someone being robbed, 35% had witnessed a stabbing, and 39% had witnessed someone being shot, and nearly 24% had witnessed someone being killed.

According to the *Diagnostic and Statistical Manual of Mental Disorder* (DSM-IV, 1994), when children witness, experience, or learn of significant others who have experienced these types of events, they have been exposed to trauma, that is, serious threats to one's physical and/or psychological well being.

While abuse of any kind, neglect, the sudden loss of a significant other, and disasters all meet the DSM-IV (1994) criteria of a trauma, so does violence exposure. When school-aged children are exposed to violence, they have been shown to experience the typical symptoms associated with trauma, post-traumatic stress syndrome (PTSD; (Dyson, 1990; Pynoos, Fredrick, Nader, Arroyo, Steinberg, Eth, Nunez, and Fairbanks, 1987; Pynoos and Eth, 1984). Therefore, violence exposure, the focus of this paper, is a type of trauma and is quite prevalent in the lives of today's children (Shakoor and Chalmers (1991).

Consequently, if educators are to successfully identify and assist children who have been exposed to the trauma of violence, they will need information, a model, that is consistent with their own concerns, goals, and expertise. Such a model could help educators with this identification of traumatically exposed children and guide them in assisting these children. Otherwise, by expecting teachers to identify traumatically exposed students using the unfamiliar language of psychologists, we risk the likelihood of identifying at-risk students when the effects of exposure have become more complex. Thus, research that would focus on the language and concerns of the educator could greatly benefit children who are exposed to traumatic events. After all, it is the educator, not the psychologist who has the greater likelihood of interacting with children on a daily basis and consequently

of identifying at-risk students and of initiating assistance on their behalf.

This study has two purposes. First, this study adds its voice to similar studies (Eckenrode, Laird, and Doris, 1993; Dyson, 1990; Wodarski, Kurtz, Gaudin, and Howing, 1990) with the goal of contributing to our understanding of children who are traumatically exposed. Second, this study marks a step in the direction of developing a teacher-friendly model to explain the educational impact of traumatic events on the achievement of children. Such a model should help teachers to identify traumatically exposed students and suggest the types of academic and social behaviors that they can target when working with children who are traumatically exposed.

The Proposed Model: A Psycho-Educational Explanation of Traumatic Exposure

This study tested the validity of a proposed model. This model represents a balance between what the trauma literature has already provided (i.e., that trauma affects mental functioning) and that which the educational field awaits (Does trauma, such as violence, impede achievement? If so, how? What role can educators play to assist traumatized children and minimize educational effects?). Drawn from empirical and clinical evidence within the literature on trauma and on achievement, the proposed model examines the relationship among four variables: a) violence exposure (one form of traumatic event that was previously discussed), (b) mental health behaviors, (c) performance-related behaviors, and (d) achievement.

Violence exposure. Violence exposure was included in this model because it is a form of trauma; violence exposure meets the criteria of being a serious threat to a child's well being. Violence exposure involves witnessing any violent act and knowing a significant other who was victimized. In addition, violence has been linked to poor school performance (Dyson, 1990; Pynoos and Eth, 1984).

Mental health behaviors. Mental health behaviors such as thought problems, social problems, delinquency, and aggression are included in this model for two reasons. First, trauma has been linked to many types of mental health problems including thought problems (Pynoos, Fredrick, Nader, Arroyo, Steinberg, Eth, Nunez, and Fairbanks, 1987), social problems (Dyson, 1990; Shanok, Welton, and Lapidus, 1989; Salzinger, Kaplan, Pelcovitz, Samit, and Krieger, 1984), delinquency (Dyson, 1990; Pynoos and Eth, 1984; Wordarski, Kurtz, Gaudin and Howing, 1990), and aggression (Dyson, 1990; Einbender and Friedrich, 1989). Second, mental health is included in this model because some studies suggest a relationship between mental health behaviors and school-related tasks (Shanok, et al., 1989) and between mental health behaviors and school achievement (Dyson, 1990; Shanok, et al., 1989; Pynoos, et al., 1987).

Operational definitions of these mental health behaviors are consistent with Achenbach (1991) subscales. Achenbach (1991) defines thought problems using the fol-

lowing criteria: "can not get mind off certain thoughts; deliberately harms self; fears certain animals, situations, places; hears things, repeats act; sees things; strange behavior; strange ideas" (p. 49). Social problems are defined as "acts young, clings, lonely, cries, not get along, unloved, out to get, worthless, get hurt, teased, not liked, clumsy, prefers young" (p. 49). Delinquent behaviors include: "doesn't seem to feel guilty after misbehaving, hangs around with bad company, lying or cheating, prefers older company, steals, swears, tardy to school or class, truant, uses alcohol or drugs" (p. 50). Aggression includes behaviors such as: "argues, defiant, brags, mean, demands attention, destroys own things, destroys property belonging to others, disobedient at school, disturbs other people, jealous, fights, talks out, attacks others, disrupts, screams, shows off, explosive temper, stubborn, mood change, talks much, teases, threatens" (p. 50).

Performance-related behaviors. Performance-related behaviors, specifically on task and oral academic participation, are included in this model for three reasons. First, there is empirical and clinical evidence that have demonstrated the relationship between on task behavior and achievement (Stallings, Johnson, and Goodman, 1985; Guida, Ludlow, and Wilson, 1985, Johnson and Butts, 1983). A similar relationship has been documented between classroom participation and achievement (Finn and Cox, 1992; Loranger and Picard, 1981; Morine-Dershimer, Galluzzo, and Fagal, 1980).

Second, the performance related behaviors of school children can be influenced by traumatic exposure. For example, in their description of an emotionally neglected preschooler, Shanok and her colleagues (1989) reported an increase in the performance-related tasks of writing, making books, dictating sentences and making pictures only after they had worked with her. In addition, Friedrich, Einbender and Luecke (1983) studied the relationship of physical abuse to on task behavior (performance-related behavior). Although the relationship between traumatic exposure and on task behavior was not supported, their inquiry suggests a possible relationship that warrants more systematic investigation especially since Shanok and her colleagues (1989) have reported a link between trauma and performance-related tasks.

Finally, performance-related behaviors are included in this model because of their potential to direct teacher efforts at helping their traumatically exposed students. If these behaviors are found to significantly impact achievement, then these and similar behaviors may provide direction for future educational interventions and ultimately prove helpful in minimizing any negative educational effects of traumatic exposure.

Student achievement. Student achievement is the ultimate dependent variable of this model. Traumatic exposure has been linked to low achievement scores in some trauma studies (Eckenrode, et al., 1993; Salzinger, et al., 1984; Pynoos and Eth, 1984). Further, achievement scores are the

major criterion for which educators are held accountable and by which students are provided with special services. Finally, achievement scores are included in the proposed model because there is a “strong relationship between successful adult adjustment and school success in our culture” (Brassard and Gelardo, 1987, p. 131).

Thus, the model proposes that there is a chain of psychological and educational reactions that occur as a result of traumatic exposure. This study tested the validity of this proposed model. If the proposed model holds true, then it will describe children using educational criteria/concerns instead of psychological jargon thereby providing teachers with a familiar language by which they can identify children exposed to trauma. In addition, if the proposed model holds true, then this model will suggest school-related behaviors that teachers can target in their attempts to assist children exposed to trauma. Finally, Model 1 hypothesizes that exposure to violence influences one’s mental health, which in turn limits engagement in performance-related behaviors, and ultimately hinders school achievement.

Methodology

This study was a secondary analysis of an extensive database provided by the Metropolitan Area Child Study (MACS; Guerra, Eron, Huesman, Tolan, and Van Acker, 1990). All participants, measures, and assessment procedures used in this study were those used by MACS.

The MACS project is a large-scale, longitudinal study conducted in selected Chicago and Aurora schools. Participants include more than 4,000 school-aged children. Focusing on high-risk urban children, MACS was designed to prevent the development of antisocial and violent behavior in children and to promote the development of social skills as a way of curbing youth violence. MACS was funded by grants from the National Institute of Mental Health (NIMH) and the Centers for Disease Control and Prevention (CDC).

Participants

The initial sample for the current study consisted of all the students for whom MACS had collected trauma data. These 880 elementary school students ranged in ages from six to thirteen and were drawn from eight inner-city, mid-western elementary schools. From these students, a sub-sample was taken. This sub-sample included all students who had data for all four variables within the proposed model. Thus, the final sample for this analysis consisted of 146 students.

Of these 146 students, 42 (29%) of them were second graders, 53 (36%) were third graders and 51 (35%) were fifth graders. Eighty-nine of these children (61%) were male and 57 (39%) were female. In addition, forty-two children (29%) were African-American, fifty-six (38%) were Caucasian, forty-seven (32%) were Hispanic, and one child (1%) was Other. Lastly, as measured by the free lunch program,

forty-five of these students (31%) were from low SES families, thirteen (9%) were from middle SES families, and sixty-six (45%) were from high SES families, and twenty-two (15%) were missing SES data (see Table 1).

Table 1
Demographic Data of Sample

Research Groups *	1	2	3	Total
N=	27	77	42	146
GENDER				
Females	41%	42%	33%	39%
Males	59%	58%	67%	61%
ETHNICITY				
African-Americans	26%	26%	36%	29%
Caucasians	63%	38%	24%	38%
Hispanics	11%	35%	40%	32%
Other	0%	1%	0%	1%
AGE				
6	0%	0%	2%	1%
7	0%	0%	0%	0%
8	15%	12%	14%	13%
9	30%	34%	29%	32%
10	30%	17%	17%	19%
11	7%	21%	7%	14%
12	19%	16%	24%	18%
13	0%	1%	7%	3%
GRADE				
2	26%	26%	36%	29%
3	48%	36%	29%	36%
4	0%	0%	0%	0%
5	26%	38%	36%	35%
SES				
Low	19%	34%	33%	31%
Middle	15%	6%	10%	9%
High	56%	45%	38%	45%
Missing	11%	14%	19%	15%

Measures/Instruments and Assessment Procedures

Three measures and naturalistic observation were used to gather data for this study. The three measures were: the *Traumatic Exposure Scale*, the *Child Behavior Checklist Form*, and the standardized achievement test subscales for mathematics and reading.

Traumatic Exposure Scale. The *Traumatic Exposure Scale* measures a child’s exposure to traumatic events through self-reports. This scale was taken from a self-report questionnaire, the *Social Stress Measure*, which was designed by Tolan, Miller and Thomas (1988; *Social Stress Measure*). The *Social Stress Measure* consisted of 15 items that were divided into 5 subscales: circumscribed events, daily hassles, violence, transitions, and school problems. Embedded within those 15 items were five items that met the criteria as indices of traumatic events, that is serious threats, to a child’s well being (DSM-IV, 1994). These five items measure vio-

lence exposure: witnessing violence, knowledge of someone who was victimized, and behavioral change prompted by fear of violence. For the purposes of the present study, these five items formed a sixth sub-scale, the *Traumatic Exposure Scale* (see Table 2). Internal consistency for this scale as measured by the Spearman-Brown coefficient was .72 and the Cronbach Alpha was .62.

Table 2
Traumatic Exposure Scale

DIRECTIONS:

Answer the following questions by circling "YES" or "NO".

1. During the last year, has someone else you know, other than a member of your family, gotten beaten, attacked or really hurt by others? YES NO
2. During the last year, have you seen anyone beaten, shot or really hurt by someone? YES NO
3. During the last year, have you seen or been around people shooting guns? YES NO
4. During the last year, have you been afraid to go outside and play, or have your parents made you stay inside because of gangs or drugs in your neighborhood? YES NO
5. During the last year, have you had to hide someplace because of shootings in your neighborhood? YES NO

Traumatic exposure data were collected in Spring 1991-92 by a team of three individuals extensively trained in the administration of this measure with classrooms of bilingual children. Students completed the index by indicating whether or not they had experienced any of the traumatic events listed during the 1991-92 school year. Based on the total number of "yes" responses, each child was assigned to either the control group or the exposure group.

Child Behavior Checklist: Teacher Report Form. The second measure used in this study was the *Child Behavior Checklist* (CBCL-TRF; Achenbach and Edelbrock, 1986) that measures the mental health-related behaviors. The CBCL-TRF describes the social/emotional development of children and adolescents. This measure contains a list of 119 items describing eight conditions: 1) withdrawn, 2) somatic complaints, 3) anxious/depressed, 4) social problems, 5) thought problems, 6) attention problems, 7) delinquent behavior, and 8) aggressive behavior. Estimates of reliability are .89 (test-retest) and .57 (inter-rater reliabilities between teachers and classroom aids) (Elliott and Busse, 1992). Finally, concurrent validity with the *Conners Revised Teacher Rating Scales* is .85 (Elliott and Busse, 1992).

For this study only, four mental health conditions were investigated: aggression, delinquency, thought problems, and social problems. These conditions were used because the literature suggests a relationship between these mental health conditions and violence (Dyson, 1990; Shanok, et al., 1989; Pynoos, et al., 1987).

By Spring of the 1992-93 school year, the primary classroom teacher for each student completed the *Teacher Report Form* (TRF) of the CBCL for each child. These teachers

considered a list of statements that described each student "now or within the past two months" and then rated each description as it applied to individual students using a 3-point Likert rating (1 = not true; 2 = somewhat or sometimes true; 3 = very true or often true). Teachers were given a small monetary award for completing each CBCL-TRF assessment.

Standardized academic achievement. Academic achievement was measured using the percentile scores of the math and the reading subscales obtained from one of three standardized tests: the Iowa Test of Basic Skills (ITBS), the California Achievement Tests (CAT), or La Prueba de Riverside en Espanol (La Prueba). The La Prueba is a standardized achievement test that is given to Spanish speaking children. All three instruments are valid and reliable measures of academic skills (Raju, 1992; Airasian, 1989; Chicago Public Schools, 1998). These achievement tests were administered at each individual school site in the spring of the school year and results were obtained from the schools for the 1992-93 school year.

Naturalistic observation. This study also included the naturalistic observations conducted by MACS personnel. Of the many student behaviors that MACS investigated, two specific behaviors were of interest to the current study: on task and classroom participation. MACS defined on task behavior as any overt motor behavior, vocal/verbal behavior, or gesture that is related to the completion of the assigned task and defined classroom participation as volunteering to answer questions, calling out, and being called upon. These behaviors are of interest because, as previously discussed, on task behavior and classroom participation are positively correlated with school achievement (Finn and Cox, 1992; Stallings, Guida, Ludlow, and Wilson, 1985; Johnson). However, the relationship between these performance-related behaviors and school achievement has not been well explored among populations of traumatically exposed children.

Collection of these data occurred during the Spring of the 1992-93 school term. Each child was observed for two 20 minute sessions during a variety of school activities (i.e., within the classroom, during lunch, during recess, etc.) to gather a range of student behaviors. All of these behaviors were coded and assigned a specific key on the keyboards of pre-programmed laptop computers. These computers were programmed for real-time, multiple entry data collection (Repp, Harman, Felce, Van Acker, and Karsh, 1989; Van Acker, Bush, and Grant, 1992). Thus, these computers were capable of calculating the total frequency, the starting, ending and duration times as well as the percentage of the total session time for coded behaviors.

Behavioral observers were extensively trained by the MACS project. These observers reported 100 percent accuracy of definitions, 95 percent accuracy of written examples, and 85 percent inter-rater reliability for more than three consecutive sessions with no individual code falling below an 80 percent agreement among raters. Agreement was estab-

lished for both duration and frequency of the responses prior to data collection.

Data Analysis

Structural equation modeling (SEM) was the statistical approach used to test the relationships among all four variables: violence exposure, mental health behaviors, performance-related behaviors, and achievement. The LISREL 8 (Joreskog and Sorbom, 1993) computer program was used in conducting this analysis.

Maximum likelihood (ML) was the estimation method that was used in all LISREL analyses because it has been shown to perform reasonably well under unfavorable conditions such as non-normality (Hoyle and Panter, 1995). Additionally, chi-square (χ^2) and Goodness-of-Fit Index (GFI) are the absolute indices that were used to determine the overall fit of all models in this study. The typical critical value of .9 or more determined the appropriateness of the GFI (Hoyle and Panter, 1995). Chi-square and GFI were used because chi-square has been shown to “exert less impact” when multivariate normality is violated (p. 95) and GFI was shown to “behave consistently across ML [and other estimation procedures] at all sample sizes” (Hu and Bentler, 1995, p. 92). Both of these indices of absolute fit are indicators of the “degree to which the covariances... specified in the [proposed or alternative] model match the observed covariances” of the sample (Hu and Bentler, 1995, p. 165).

According to the LISREL manual (Joreskog and Sorbom, 1993), generating models based on a tentative initial model is an acceptable use of SEM. The goal of such an inquiry is to find a model that not only fits the data well from a statistical point of view but also has substantive meaning. For the purposes of this study, a model of “substantive meaning” will be any model in which violence exposure is linked to any school-related behavior that teachers could target, such as on task behavior, classroom participation, and academic achievement, in an effort to minimize the adverse influence of traumatic exposure in the lives of their students.

Five steps were taken in order to determine if a model with the proposed variables could both statistically and substantively describe the sample of students in this study. First, a model was defined as suggested by relevant literature and by variables of interest. Second, the correlations, standard deviations, and means of the variables in the proposed model were inputted into the LISREL program, the relationships among the variables were described as LISREL equations, and the program was run. Third, statistical results of the proposed model were examined and any relationships (paths) that were not statistically significant (t -value $< \pm 1.96$), were removed, creating an alternative model which was then re-run, or re-estimated, using the LISREL program (A. Conway, personal communications, March 31, 1997 and April 7, 1997). Fourth, results of the alternative model were examined for statistical criteria of fit. If this model was found to be statistically representative of the sample, then this model

was examined further for substantive meaning. Fifth, only models with both statistical and substantive meaning were retained.

Descriptive statistics. For each of the variables within the proposed model, a summary of the means, standard deviations, ranges, skew and kurtosis are presented (see Table 3). Examination of these data suggests that several variables deviate from the normal curve; some of these deviations demonstrate a curve that is skewed to the right of what is typically thought of as a normal curve and other deviations demonstrate a curve with an abnormal height (kurtosis). All of these variables show some skew, however, only two variables demonstrate excessive skew (>2): thought problems and social problems. Additionally, only the social-problems variable is without kurtosis and only the thought problems variable shows excessive kurtosis (>7). To lessen these types of problems with behavioral data, square-root transformations are customarily performed. However, this corrective procedure was not deemed necessary for this study. To the contrary, extreme behavioral scores were anticipated and explored for any potential relationship that they might have with traumatic exposure.

Table 3
Summary of Means, Ranges, Skew, and Kurtosis

	Mean	Range	Skew	Kurtosis
Violence	2.1	1-3	-.13	-.84
CBCTHO3S	.9	0-12	3.08	9.65
CBCSOC3S	2.4	0-18	2.16	5.05
CBCDEL3S	2.0	0-12	1.30	1.58
CBCAGG3S	9.9	0-41	1.10	.29
ONTASK23	743.2	0-1198	-.65	-.09
PARTICI23	1.8	0-12	1.66	2.76
MATH23	52.0	1-99	-.03	-1.06
READ23	45.7	1-99	.19	-.97

Note: Violence = Violence Exposure; CBCTHO3S = Thought problems; CBCSOC3S = Social problems; CBCDEL3S = Delinquency; CBCAGG3S = Aggression; ONTASK23 = Ontask behavior; PARTICI23 = Classroom participation; MATH23 = Mathematics achievement; and READ23 = Reading achievement.

SKEW ³ 2 = highly non-normal; KURTOSIS ³ 7 = highly non-normal.
n=145

Correlation matrix. Correlation among the variables within the proposed model was explored. Results, presented in Table 4, revealed several significant relationships ($p < .05$). A significant, but low, positive correlation (McMillian, 2000) was revealed between violence exposure and delinquency ($r = .16$). In addition, significant, but low, negative correlations were shown between delinquency and reading achievement ($r = -.24$) and between social problems and math achievement ($r = -.15$). These findings suggest that increases in violence exposure are associated with increases in delinquent behavior; that elevated levels of delinquent behavior are related to lower reading achievement scores; and, lastly,

that increases in social problems are correlated with lower math achievement.

Finally, significant, positive inter-correlations were also found among the mental health behaviors, between on task and participation ($r = .22$), and between reading and math subscales of achievement ($r = .70$). These findings suggest several significant relationships: the more thought problems children have, the more they will also have social problems, delinquency problems, and aggression problems; the more social problems children have, the more they will exhibit delinquency problems; and the more delinquency problems children manifest, the more they will also demonstrate aggression problems. Additionally, the more problems children have participating in the classroom the more they will have difficulty being on task in this classroom. Lastly, the higher reading achievement children have, the higher their mathematics achievement will be.

Results

Ten models were tested for both statistical and substantive meaning in this analysis. With the proposed model as

the initial model, four more models were proposed and tested (Models 3, 5, 7 and 9). Each of the four proposed models were saturated, that is, models that indicate relationships among all variables. The other five (Models 2, 4, 6, 8, and 10) were alternatives to the proposed and saturated models and resulted from the removal of insignificant paths within the saturated models. A comparison of the data for statistical overall-fit suggests that all of the alternative models except Model 2 are plausible representations of the students in this sample (see Table 5). However, an examination of the relationships (paths) within the plausible models suggests that only one of these is both statistically and substantively meaningful, Model 8.

Statistically, the overall fit measures for Model 8 suggest a good fit. The overall chi-square value was not significant, $\chi^2(7, N=145)=2.61, p>.05$, indicating that the observed model is a plausible one for this population. In addition, the second absolute index of fit was appropriately above .9 (GFI = .99). The Type 2 and 3 indices of fit suggest that Model 8 is a better fit than the independence model (IFI=2.26, CFI=1.00) and the root mean square residual (RMR) was appropriately low (.05).

Table 4
Correlation Matrix

	1	2	3	4	5	6	7	8	9
1. Violence	1.000								
2. CBCTHO3S	.112	1.000							
3. CBCSOC3S	-.027	.712*	1.000						
4. CBCDEL3S	.167*	.389*	.361*	1.000					
5. CBCAGG3S	.075	.470*	.504*	.740*	1.000				
6. ONTASK23	-.010	.014	-.132	-.054	-.100	1.000			
7. PARTICI23	.005	-.081	-.086	-.095	.063	.223*	1.000		
8. MATH23	.035	-.060	-.158*	-.147	-.006	.037	.025	1.000	
9. READ23	-.008	-.069	-.044	-.241*	-.013	.050	.097	.701*	1.000
<i>SD</i>	.6	2.1	3.5	2.3	10.5	274.8	2.4	27.2	26.1

Note: Violence = Violence Exposure; CBCTHO3S = Thought problems; CBCSOC3S = Social problems; CBCDEL3S = Delinquency; CBCAGG3S = Aggression; ONTASK23 = Ontask behavior; PARTICI23 = Classroom participation; MATH23 = Mathematics achievement; and READ23 = Reading achievement.

n = 145

* $p < .05$

Table 5
Indices-of-Fit for Alternative Models

Model	χ^2	GFI	IFI	CFI
2. Violence, mental health, performance behaviors, and achievement.	110.78	.86	.77	.76
4. Violence, aggression, ontask, participation, and reading.	4.91*	.99	2.26	1.00
6. Violence, aggression, ontask, participation, and math.	3.8*	.99	3.42	1.00
8. Violence, delinquency, ontask, participation, and reading.	2.61**	.99	1.28	1.00
10. Violence, aggression, ontask, participation, and reading.	5.35*	.99	1.30	1.00

Note: GFI – Goodness-of-Fit Index; IFI = Incremental Fit Index; CFI = Comparative Fit Index. Models 3, 5, 7, and 9 are saturated models. Their fit-statistics include a chi-square of 0, with 0 degrees of freedom, and $p = 1.00$.

n = 145

* Plausible models based on statistical results ($p > .05$).

** Plausible models based on both statistical and substantive meaning.

Examination of other fit criteria lends more support for Model 8. First, the chi-square for the independence model was rejected, $\chi^2(10, N=145)=22.64$, which suggests that the observed variables are uncorrelated with each other. Second, the root mean square error of approximation (RMSEA) is 0, the 90 percent confidence intervals are 0 to .04 [within the .05 - .08 limit suggested by Browne and Cudeck (1993)], and the p-value for the test of close fit (RMSEA < .05) is .97. Together, these indicators of fit suggest that there is very little discrepancy between the population and the precision of the fit measure itself.

One final indication that Model 8 describes this population of students well is the index value of the Expected Cross-Validation Index (ECVI). Specifically, the ECVI for Model 8 (.13) is less than the ECVI for the saturated model (.21) and the independence model (.23), which suggests that it is highly likely that Model 8, which was produced by data-driven modifications, would be replicated in other samples of this size (Hoyle, 1995; pp. 174 and 86).

Based on all of these indicators and on the aforementioned indices of fit, Model 8 appears to fit well and to represent a reasonably close approximation of the population. Thus, Model 8 was retained as a final solution because it fits the data well from a statistical point of view. Further examination of Model 8 suggests that it also has substantive mean-

ing for educators who work with traumatically exposed children.

The standardized path estimates of Model 8 are presented in Figure 1. According to this model, violence exposure has a significant, positive relationship with delinquency ($\gamma = .17$) and delinquency has a significant, negative relationship with reading achievement ($\beta = -.24$). Additionally, classroom participation behavior has a significant, positive relationship with on task behavior ($\beta = .22$). The former finding suggests that violence exposure is associated with increased delinquent behaviors, which in turn are associated with lower reading achievement. The latter finding suggests that the more a student participated in class the more he/she was on task. No other significant relationships were suggested by the results of this model.

Implications for Education and Research

The validation of Model 8 suggests that children who are exposed to trauma have a high likelihood of exhibiting more mental health problems (delinquency) and academic failure (reading achievement) than their peers who have not been exposed to traumatic events. This finding has implications for education and research.

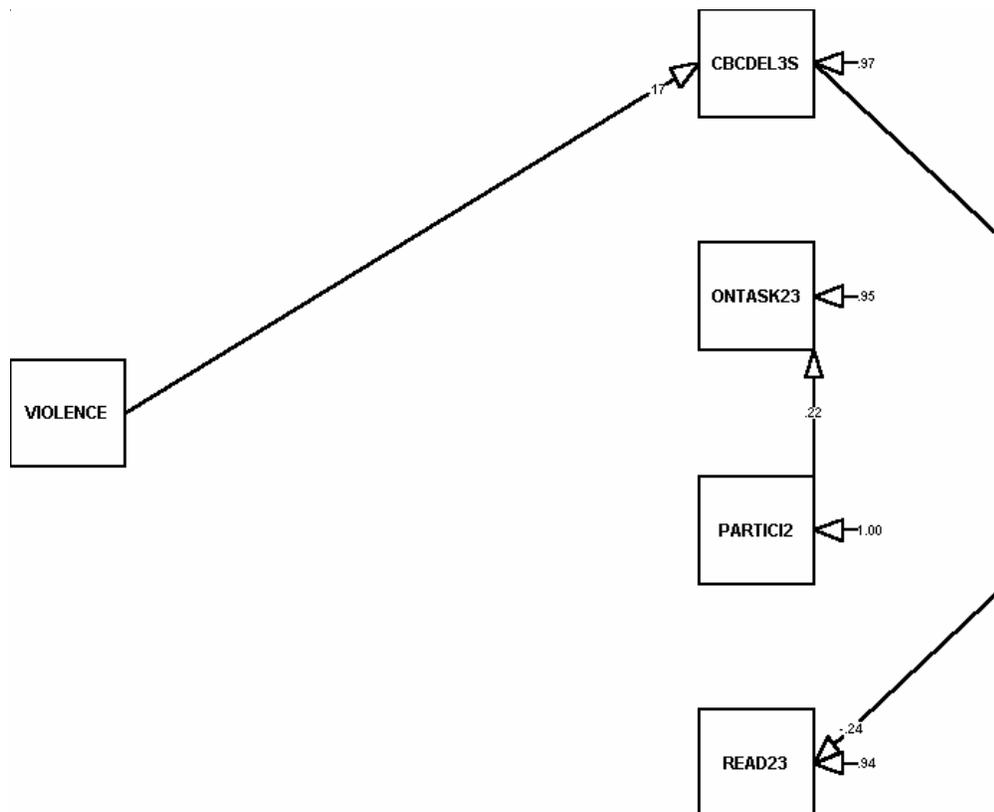


Figure 1. Model 8 represents the one model that has both statistical and substantive meaning for educators who work with children who are exposed to violence.

Note: Violence = Violence Exposure; CBCDEL3S = Delinquency; ONTASK23 = Ontask behavior; PARTICI23 = Classroom participation; and READ23 = Reading achievement.

The model suggests that children who are traumatically exposed exhibit problems with delinquency and achievement, behaviors that were observed during school hours. Consequently, this model also suggests that schools and educators can play a critical role in identifying and assisting traumatically exposed children. For example, knowing that violence exposure increases delinquent behaviors and ultimately decreases reading achievement, educators can understand and see how violence exposure might impact their students and their goals and objectives for their students, namely the goals of achievement and appropriate social behavior. With this understanding, when teachers notice a child who is exhibiting delinquent behavior and lower reading achievement, it is quite possible that this child has been exposed to a serious violent event. On the other hand, with this understanding, educators and schools may decide to periodically screen their students in order to determine if they have been traumatically exposed. Either way, understanding the impact that traumatic exposure can have on student behaviors and achievement is the first step in enabling educators and schools to assist these students.

Once educators and schools understand the impact of traumatic exposure on the academic and social goals they have for their students, they are then in the position to anticipate how they can further assist traumatically exposed children. Assistance to children who are traumatically exposed can come in various forms. One form of assistance may come through a counseling referral. While a referral is not the only way to assist traumatically exposed children, past research has shown that counseling has increased academic achievement of children who have been traumatically exposed (Dyson, 1990; Shanok, et al., 1989). Thus, a counseling referral is a good start for educators who want to help their students who have been traumatically exposed. Schools and educators can also help students cope with traumatic exposure by providing mentors who are caring adults such as a teacher or other school personnel and by providing school opportunities in which a traumatically exposed student feels responsible for someone or something else (Zimrin, 1986). All of these are forms of assistance that will help these students understand and cope with the traumatic event that they have experienced.

Additionally, given that lower achievement was a consequence of traumatic exposure, another form of assistance may come in the form of traditional academic assistance. Any strategies that have been shown to assist academic achievement, in this case reading achievement would be appropriate forms of assistance. For example, strategies such as tutoring, providing extra practice, using academic games for reinforcing sight words, teaching phonics and building appropriate vocabulary, and drawing on students' background experiences as a way of motivating them and connecting reading to real life are some of many tried-and-true strategies that research and practice have shown to be effective. Assistance can also come in the form of educators being aware that the effects of traumatic exposure can mask them-

selves as special education needs. Therefore teachers need to be careful to distinguish between special education needs and the effects of trauma since both are based on cognitive and/or behavioral problems. Otherwise, traumatically exposed children who exhibit these types of problems may be "inaccurately diagnosed, and inappropriately treated" (Bell and Jenkins, 1991, p.175). Thus, a teacher's understanding, a referral to a social worker, and a little time may be more helpful to this type of student than a referral to special needs services.

Lastly, given that delinquency is associated with lower achievement, another form of assistance may come in the form of curriculum additions and school activities and programs designed to directly address delinquent behaviors. For example, Shakoor and Chalmers (1991) suggest that schools address the cycle of violence, violence prevention, and conflict resolution through curriculum additions and school activities and programs. Also, Sroufe (1983) suggests that caring teachers helped traumatically exposed children develop appropriate ways of interacting with them. Additionally, perhaps open dialogs among school personnel and with parents and students can be attempted so that everyone is better educated about traumatic exposure and its impact on student behaviors and achievement. Thus, perhaps if all three of these recommendations could be met then delinquent behavior could be directly addressed and lessened, consequently, student achievement could be addressed and increased, and ultimately the adverse impact of traumatic exposure minimized.

Therefore, the findings of Model 8 suggest that if educators and schools want to be successful, they must realize the impact that traumatic exposure and delinquency have on student achievement. Consequently, they have a stake in understanding these relationships so that they will be prepared to take on the tasks of identifying and assisting their traumatically exposed students. But if educators and schools are to be ready for the tasks related to identifying and assisting these children, they will need the help of researchers.

Researchers can help educators by developing models that show how traumatic exposure and subsequent mental health conditions can impact achievement. To do this, future research on the effects of trauma on the education of children must be conducted in a school environment. That way, researchers can develop models of traumatic exposure that can help to inform educators about the consequences of exposure that are related to education. In turn, this type of information will help educators identify traumatically exposed children who are at-risk and strategically direct their efforts at minimizing the adverse educational effects of traumatic exposure. On the other hand, the lack of continued inquiry regarding the effects of childhood traumatic exposure may not only impede education and jeopardize the likelihood of success for both the student and the teacher, but may also increase the likelihood that the effects of traumatic exposure may become more complex.

References

- Achenbach, T. (1991). Manual for the child behavior checklist/4-18 and 1991 profile. Burlington, VT: University of Vermont Department of Psychiatry.
- Achenbach, T., and Edelbrock, C. (1986). *Manual for the Child Behavior Checklist and Revised Child Behavior Profile*. Burlington: University of Vermont.
- Airasian, P. (1989). [Review of the California Achievement Tests, Forms E and F]. In J. C. Conoley and J. J. Kramer (Eds.), *The tenth mental measurements yearbook* (pp. 126-133). Lincoln, NE: Buros Institute of Mental Measurements.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Bell, C., and Jenkins, E. (1991). Traumatic stress and children. *Journal of Health Care for the Poor and Underserved*, 2(1), 175-185.
- Brassard, M., and Gelardo, M. (1987). Psychological maltreatment: The unifying construct in child abuse and neglect, *School Psychology Review*, 16(2), 127-136.
- Browne, M., and Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen and J.S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Chicago Public Schools. (1998). *Reliability, validity and fairness of La Prueba Riverside de Realizacion en Espanol*. Department of Research, Evaluation and Planning Bureau of Student Assessment. Chicago Public Schools: Author.
- Conte, J., and Schuerman, J. (1987). The effects of sexual abuse on children: A multidimensional view. *Journal of Interpersonal Violence*, 4(2), 204-219.
- Dyson, J. (1990). The effect of family violence on children's academic performance and behavior. *Journal of the National Medical Association*, 82(1), 17-22.
- Eckenrode, J., Laird, M., and Doris, J. (1993). School performance and disciplinary problems among abused and neglected children. *Developmental Psychology*, 29(1), 53-62.
- Einbender, A., and Friedrich, W. (1989). Psychological functioning and behavior of sexually abused girls. *Journal of Consulting and Clinical Psychology*, 57(1), 155-157.
- Elliott, S., and Busse, R. (1992). [Review of the Child Behavior Checklist]. In J. C. Conoley and J. J. Kramer (Eds.), *The eleventh mental measurements yearbook* (pp. 166-169). Lincoln, NE: Buros Institute of Mental Measurements.
- Finn, J., and Cox, D. (1992). Participation and withdrawal among fourth-grade pupils. *American Educational Research Journal*, 29(1), 141-62.
- Friedrich, W., Einbender, A., and Luecke, W. (1983). Cognitive and behavioral characteristics of physically abused children. *Journal of Consulting and Clinical Psychology*, 51(2), 313-314.
- Guerra, Eron, Huesmann, Tolan and Van Acker. (1990). A cognitive/ecological approach to the prevention and mitigation of violence and aggression in inner-city youth. University of Illinois at Chicago.
- Guida, F., Ludlow, L., and Wilson, M. (1985). The mediating effect of time-on-task on the academic anxiety/achievement interaction: A structural model. *Journal of Research and Development in Education*, 19(1), 21-26.
- Hoyle, R. (Ed.). (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage Publications.
- Hoyle, R., and Panter, A. (1995). [Writing about structural equation models]. In R. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158-176). Thousand Oaks, CA: Sage Publications.
- Hu, L., and Bentler, P. (1995). [Evaluating model fit]. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76-99). Thousand Oaks, CA: Sage Publications.
- Johnson, T., and Butts, D. (1983). The relationship among college science student achievement, engaged time, and personal characteristics. *Journal of Research in Science Teaching*, 20(4), 357-66.
- Joreskog, K., and Sorbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Loranger, M., and Picard, L. (1981). *The special education student: His behavior and academic achievement*. Paper presented at the Annual Meeting of the American Psychological Association (Los Angeles, CA, August, 1981).
- McMillian, J. H. (2000). *Educational research: Fundamentals for the consumer* (p. 132). New York: Addison Wesley Longman.
- Morine-Dershimer, G., Galluzzo, G., and Fagal, F. (1980). *Participant perspectives of classroom discourse*. Part III: Rules of discourse, classroom status, *pupil participation, and achievement in reading: A chaining of relationships*. Final report. Eric Documents (ED) # 210 104.
- Pynoos, R., and Eth, S. (1984). The child as witness to homicide. *Journal of Social Issues*, 40(2), 87-108.
- Pynoos, R., Fredrick, C., Nader, K., Arroyo, W., Steinberg, A., Eth, A., Nunez, F., and Fairbanks, L. (1987). Life threat and posttraumatic stress in school-age children. *Archives of General Psychology*, 44, 1057-1063.
- Raju, N. (1992). [Review of the Iowa Tests of Basic Skills, Form J]. In J. C. Conoley and J. J. Kramer (Eds.), *The eleventh mental measurements yearbook* (pp. 423-424). Lincoln, NE: Buros Institute of Mental Measurements.

-
- Repp, A., Harman, M., Felce, D., Van Acker, R., and Karsh, K. (1989). Conducting behavioral assessments of computer collected data. *Behavioral Assessments, 11*, 249-268.
- Salzinger, S., Kaplan, S., Pelcovitz, D., Samit, C., and Krieger, R. (1984). Parent and teacher assessment of children's behavior in child maltreating families. *Journal of the American Academy of Child Psychiatry, 23*(4), 458-464.
- Shakoor, B., and Chalmers, D. (1991). Co-victimization of African-American children who witness violence: Effects on cognitive, emotional and behavioral development. *Journal of the National Medical Association, 83*(3), 233-238.
- Shanok, R., Welton, S., and Lapidus, C. (1989). Group therapy for preschool children: A transdisciplinary school-based program. *Child and Adolescent Social Work, 6*(1), 72-95.
- Sroufe, L. A. (1983). Infant caregiver attachment and patterns of adaptation in preschool: The roots of maladaptation and competence. The Minnesota symposium on child psychology, *16*, 41-83. Hillsdale, NJ: Erlbaum.
- Stallings, J., Johnson, R., and Goodman, J. (1985). *Engaged rates: Does grade level make a difference?* Paper presented at the Annual Meeting of the American Educational Research Association (69th, Chicago, IL, March 31-April 4, 1985).
- Tolan, P., Miller, L., and Thomas, P. (1988). Perceptions and experience of types of social stress and self-image among adolescents. *Journal of Youth and Adolescence, 17*, 147-163.
- Van Acker, R., Bush, J., and Grant, S. (1992). Observer accuracy under two different methods of data collection: The effect of behavior frequency and predictability. *Journal of Special Education Technology, 11*(3), 1-12.
- Wordarski, J., Kurtz, P., Gaudin, J., and Howing, P. (1990). Maltreatment and the school-age child: Major academic, socioemotional and adaptive outcomes. *Social Work, 35*(6), 506-513.
- Zimrin, H. (1986). A profile of survival. *Child Abuse and Neglect, 10*, 339-349.
-

Appendix A

Things That Happen (MACS Stress Questionnaire: Tolan, Miller, & Thomas, 1988)

DIRECTIONS: Answer the following questions by circling "YES" or "NO".

During the last year,

1. did you get poor grades on your report card? YES NO
2. have you gotten into trouble with a teacher or principal at school? YES NO
3. did you get suspended from school? YES NO
4. did your family move to a new home or apartment? YES NO
5. has your family had a new baby come into the family? YES NO
6. has anyone moved out of your home? YES NO
7. has someone else you know, other than a member of your family, gotten beaten, attacked or really hurt by others? YES NO
8. have you seen anyone beaten, shot or really hurt by someone? YES NO
9. have you seen or been around people shooting guns? YES NO
10. did you change where you go to school? YES NO
11. have you been afraid to go outside and play, or have your parents made you stay inside because of gangs or drugs in your neighborhood? YES NO
12. have you had to hide someplace because of shootings in your neighborhood? YES NO
13. did a family member die? YES NO
14. has a family member become seriously ill, injured badly, and/or had to stay at the hospital? YES NO
15. did another close relative or friend die? YES NO

Community Awareness Project: A Tool for Preparing Culturally Sensitive Teachers

Barbara A. Rucker
Cleveland State University

Abstract

Of the innumerable problems facing American schools today, one of the most pressing is the shortage of teachers equipped with the skills necessary to help diverse students achieve academic success. Lack of aggressive efforts to include strategies to help preservice teachers understand and develop culturally sensitive and relevant pedagogy will result in these children continually being excluded from full participation in the educational discourse of American schools. This article looks at a community awareness project, which was designed as an approach to enhance preservice and inservice teachers' level of cultural sensitivity toward culturally and linguistically diverse children.

While the enrollment of culturally and linguistically diverse children in the nation's schools slowly creeps upward (Hodgkinson, 1998; U.S. Census, 1999), the chasm between the cultures of the children and those of the teachers they are meeting in the classrooms correspondingly continues to widen. Accompanying this phenomenon is evidence of an academic achievement gap, seemingly impervious to change, between European American students and students of color (CGCS, 2001; Racial Gap, 2000, U. S. Census, 1999). As suggested in the literature, American schools and colleges of education today are confronted with numerous issues and challenges involving pedagogy for children from ethnically diverse cultures. If the issues and challenges are to be addressed effectively, there must be changes in the way educators view ethnically and linguistically diverse children and the children's homes and communities. Of top priority for the two institutions are the challenges of (a) reversing the trend of failure for diverse students, (b) closing the disparity gap between children from diverse populations and those possessing the characteristics of mainstream America, and (c) decreasing the cultural chasm between the children and the teachers. Meeting these challenges calls for a commitment by the schools of America and the colleges of education to aggressively develop and implement, separately and collaboratively, programs that result in a transformation of the present agenda influencing how instruction is delivered to diverse students.

In overcoming the obstacles impeding educational attainment by diverse youth, school systems face the task of finding persons capable, willing, and able to put aside personal biases in order to become more knowledgeable about the homes, communities, and worldviews of the diverse youth entrusted to their care. Teachers reflecting these attributes must also have the competencies to plan and implement culturally responsive instruction. Effective planning that incorporates the children's world results in the children acquiring the essential tools necessary to succeed in an environment, which previously has been hostile to the cultural contributions brought from their homes and communities (Gay, 2000;

Hale, 1982; Irvine and Armento, 2001; Ladson-Billings, 1995; Viadero, 2000). As the Holmes Group (1995) so compellingly wrote "Educators can no longer see or relate to the next generation through a singular set of lenses. They must respect and reach all sorts of youngsters—boys and girls, rich and poor, those of different races, ethnic backgrounds, and disabling conditions" (p. 38).

For colleges of education the challenge is to fulfill the needs of the schools, especially urban schools, by implementing strategies to help those in teacher training courses develop enhanced levels of sensitivity toward culturally and linguistically diverse students. This issue was recognized as early as the 1970s by the National Council for Accreditation of Teacher Education (NCATE) when it required that multi-cultural education courses be infused into the teacher preparation curricula of colleges of education seeking accreditation (Goodwin, 1997; Mahan, 1982). In sharing NCATE's view on this issue, Garcia and Pugh (1992) wrote, "We need a new kind of knowledge base and the ability to translate it into teacher education programs that enable students to both examine critically and experience pluralism that characterizes their world and that of their future students" (p. 218). While Villegas and Lucas (2002) support this position, they also speculate that it will take some time to bring about change.

The study presented in this article had a dual purpose. The primary focus was to explore the relationship between a community awareness project and preservice and inservice teachers' level of cultural sensitivity. The secondary purpose was to use the community awareness project to expand upon the studies of Lark (1990) and Deering and Stanutz (1995). The instrument used in the three studies was Henry's (1985; 1991) Cultural Diversity Awareness Inventory (CDAI).

Larke (1990) used the CDAI "to discover the effect of a multicultural course on the cultural sensitivity of female elementary preservice teachers" (p. 39). To expand upon the work of Larke (1990), Deering and Stanutz (1995) used the CDAI as a pretest and a posttest for the intervention of a

field experience by 16 secondary education research subjects (10 males and 6 females) placed in a culturally diverse middle school. The research questions for the study were: "How culturally sensitive are preservice teachers?" and "What effect does a pre-student teaching field experience in a multicultural setting have on their cultural sensitivity?" (p. 391). The researchers reported mixed results. They found "Significant changes in attitudes occurred in some areas, but the field experience seemed to have no impact other areas, perhaps a negative impact in a few" (p. 392). The two researchers concluded that "raising the cultural sensitivity of preservice teachers is more difficult than originally thought" (p. 392).

The community awareness study presented in this article used Henry's (1991) CDAI to expand upon the work of Deering and Stanutz (1995), who expanded upon the work of Larke (1990). This study differs from the previously mentioned studies because it (a) uses statistical exploration and not percentages to report responses and changes in responses, (b) uses a community awareness project, and (c) gives the teacher education students a voice through reflection papers and focus groups. The researcher of the study for this article anticipated more positive results than those achieved by the previous researchers. The assumption was that the employment of a community awareness project and the encouragement of their voice (collectively or as individuals) would result in the acquisition of a measurable enhanced level of cultural sensitivity for the preservice and inservice teachers.

Theoretical Framework

Along with the family, the community plays an important role in shaping the attitudes, values, and worldview of most children (Dodge and Colker (1996), Grant and Gomez (1996), Hale (1982), Ladson-Billings (1992), McCaleb (1994) and Seefeldt (2001). Pang (1994) asserts that when teachers know the school neighborhood and issues important to the community, they display more competence in developing curricula that makes a connection to the lives of the students. It must therefore be accepted that knowledge of the community leads to knowledge of the cultural frame of reference of the children residing within the boundaries of the community. Teacher education students must not only be provided with knowledge in the content areas and strategies to impart the knowledge, but the affective domain must also be prepared to be receptive to the diverse school populations that they will encounter. Viadero (2000) quotes the following from an interview with Jacqueline Jordan Irvine:

Teachers of minority children must be able to draw on their students' prior knowledge and to link it to concepts they are trying to get across. One way education schools can produce teachers with that skill, she said, is to provide them with internships that take them beyond the school walls and into the

neighborhoods, churches, temples, and mosques of their students. (p. 19)

Knowledge of community norms and values is critical if barriers leading to cultural dissonance are to be removed. Cultural incongruence should not be allowed to have the power in this century as it had in the last one. This power can be weakened with knowledge, understanding and respect. Acknowledgment and validation of a community's way of life can be achieved through sincere and aggressive programs designed to provide opportunities for exploration and participation in the community by present and future teachers.

Cultural Incongruence

Fordham (1988), Hilliard (1995) and Ladson-Billings (2001) theorize that the lack of academic achievement of diverse children is due to incongruence between the cultures of the homes and communities and that of the schools. The school's Eurocentric curriculum does very little to validate the cultural behaviors and learning styles of these children. Bennett (1999) noted, "Given a choice of fitting in at home or school, most children choose the former" (p. 236).

Foster (1992) and Heath (1983), in looking at African American children, assert that for some of these children there is a lack of congruence between their literacy patterns and the Eurocentric expectations of the schools. An example of this premise is the storytelling pattern of the children. Foster (1992) and Hilliard (1995) found that African American children have a disjointed storytelling pattern, which Eurocentric trained teachers find disconcerting. According to Hilliard (1995):

Many teachers from Eurocentric cultures have a linear storytelling style. Many African American children, on the other hand, exhibit a spiraling storytelling style, with many departures from an initial point, but with a return to make a whole. Many teachers of these children are unable to follow the children's incoherent stories. Some teachers even believe the children's stories have no order at all. Some lose patience with the children and indicate that they are doing badly. (p. 176)

Teachers who do not understand the perspectives of diverse students will have low expectations for them, and therefore will be unable to plan effective instruction to help the children master content and to achieve to their highest potential. The task of teacher education programs is clear, to begin to reverse the plight of diverse children, the teacher preparation curricula should have as a major goal the training of knowledgeable, competent and culturally sensitive teachers. Getting to know the community in which a child resides is one viable strategy to acquiring an understanding of the whole child.

Community Awareness Projects

By developing substantive family and community awareness curricula for preservice teachers, teacher preparation

programs have the capacity to play a vital role in addressing the issue of cultural incongruence. The implication therefore is that programs training teachers of today should consider the community as a resource in preparing them to recognize and respect the family and the community's contribution to the education of the diverse child. Community awareness experiences must be a required component of the curricula of teacher training programs.

Over the past couple of decades, recognition and acceptance of the need for community awareness experiences have prompted many colleges of education to incorporate more substantive community awareness activities into their curricula. Gradually, along with service learning which was already apart of some curricula, there was more emphasis on multicultural classes and community awareness activities. The many community awareness projects springing up in schools of education ranged from minimal contact to full engagement. Following are examples of projects that are most representative of those in the colleges of education. The list is in order of minimal to full immersion.

1. Persons from the community would visit the college classroom to explain the characteristics of their community and the associated cultural behaviors (Colville-Hall, MacDonald and Smolen 1995).
2. In an hour-long immersion in a diverse community or environment, preservice teachers were asked to visit self-selected activities (religious services, welfare lines, gay bars, etc.) and then write about their experiences in journals (Weist, 1998).
3. In a full-day activity teacher education students visit a school in a diverse community and are assigned a student to shadow. At the end of the day the teacher education student interviews the shadowed student (Tamura, et al., 1996).
4. Teacher education students enrolled in a summer course were involved in self-reflective dialogues, community field trips, service learning, and curriculum development (Pang, 1994).
5. In a special master's program preservice teachers had to (a) spend 10 hours a week the first year involved in service learning in an urban environment, (b) daily spend the summer with a family in the school community, and (c) observe, assist and teach in the diverse school community during the following (Ladson-Billings, 2001).
6. Full engagement is reflected in a program that covered a period of four months. The teacher education students had the opportunity to live with residents in Native American and Hispanic communities. However, it must be noted that those assigned to urban African American schools did not have to reside in the neighborhood (Mahan, 1981; 1982).

Since the body of literature reflects that the community plays such a vital role in the life of a child, the researcher of this study chose to pursue the community awareness topic

and hypothesized that there is a relationship between participation in a community awareness project and preservice and inservice teacher education students' level of cultural sensitivity toward culturally and linguistically diverse children. The general research question to be explored was, Is there a relationship between participation in a community awareness project and preservice and inservice teacher education students' level of cultural sensitivity?

The specific research questions were:

1. Can an experience of one or two visits to a diverse community be beneficial in helping teacher education students develop a level of comfort in a diverse community?
2. Can participation in a community awareness project be beneficial in helping preservice and inservice teacher education students develop multiple perspectives?
3. Can a community awareness project help preservice and inservice teacher education students develop culturally relevant lesson plans for culturally and linguistically diverse students?

Methodology

The community awareness research project was conducted at a large university in an urban Midwestern setting. The subjects were preservice and inservice teacher education students enrolled in a diversity class. Of the 35 students assigned to the class, data were gathered from the 20 students who fulfilled the criteria of taking two pretests and one posttest.

Research Design

The research design was both qualitative and quantitative. The mixed research design addressed the general question "Is there a relationship between participation in a community awareness project and preservice and inservice teacher education students' level of cultural sensitivity toward culturally and linguistically diverse children?" The qualitative component of the design was used to clarify and enhance the quantitative data. Data were triangulated to enhance their value and to corroborate or call into question their trustworthiness.

Quantitative component. The quantitative component employed the Cultural Diversity Awareness Inventory (CDAI) by Henry (1991), a self-administered questionnaire designed to examine educators' attitudes, beliefs, and behaviors toward culturally diverse children and their families. The instrument consists of 28 items on a 5-point Likert type scale with a range of strongly agree to strongly disagree and neutral. According to the author, "Internal consistency validity for the CDAI showed that 92% (N=26) of the 28 total statements correlated at a level above 0.52. Results of Cronbach's test for internal consistency reliability evidence an overall alpha coefficient of 0.90, and 26 of the 28 statements correlated at 0.52 or better with the whole of the test. Test-retest reliability for the entire test was .66"

(p. 5). The CDAI was used to gauge the cultural sensitivity of the preservice and inservice teacher education students before and after participation in the community awareness project.

The research design framework utilized repeated measures of pretest 1, pretest 2, and the posttest. The study used SPSS 10.0 GLM Repeated Measures Polynomial Tests of Within-Subjects Contrast. This specific test was used because of the comprehensive nature of the GLM and because the polynomial test analyzed for curvilinear as well as linear relationships between the scores of the three tests. In accordance with standard statistical practice, so that the p-value would be less than or equal to 0.05 for the instrument as a whole, the Bonferroni Correction for Multiple Tests (SPSS, 1996, p.193) was applied to the GLM analysis ($0.05/28=0.00178$), therefore $p < 0.002$ indicated statistical significance for individual questions. Consequently each question could be treated as an individual test within the instrument. This was necessary in order to be able to use questions selectively for the qualitative analysis. Item 22 was reverse scored because it is stated as a negative statement. The p-value was set at 0.05. Since the study was looking for positive changes, the directional alternative hypothesis was one-tailed.

Qualitative component. The qualitative component included data generated by the CDAI, focus groups, a class set of lesson plans, and selected reflection papers. The lesson plans, along with the focus groups served to provide methodological triangulation to the qualitative data. The focus group statements were derived from CDAI items showing the greatest change between pretest 2 and the posttest. SPSS 10.0 Paired Sample Test was used to determine changes in gain scores between pretest 2 and the posttest. The lesson plans were to be structured to include culturally relevant and culturally authentic aspects of a diverse community. The lesson plans, based on Stockard's (1995) modified version of the Madeline Hunter Model, were examined for cultural sensitivity through the use of a rubric developed collaboratively by the instructor and the teacher education students. Selected reflection papers were analyzed to gauge the impact of the project upon the teacher education students' attitudes.

Sample

The group of 20 was comprised of 17 females and 3 males; of that number 11 identified themselves as European Americans, 4 as Hispanics, 3 as African Americans, 1 as Asian American, and 1 as Other. They were selected because they took all three CDAI inventories.

Community Awareness Project

A large, urban city in the Midwest was selected for the school community awareness project. Six school communities reflecting four traditional elementary schools, an elementary magnet school for the gifted, and a combined middle/senior high school were selected by the instructor/

researcher for the project. Based upon grade level, content interest or specialty, preservice and inservice teacher education students were assigned by the instructor/researcher to racially/ethnically diverse groups of five or six members. The groups were instructed to visit an assigned school where they were to collect data relevant to the staff, students, and interior and exterior physical plant. In the community they were required to note the resources and general appearance of the community within a radius of one mile of the school. This information was to be used to develop culturally relevant and authentic lesson plans.

Implementation/Timeline

1. The CDAI pretest 1 was administered for baseline data for the course at the first class meeting.
2. The CDAI pretest 2 was administered for pre-intervention data during the fifth week of class.
3. The Community Awareness Project was introduced and explained by the instructor/researcher during the fifth week.
 - (a) The instructor assigned the members of the class to one of six groups that were comprised of five to six class members.
 - (b) The instructor assigned community school sites.
 - (c) To fulfill the requirements spelled out in the guidelines, the teacher education students were given a time frame of six weeks to visit their assigned school and the surrounding radius of one mile within the school community. This task was to be completed outside of class time.
4. During the sixth week, a rubric to ascertain the presence of culturally sensitive teaching strategies was developed collaboratively by the instructor/researcher and the preservice and inservice teacher education students.
5. The CDAI was administered as the posttest during the 11th week.
6. The focus groups met the 12th week to bring clarity to CDAI items reflecting the most significant changes between pretest 2 and the posttest.
7. Lesson plans were collected the 13th week.
8. Reflection papers were collected during the semester as a course requirement.

Data Collection and Analysis

Quantitative

The GLM analyzed the CDAI results of pretest 1 and pretest 2 to determine if there had been a shift in the baseline data before the community awareness project due to the introductory course work. Changes between CDAI pretest 2 and the posttest results were evaluated for the relationships between a community awareness project and the cultural

level of sensitivity of preservice and inservice teacher education students.

Qualitative

Qualitative data were analyzed in the following three ways:

1. Focus groups—Analysis of statements generated by the CDAI items reflecting the most significant changes between pretest 2 and the posttest were used for the focus group sessions.
2. Lesson plans—Lesson plans were collected and analyzed for evidence of the incorporation of culturally sensitive elements.
3. Reflection papers—Statements in reflection papers were examined for signs that students had had their preconceptions about diverse populations challenged.

Results

The scores for pretest 1, pretest 2 and the posttest were analyzed using SPSS 10.0 GLM Repeated Measures Polynomial Tests Within-Subjects Contrast. At $\alpha=0.05$ and with the application of the Bonferroni Correction for Multiple Tests ($0.05/28=0.00178$) p was determined statistically significant at < 0.002 . None of the statements showed statistically significant changes. The hypothesis, “There is a relationship between a community awareness project and preservice and inservice teacher education students’ level of cultural sensitivity toward culturally and linguistically diverse children.” was rejected.

The qualitative data, at times, supported the statistical analysis and at times showed great contradiction. For example, the focus groups showed that when allowed to explore the statements with peers, in contrast to individually responding, the results were quite different from the statistical analysis. Focus group statement #11, as shown in Table 1, is an example of how much difference occasionally occurred between the qualitative and quantitative data. Statement #11 addresses the translation of standardized tests and has fairly comparable numbers between pretest 1 (70% strongly disagree/disagree), pretest 2 (75% strongly disagree/disagree) and the posttest (70% strongly disagree/disagree). However, in the focus groups 95% of the participants felt that the translation of a standardized achievement test or intelligence test into the child’s dominant language does not give the child an added advantage or impact peer comparison. This difference may be explained by strong opinions expressed by experienced inservice bilingual teachers swaying non-bilingual and/or inexperienced teacher education students, but since participants were not identified by name, it was difficult for the researcher to state this conclusively. However, the opening remarks for the statement are strongly expressed and appear to have been made by experienced teachers, and a number of the experienced teachers in the diversity class were bilingual teachers.

Table 1
Response Percentages for Focus Group Statements

Statement # 11:

I believe translating a standardized achievement or intelligence test to the child’s dominant language gives the child an added advantage and does not allow for peer comparison.

Tests	Strongly Agree/ Agree N (%)	Strongly Disagree/ Disagree N (%)	Neutral N (%)	Blank Response N (%)
Pretest # 1	3(15)	14(70)	3(15)	
Pretest # 2	4(20)	15(75)	1(5)	
Posttest	1(5)	14(70)	5(25)	
Focus groups	19(95)	1(5)		

General Research Question

Is there a relationship between participation in a community awareness project and preservice and inservice teacher education students’ level of cultural sensitivity toward culturally and linguistically diverse children?

There was no statistically significant change on the level of cultural sensitivity toward culturally and linguistically diverse children for the preservice and inservice teacher education students participating in the community awareness project.

Specific Research Question # 1

Can an experience of one to two visits to a diverse community be beneficial in helping preservice and inservice teacher education students develop a level of comfort in a diverse community?

As delineated above, the empirical data shows that there was no significant statistical change in the level of comfort in the diverse community for the teacher education students. Between pretest 2 and the posttest the teacher education students were involved in activities planned and structured to introduce them to world of diverse children. There were times for visits to the school and the community, and also class time for reflections and sharing. Results showing the comparison between pretest 2 and the posttest of the CDAI revealed that participation in the community awareness project did not significantly affect the preconceived notions the teachers held about children and families in diverse communities; there was very little movement toward increased cultural sensitivity.

Question #2

Can participation in a community awareness project be beneficial in helping preservice and inservice teachers develop multiple perspectives?

Reflection papers gave a glimmer of hope that some were becoming more aware and sensitized to the cultural diversity within the city schools and communities. In order to take the students out of their “comfort zone”, the instructor/researcher assigned students to communities, which were considered different from those in which the students were

raised and/or were presently residing. African Americans who primarily resided on the east side of the city were assigned to far west, European American, or Hispanic communities. The European American students who were mainly from the far western and eastern suburbs were assigned to inner city and near west communities. The students were encouraged to express their feelings about the community awareness project through reflection papers and small group and whole class discussions. The reflections in the examples below are indications that some of the teacher education students were willing to consider selected aspects of the views of others different from themselves.

In a reflection paper, one white female inservice teacher wrote:

I found it ironic that 99% of the student population is African American and 62% of the professional staff is white. An interesting fact is that 95% of the paraprofessional and cleaning staff are Black. This seems to be an example of the inequity that African Americans find in the job market.

Other teacher education students wrote that they were surprised by the large number of churches in the African American and Hispanic communities and the minimal number of banks and large chain supermarkets within the diverse communities.

Another group of primarily European American students took delight in sharing in class and in the reflection papers the activist role which they had assumed while visiting their community. This group noted that within a one mile radius of the elementary school for which the group was assigned, they had found only one supermarket. Assuming that the neighborhood surrounding the school would reflect similar demographics of the school (50% European American, 40% Hispanic Americans, and 10% Native Americans, Asian Americans, and multi-racial children), the group members reported that they were confused when they discovered that the supermarket did not stock products that reflected the diversity of the school and probably the surrounding neighborhood. In a phone interview with the assistant manager, they were told that “the store doesn’t have a variety of ethnic foods simply because it is located in an older mixed neighborhood.” The group wrote that they surmised that the term “older mixed neighborhood” referred to older European American population residing in the community. After more investigation of supermarkets of the same chain in adjacent communities, the group members discovered that those supermarkets stocked the Goya brand for the Hispanic American population and the Glory brand for the African American population. When they called the assistant manager and shared their findings with him, he was at a loss to explain the difference between what his market stocked and that of the others. The group was confident that they had uncovered a form of subtle discrimination against people of color.

It must also be noted that for a least one European American female student the project only served to reaffirm her stereotypical attitudes about the cultural group of the

assigned community. Her comments were generally negative about the people she observed in the community, the school, the community resources, the project, and the diversity class. In her reflection paper she wrote that by requiring the class to visit a diverse community, the instructor/researcher had placed them in a dangerous position.

Specific Research Question #3

Can a community awareness project help preservice and inservice teachers develop culturally relevant lesson plans for culturally and linguistically diverse children?

The teacher education students made every attempt to follow the lesson plan model, but with a few exceptions, the majority of the lesson plans did not reflect culturally relevant or authentic resources that were unique or specific to the community of the diverse populations which the teacher education students were to address. Most of the teacher education students elected to use ideas and strategies, which they found to be comfortable, familiar, and readily available, e.g. field trips to restaurants popular with mainstream society, fast food establishments, and large supermarkets. They ignored the neighborhood-clothing store, the small ethnic restaurants and bakeries, as well as small grocery stores that stocked the ethnic foods and spices preferred by members of the community. Those who did choose to seek input from parents or community volunteers did not reflect that they requested substantive input, support, or contributions. The researcher could only speculate that the students might have felt that since the lessons would not be implemented in a k-12 educational setting, they were not worth putting forth a full effort.

While many of the lesson plans used food for content, the student who submitted “Making Habichuelas for a Monthly Multicultural Luncheon” truly reflected movement toward an enhanced level of cultural sensitivity. In order to insure authenticity, she researched a culture different from her own before attempting to develop the lesson plan. The student, an inservice African American female, planned a lesson for children in a bilingual school with a large Hispanic American enrollment, and of that enrollment most of the children were of Puerto Rican descent.

The student reported that after finding the recipe in a Hispanic cookbook, she decided to consult two colleagues, a Puerto Rican and a Mexican American, to further confirm the authenticity. She further shared that the Mexican American stated that since the dish was Puerto Rican, she could not vouch for the authenticity. The Puerto Rican colleague said that the ingredients were authentic, and even though he does not cook, he recognized words that he had heard all his life. The fact that the recipe was particularly Puerto Rican showed that the student was aware that the Hispanic culture is not just one culture, but many cultures. Therefore when shopping for the ingredients she was able to list the names of two neighborhood stores that stocked the items needed for the dish. Validating the specific culture of the school’s community is the heart of culturally relevant teaching.

Conclusion

The community awareness project covered 10 weeks. The focus of the study was a project designed to enhance preservice and inservice teacher education students' level of cultural sensitivity toward culturally and linguistically diverse children. The CDAI assessed the level of cultural sensitivity, the lesson plans served to gauge how the level of cultural sensitivity affected instructional strategies, and the reflection papers provided a forum for individual and group voice reaction toward a project designed to increase cultural sensitivity. Although the qualitative data showed glimmers of the emergence of an enhanced level of cultural sensitivity, it overall supported the lack of statistical significance in the quantitative design. There was not enough significance expressed in the qualitative to contradict the quantitative significance.

While not a focus of this article, it is imperative that attention also be given to the level of cultural sensitivity possessed by those training teacher education students. In doing so, the issues behind the colleges of education's role in preparing culturally sensitive teachers are further highlighted. This concern has been documented by several researchers. Davis (1995) and Reed (1998) found that the attitudes of the faculty toward diverse children strongly impact the frame of reference of teacher education students. While conducting a study that looked at the worldviews of undergraduates, Davis (1995) discovered that the faculty of the teacher education students held "a wide range of views on the education of minority children including genetic deficiency, cultural deficiency, and home/difference" (p. 554). It is her position that these types of views affect instructional methods for preservice and inservice teachers. In expressing concern about the quality of culturally sensitive curriculum and instruction in higher education, Pang, Anderson, and Martuza (1997) wrote, "Although it is possible to create programs that 'sensitize' professors to the need for cultural literacy, the basic structure of universities remains largely dominated by Western thought and an entrenched system of control" (p. 69).

Teaching teachers to be more culturally sensitive is a difficult task. The deeply ingrained mainstream values brought into the college classroom presents a formidable foe when asking preservice and inservice teachers to look at the impact of those values upon children from diverse cultures (Bennett, 1999; Sleeter, 1995). The results of the study presented in this article support the findings of other researchers, e.g. Larke (1990) and Deering and Stanutz (1995) who also found that teaching teacher education students to be more culturally sensitive is a difficult task. However, the glimmers of positive change reflected in the qualitative data were encouraging and affirming for the need to have ongoing attempts by all involved with teacher education. Teacher education students are on the threshold of being receptive to change; the task now is for professors in higher education to examine not only their personal attitudes to-

ward ethnically diverse students, but also the pedagogical principles that they have so dearly espoused in the past. The method used by a professor to provide guidance and deliver instruction has a profound affect upon what preservice teachers will model in the future. If underachieving children from diverse backgrounds are to experience the same educational success as their mainstream counterparts, colleges of education will have to alter the prevailing ethos and consider a new paradigm for teacher training (Rucker, 2001).

Recommendations

While the statistical data do not reflect a significant difference, the community awareness project study is not without merit. Some of the preservice and inservice teacher education students felt that the opportunity to visit a culturally and linguistically diverse school and community added to their understanding of the different perspectives children bring to the classroom. The lack of a statistically significant change indicates the need for more and better programs to be introduced to the mainstream teachers who will one day be placed in classrooms with children from diverse populations. Specific recommendations resulting from this study are:

1. All preservice and inservice teacher education students should be administered an inventory to measure their level of cultural sensitivity toward diverse populations. This could be done at admittance, at the same time that their math and English abilities are being tested. The results would then be used to structure programs to help them attain the needed level of sensitivity to be successful with culturally and linguistically diverse children.
2. Recruitment efforts must be expanded to aggressively target teachers of color and mainstream teachers interested in working in diverse communities.
3. "Grow your own" collaborative programs should be expanded to more local school districts.
4. Multicultural education concepts must be a part of the content of all education classes. Teacher education students must be taught how to integrate multicultural education concepts into the content areas.
5. Pre-service teacher education students must be introduced to diverse populations early in their teacher education program. The researcher recommends that service learning not only be a component of the first education class in which a teacher education student enrolls, but also in remaining classes until practicum or student teaching.
6. Colleges of education must give consideration to collaborating with local urban school systems to develop professional development schools and/or hold teacher education classes on-site at the schools.
7. Professors of teacher education courses must have professional development with emphasis on (a) culturally

sensitive pedagogy, (b.) culturally relevant pedagogy, and (c) an understanding of the uniqueness of cultural and linguistic plurality.

References

- Bennett, C. (1999). *Comprehensive multicultural education: Theory into practice*. (4th ed.). Boston: Allyn and Bacon.
- Colville-Hall, S., MacDonald, S., and Smolen, L. (1995). Preparing preservice teachers for diversity in learners. *Journal of Teacher Education*, 46(4), 295-303.
- Council of the Great City Schools. (2001, April 27). *New studies detail urban achievement* [Press release]. Washington, DC: Author. Retrieved April 27, 2001, from http://www.cgcs.org/services/media/4_27_01.htm
- Davis, K. A. (1995). Multicultural classrooms and cultural communities of teachers. *Teaching and Teacher Education*, 11(6), 553-563.
- Deering, T. E., and Stanutz, A. (1995). Preservice field experience as a multicultural component of a teacher education program. *Journal of Teacher Education*, 46(5), 390-394.
- Dodge, K. T., and Colker, L. J. (1996). *The creative curriculum for early childhood*. (3rd ed.). Washington, DC: Testing Strategies, Inc.
- Fordham, S. (1988). Racelessness as a factor in Black students' school success: Pragmatic strategy or pyrrhic victory? *Harvard Educational Review*, 58(1), 54-84.
- Foster, M. (1992). Sociolinguistics and the African American community: Implications for literacy. *Theory Into Practice*, 31(4) 303-310.
- Garcia, J., and Pugh, S. L. (1992). Multicultural education in teacher preparation programs: A political or an educational concept? *Phi Delta Kappan*, 74(3) 214-219.
- Gay, G. (2002). Preparing for culturally responsive teaching. *Journal of Teacher Education*, 53(2), 106-116.
- Grant, C. A., and Gomez, M. L. (1996). (Eds.). *Making schooling multicultural: Campus and classroom*. Englewood Cliffs, NJ: Prentice Hall, Inc.
- Hale, J. E. (1982). *Black children: Their roots, culture, and learning styles*. Provo, UT: Brigham Young University Press.
- Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. New York: Cambridge University Press.
- Henry, G. B. (1985). *Cultural diversity awareness inventory*. Hampton, VA: Hampton University Mainstreaming Outreach Project. (Available from Educational Testing Service, Rosedale Road, 30-B, Princeton, NJ 08541-001, TC020014)
- Henry, G. B. (1991). *Cultural diversity awareness inventory*. Hampton, VA: Hampton University Mainstreaming Outreach Project. (Available from Educational Testing Service, Rosedale Road, 30-B, Princeton, NJ 08541-001, TC020014)
- Hilliard, A. (1995). *The maroon within us: Selected essays on African American community socialization*. Baltimore: Black Classic Press
- Hodgkinson, H. (1998). The demographics of diversity. *Principal*, 78(1), 26-28, 30 -32.
- Holmes Group (1995). *A report of the Holmes Group: Tomorrow's schools of education*. East Lansing, MI: The Homes Group.
- Irvine, J. J., and Armento, B. J. (2001). *Culturally responsive teaching: Lesson planning for elementary and middle grades*. Boston: McGraw Hill.
- Ladson-Billings, G. (1992). Reading between the lines and beyond the pages: A culturally relevant approach to literacy reading. *Theory Into Practice*, 31(4), 312-320.
- Ladson-Billings, G. (1995). But that's just good teaching! The case for culturally relevant pedagogy. *Theory Into Practice*, 34(3), 159-165.
- Ladson-Billings, G. (2001). *Crossing over to Canaan: The journey of new teachers in diverse classrooms*. San Francisco: Jossey-Bass, Inc.
- Larke, P. J. (1990) Cultural diversity awareness inventory: Assessing the sensitivity of preservice teachers. *Action in Teacher Education*, 12(3), 23-30.
- Mahan, J. M. (1981, April). *Native Americans as Teacher Trainers: Anatomy and Outcomes of a Cultural Immersion Project*. Paper presented at the Annual Meeting of the American Educational Research Association, Los Angeles, CA.
- Mahan, J. M. (1982, February). *Community Involvement Components in Culturally-Oriented Teacher Preparation*. Paper presented at the Annual Meeting of the Association of Teacher Educators, Phoenix, AZ.
- McCaleb, S. P. (1994). *Building communities of learners: A collaboration among teachers, students, families, and community*. New York: St. Martin Press.
- Pang, V. O. (1994). Why do we need this class? Multicultural education for teachers. *Phi Delta Kappan*, 76(4), 289-292.
- Pang, V. O., Anderson, M. G., and Martuza, V. (1997). *Removing the mask of academia: Institutions collaborating the struggle for equity*. In King, J. E., Hollins, E. R. and Hayman, W. C. (Eds.) *Preparing Teachers for Cultural Diversity* (pp. 53-70). New York: Teachers College Press.
- Racial gap widens on test scores: Disparity greatest among children of America's best-educated parents. (2000, August 25) *The Plain Dealer*, pp. A1, A10.
- Reed, D. F. (1998) Speaking from experience: Anglo-American teachers in African American schools. *The Clearing House*, 71(4), 224-230.

-
- Rucker, B. A. (2001). The relationship between a community awareness project and the level of cultural sensitivity of preservice and inservice teacher education students (Doctoral dissertation, The University of Akron, 2001). *Dissertation Abstracts International*, 62, 10.
- Seefeldt, C. (2001). *Social studies for the primary child* (6th ed.). Upper Saddle River, NJ: Prentice Hall, Inc.
- Sleeter, C. E. (1995). White preservice teachers and multicultural education coursework. In J. M. Larkin and C. E. Sleeter (Eds.), *Developing multicultural teacher curricula*. New York: State University of New York Press.
- SPSS: *SPSS base 7.0 for windows. User's guide*. (1996). Chicago: SPSS Inc.
- Stockard, J. W., Jr. (1995). *Activities for elementary school social studies*. Prospect Heights, IL: Waveland Press.
- Tamura, L., Nelson, P., and Ford-Stevenson, T. (1996, February). *Preparing Teachers to Recognize Multiple Perspectives*. Paper presented at the Annual Meeting of the American Association of Colleges for Teacher Education, Chicago, IL.
- Weist, L. R. (1998). Using immersion experiences to shake up preservice teachers' views about cultural differences. *Journal of Teacher Education*, 49 (5), 358-365.
- U. S. Census Bureau, Statistical Abstract of the United States: 1999. (119th edition). Washington, DC.
- Viadero, D. (2000, March 22). Lags in minority achievement defy traditional explanations. *Education Week*, pp. 1, 11.
- Villegas, A. M., and Lucas, T. (2002) Preparing culturally responsive teachers: Rethinking the curriculum. *Journal of Teacher Education*, 53(1), 20-31.
-

Call for Editors

Mid-Western Educational Researcher

Journal of the Mid-Western Educational Research Association

Proposals are currently being sought for the Editorship of the *Mid-Western Educational Researcher*. The *Researcher* is the quarterly publication of the Mid-Western Educational Research Association. The journal serves the dual function of providing MWERA members with timely information about the organization and of providing a vehicle for dissemination of scholarly work in education or education related fields. This dual mission reflects the growth and change of the organization itself in recent years.

The appointment of the next editor or editorial team will be from January 2006, through October, 2009. Proposals are sought from individuals and teams interested in assuming responsibility for the operation and direction of the *Researcher* for a three-year period. The format for proposals is open, but each proposal should include at least the following:

- 1) Name, institutional affiliation, address, telephone and FAX numbers, and e-mail address of each member of the proposed editorial team;
- 2) A vision statement indicating the editorial team's intended goals for the journal, and an explanation of how this vision reflects the membership, perspectives, and direction of MWERA;
- 3) A proposed plan for promoting this vision; and
- 4) An explanation of the expertise and qualifications of the editorial team which are likely to encourage the continued improvement and development of the *Researcher*.

Proposals should be submitted no later than October 1, 2004, to the President of MWERA, Dr. A. William Place, at:

A. William Place, Ph.D.
Director of Doctoral Studies
University of Dayton
300 College Park,
Chaminade Hall, Room 207
Dayton, OH 45469-0526

Questions may be directed to A. William Place, Ph.D., Director of Doctoral Studies, at (937) 229-2640, FAX (937) 229-4824, or e-mail: will.place@notes.udayton.edu

The Mid-Western Educational Research Association's

Annual Meeting

October 13–16, 2004

The Westin Great Southern
Columbus, Ohio

Theory and Practice: Two Sides of the Same Coin

The 2004 Annual Meeting of the Mid-Western Educational Research Association will be held in Columbus with an exciting program of invited speakers, focused workshops, and peer-reviewed papers presented in a variety of session formats. The 2004 program will feature dynamic speakers of interest to both researchers and practitioners. Teachers, administrators, and other school personnel are especially invited to come and share their school-based research and experiences at the 2004 MWERA conference. Educational researchers across North America return to MWERA to renew acquaintances, make new contacts, and engage in exciting conversation in a collegial atmosphere. Come and be a part of MWERA!

Look for us on the World Wide Web at our new location!

<http://www.mwera.org>

Featured Speaker

Thursday Keynote Address



Dr. J. Q. Adams

J. Q. Adams is an accomplished educator, author and presenter whose in-depth research into multicultural education, preservice/in-service teacher education, and social interaction in the culturally diverse workplace has earned him a reputation as one of the premier figures in intercultural and multicultural issues in education. His popular PBS Telecourse “Dealing with Diversity” is taught at over 150 colleges and universities nationwide. Adams is currently a professor of Educational and Interdisciplinary Studies at Western Illinois University.

Adams has also been active in his work as a consultant for numerous school districts, businesses like the U.S. Postal Services, R. R. Donnelly Publishers, Burlington Northern and Santa Fe Railroads and various other organizations like the National Education Association, Illinois Department of Corrections, and the United States Navy.

Adams graduated with a Ph.D. in Educational Psychology from the University of Illinois in 1989. Over the past two decades he has established himself as a notable presenter and author in the area of multicultural education. One of the main tenets of Adams’ research that he brings to his work is, “How do we develop and prepare ourselves to live in the 21st century, in a diverse and ever more closely intermingled global world?” In continuing his excellence as a researcher and educator, Adams brings his dedication to education into the classroom. As a testament to this, he is a five-time recipient of the Western Illinois University “Faculty Excellence Award” and in 2002 he received the Distinguished Alumni Award from the University of Illinois’ College of Education.

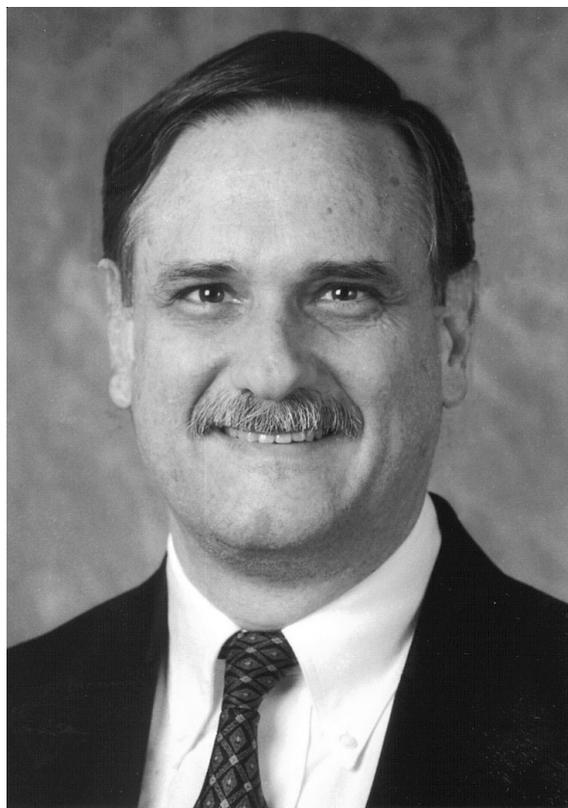
Join us for a **Fireside Chat** with **Dr. J. Q. Adams**
Wednesday evening in a casual atmosphere.
Refreshments will be provided.

Featured Speaker

**Friday Keynote/Luncheon Address:
Scientifically-Based Research and Research-Based Practice**

Dr. Robert Slavin

Robert Slavin is currently Co-Director of the Center for Research on the Education of Students Placed at Risk at Johns Hopkins University and Chairman of the Success for All Foundation. He received his B. A. in Psychology from Reed College in 1972, and his Ph.D. in Social Relations in 1975 from Johns Hopkins University. Dr. Slavin has authored or co-authored more than 200 articles and 18 books, including *Educational Psychology: Theory into Practice* (Allyn & Bacon, 1986, 1988, 1991, 1994, 1997, 2000), *Effective Programs for Students at Risk* (Allyn & Bacon, 1989), *Cooperative Learning: Theory, Research, and Practice* (Allyn & Bacon, 1990, 1995), *Preventing Early School Failure* (Allyn & Bacon, 1994), *Every Child, Every School: Success for All* (Corwin, 1996), *Show Me the Evidence: Proven and Promising Programs for America's Schools* (Corwin, 1998), and *Effective Programs for Latino Students* (Erlbaum, 2000). He received the American Educational Research Association's Raymond B. Cattell Early Career Award for Programmatic Research in 1986, the Palmer O. Johnson award for the best article in an AERA journal in 1988, the Charles A. Dana award in 1994, the James Bryant Conant Award from the Education Commission of the States in 1998, the Outstanding Leadership in Education Award from the Horace Mann League in 1999, and the Distinguished Services Award from the Council of Chief State School Officers in 2000.



Special Conference Highlights

- Division Meetings with Invited Speakers.
- New Member Welcome—A time to meet the MWERA Board of Directors and learn about the organization.
- Division L—Educational Policy and Politics—Look forward to sessions from our newest MWERA division!

Call for Papers

Other Voices in Educational Research

In December of 1987, Paulo Freire, considered by many people to be “the most significant educator in the world in the last half of the twentieth century” (Kohl, 1997), and Myles Horton, the founder of Tennessee’s Highlander School, sat down together for the purpose of “speaking a book” (Horton and Freire, 1990, p. viii). Recognizing the difficulty many students experience when first encountering texts concerning critical theory and teaching for empowerment, especially Freire’s *Pedagogy of the Oppressed* (1990), Freire and Horton set out to make their ideas more accessible by inviting the world to sit in on their conversation. The resulting text, *We Make the Road by Walking*, represents the potential for powerful and empowering classroom dialog.

Along with a colleague, Annie taught a course titled Emancipatory Pedagogies for the first time in the summer of 2002. The reading list consisted of several chapters from *The Paulo Freire Reader* (Freire and Macedo, 2001), a number of articles from *Breaking Free: The Transformative Power of Critical Pedagogy* (Leistyna, Woodrum, and Sherblom, 1999), as well as the aforementioned *We Make the Road by Walking*. They employed numerous strategies to make the texts accessible and to facilitate dialog, among them a unique team-teaching approach that eliminated the power dynamic found in most classrooms: The “class” consisted of two professors, three students, and three A’s assigned on the first afternoon of class. The teaching methodology employed in the course modeled critical theory; both students and faculty were engaged in reading, writing, reflecting, and reporting on their own lives in the classroom. Instead of a final exam or culminating paper, they decided to “talk” an article. Then began the process of shopping the resulting manuscript out to prospective publishers. Although the manuscript’s format represented a practice of empowerment, at every turn, they were asked to revise the manuscript to follow a more traditional format. This special issue of the *Mid-Western Educational Researcher* is meant to offer an outlet to authors who find themselves in similar circumstances.

This special issue of the journal will present a variety of voices and perspectives that might not otherwise be heard. The editors call for papers that address alternative responses to educational scholarship. Manuscripts are due to the editors by July 1, 2004.

Anne D’Antonio Stinson, University of Wisconsin-Whitewater

Glenda Moss, Indiana University Purdue University Fort Wayne

Anne D’Antonio Stinson, EdD
Associate Professor
Curriculum and Instruction Department
University of Wisconsin-Whitewater
800 Main Street
Whitewater, WI 53190
stinson@mail.uww.edu

Creating the Rosetta Stone: Deciphering the Language of Accountability to Improve Student Performance

Donna Snodgrass
Cleveland Municipal School District

Abstract

The objectives of this paper are three-fold. First, a model is proposed for unifying massive amounts of conceptual and numerical information flowing from the measures of the accountability movement in Ohio and the materials that are publicly available to educators. Second, this model is translated into useable forms of information that help teachers drive instructional practices in their classrooms. Third, the efficacy of this model to improve state-mandated proficiency scores at the district level is discussed. Ultimately the authors attempt to provide the readers with a rudimentary model of an educational Rosetta Stone to help educators decipher the contents of the Ohio proficiency test at a deep and complex enough level to identify the basic and fundamental instructional needs of students.

Introduction

In 1799, soldiers in Napoleon's army stumbled upon a block of stone as they were doing engineering work at Fort Julien, near Rosetta, approximately 35 miles northeast of Alexandria. The inscriptions on the stone dated to 196 BC to the reign of Ptolemy V. The decree was written in two languages, Egyptian and Greek, with the Egyptian being written twice, once in hieroglyphics and once in an Egyptian cursive script which evolved from hieroglyphics. The assumption was that all three languages had the same message. Though both Egyptian writings were indecipherable at the time, it followed that, working from the Rosetta Stone, scholars of Greek could use it to translate Egyptian hieroglyphics. From the discovery, however, it still took over 20 years before Jean Francois Champollion was able to crack the code and decipher the messages on Egyptian antiquities.

In some ways, I suppose that I have a sense of the task Champollion undertook when he set out to uncover the meanings of the hieroglyphs on the Rosetta Stone. In educational circles, high-stakes tests have provided teachers and other stakeholders with their own set of "decrees" written in a language that is neither spoken nor read, and certainly little understood, by the vast majority of professionals affected by them. As an educational community we also are fortunate to find ourselves a little ahead of Champollion; it's only taken us a little over ten years to wade through the morass which has been the high stakes testing in the state of Ohio and develop ways of helping teachers decipher this little understood language. My intent in this paper is to share some of these discoveries as many of us have attempted to help teachers make sense of high-stakes tests.

The objectives of this paper are three-fold. First, I will propose a model that has evolved for unifying massive amounts of conceptual and numerical information flowing

from the measures of the accountability movement in Ohio and the materials that are publicly available to educators. Second, I will demonstrate how this model was translated into usable forms of information that help teachers drive instructional practices in their classrooms. Third, I will demonstrate the efficacy of this model to improve state-mandated proficiency scores at the district level in one district that used this model.

Theoretical Perspective

Forty-nine of the 50 states have content standards which drive instruction within each state. Most of those states have a high-stakes testing program which compels teachers and districts to spend a great deal of time and energy preparing students to meet the objectives of the test. Some (see Kohn, 2000; Popham, 2001; Sacks, 1999; Smith 1998) would say that this is to the detriment of children educationally and psychologically. Others (see Tucker and Coddling, 1998) would argue that without standards-based assessment, there can be no accountability. This view is also reflected in the latest education proposal to come out of the White House (No Child Left Behind Act 2001). Regardless of which side of the debate one finds oneself, high-stakes assessment is a reality with which most schools nationwide are faced. And one in which educators play an important role, not necessarily as advocates pro or con, but as gatekeepers to determine that tests are being used accurately and fairly for all who are impacted by them (Newman and Benz, 1996).

Ohio's High-Stakes Accountability System

The situation in Ohio is similar. Schools are now issued a report card in the spring of the year in which they are placed in one of five categories, based on how many of 27 report card indicators the district obtains. The five catego-

ries are ranked, from lowest to highest, as follows: *Academic Emergency*, *Academic Watch*, *Continuous Improvement*, *Effective*, or *Excellent*. Of the 27 indicators currently on the district report card, 25 of the indicators are based on student performance on the previous year's 4th, 6th, 9th and 12th grade Ohio proficiency tests in *mathematics*, *science*, *writing*, *reading*, and *citizenship*. This will change in the near future since 2001 was the final year that the twelfth grade proficiency test was administered; thus this year is the last year that twelfth grade test results will be part of the report card. These "report cards" are sent home to parents and results are widely published in Ohio newspapers. Ohio educators are under constant pressure to raise or maintain high proficiency test scores.

Although the Ohio Proficiency Tests began in 1990, until recently Ohio educators were not provided with a number of tools that facilitate teachers' abilities to decipher the contents and structure of the proficiency tests in a way that promotes teaching and learning. Test results were reported for districts, buildings, classrooms and individual children, but the results did not allow teachers to locate essential information about the instructional needs of students. Student test results did allow politicians and the news media to publicly rank schools based on the numbers of children who performed at state proficiency standards. However, the test results that were provided were too limited and presented in a way that at best were confusing and at worst were misleading for instructional purposes.

The Rosetta Stone of Accountability: The Keys to Reading Proficiency Test Hieroglyphics

The discovery of the Rosetta Stone allowed linguists to uncover the language of Egyptian hieroglyphics and ultimately to decipher the messages on a variety of antiquities. Though the antiquities were known of for a long time, the key to decoding them was missing. For high stakes tests, there are several keys that teachers and other stakeholders need. As I have conceptualized them here, at minimum, teachers need:

1. Specific content standards around which the test will be organized.
2. Performance standards and an adequate number and variety of specific concrete examples of the types of items on the tests.
3. Item analyses reports that reveal the proportion of students who are able to pass each item on the test.
4. Easy access to graphically organized student data.

The Need for Content Standards

For the past ten years the Ohio Proficiency Tests (OPT) have not been based on content standards. Content standards provide educators with an understanding of the fundamental knowledge and skills that are required of students taking the tests. They provide teachers with information about the

cognitive capacities and essential thinking strategies that students must demonstrate to pass a test. They provide the schematic structure around which test items can be designed, curriculum can be developed and instructional practices can be organized. In their current state, the OPTs were built **before** the standards in most of the content areas. Add to this the fact that Ohio is revisiting their standards, already adopting new English Language Arts (ELA) and mathematics standards in December 2001 (so that there will be new standards adopted while the old tests are being used) and one can imagine the uncertainty and anxiety facing Ohio's teachers.

The Need for Performance Standards and Concrete Examples of Test Items

Ohio's proficiency tests not only lack content standards, but there are no performance standards associated with the tests. Performance standards describe the types of behavior students must demonstrate to show mastery of the content standards. Performance standards highlight what tasks or behaviors the students must demonstrate and how well the students must be able to demonstrate them to pass the test. The situation was further complicated by the fact that until a court order in July 1998, teachers did not have available to them past test items, which would provide an adequate number and variety of sample test items to employ as reference guides.

For the vast majority of that time, Ohio teachers were provided with only "theoretical statements" about the contents of the tests, written in the form of abstract definitions of *strands and outcomes*. From these theoretical statements, teachers were asked to divine the particular problems the state intended to test. For example, educators were not provided with performance standards or exemplary items for the strand *Numbers and Number Relations*. Rather, they were told that fourth grade students should be able to "decompose, combine, order and compare numbers" (*Numbers and Number Relations*; Outcome #6). Teachers would have to play a guessing game with the state as to which specific items would be tested in the domain of all possible items that require a fourth grader to *decompose, combine, order and compare numbers*. Without content standards, performance standards or a series of exemplary items this situation is, at best, a daunting task.

In their report, *Illustrative Language for an RFP to Build Tests to Support Instruction and Accountability*, The Commission on Instructionally Supportive Assessment (2001), chaired by W. James Popham, noted: "A state's high-priority content standards must be clearly and thoroughly described so that the knowledge and skills students need to demonstrate competence are evident" (p. 8). However, Ohio teachers were presented with neither clear and thorough content standards, nor performance standards. Moreover, for the majority of the past decade, Ohio educators were not provided with a variety of well-targeted sample items around which teachers could organize their instructional practices.

Without these three elements teachers do not know what they should teach and children do not know what they should learn.

The Need for Item Analyses

For most of the past 11 years, educators also were not provided with any information about how students performed on individual test items. They were merely provided with composite indicators of how their students performed relative to general strands. Moreover, there were no differential indicators in proficiency reports of how students performed on multiple-choice versus constructed response items.

Item analyses reports were not available to Ohio educators until February 2000. Even when they were available, their existence was not well publicized. They could only be obtained from the scoring contractor at an additional cost to the school districts, and they were only available to school districts for a limited period of time.

Once they were in the hands of educators, the item analyses ultimately revealed the logical and statistical fallacy of providing educators with a single composite indicator of how students performed on strands, rather than providing educators with information about how students performed on individual items within strands. The reports showed the degree to which student mastery of different types of items varied within strands.

As an illustrative example, we present Table 1, which shows the percentage of district students who were able to pass each of the four items from the strand *Geometry* on the 2000 Fourth Grade Mathematics Proficiency.

Table 1
Percentage of District Fourth Graders Who Passed the Four Geometry Items on the 2000 Ohio Fourth Grade Mathematics Proficiency Test by Item Type and Item Point Value

Items	Percentage Passing the Item	Type of Item	Points Value
1.	67%	Multiple-choice	1
2.	36%	Constructed-response	2
3.	95%	Multiple-choice	1
4.	67%	Multiple-choice	1

Rather than reporting individual rates of student passage on the four items that measure the strand *Geometry*, the state collapsed the four passage rates (67%, 36%, 95% and 67%) and reported them as single composite indicators on *Subscale (strand) Frequency Distributions*. These composite indicators masked the fact that students performed markedly differently on the different items within the strand *Geometry* (as well as most other strands). In addition, without item analysis reports, educators could not see the obvious difficulty students had correctly responding to constructed-response items. Not only did the students perform more poorly on the single constructed-response item in *Geometry* (38%), the item had the highest point value (2 points). The item analyses reports showed a clear pattern of

lower scores among items requiring constructed-responses. Creating constructed-responses in *Geometry*, as well as other areas was a skill that needed to be addressed, but without a report that disaggregated this information, this instructional implication was hidden from teachers.

One of the cardinal rules of data management is that categorically collapsed data always wipes out detail and reduces the specificity of information. Only when teachers have detailed information about their students' performance can they assess the instructional needs of their students and evaluate the effectiveness of their teaching methods. The item analyses reports, tied to the newly released items, would for the first time provide educators with the beginnings of a Rosetta Stone, which would ultimately lead educators to more defensible instructional decisions.

Item analyses would not only highlight what skills needed to be taught, but these reports would also provide educators with a number of useful comparison groups. Using item analysis reports, educators could see the relative passage rates of items for students *across the state*, students *in their districts*, students *in their buildings*, and students *in their classrooms*. Meaningful data always exist in context rather than isolation, and these comparison groups provided reference points, which allowed educators to better understand the relative effectiveness of their instruction, at least as it was measured on one high-stakes test.

The Need for Graphical Displays of Data

Aggregate and disaggregate data about student performance on the Ohio Proficiency Tests comes in the form of basic descriptive statistics on vendor printouts. Most of these data and their associated data analysis techniques are abstract and dependent upon various computations. Unfortunately, most educational stakeholders are not comfortable with numerical representations of test scores, nor familiar with statistical analyses. In order to communicate effectively with teachers, there is precedent to de-emphasize statistical representations of data in favor of graphical representations (Landwehr and Watkins, 1986; Tufte, 1983). Landwehr and Watkins, for instance, cite the Quantitative Literacy Movement, based on Tukey's (1977) work, that data should be used to find "patterns and surprises" (preface).

For those who are responsible for reporting on these patterns and surprises, those who are responsible at the district or building levels for evaluating educational programming, it is important that their reports and resources are tailored to the audiences for whom they are meant (Worthen and Sanders, 1987). In schools, the greatest impact will be made when that information is able to inform both classroom teachers and building principals about student achievement and needs. In order to be heard by and to make a difference with those on whom they impact, Weiss and Bucuvalas (1980) noted that the dissemination of this information must meet two criteria. It must have both *truth value* and *utility value* for the recipients. In other words, results

must be reported in such a way that those who receive them recognize their own experiences and expectations in them, as well as seeing that they provide recipients with a direction that they can take to improve on the situation.

In order to begin to understand a multivariate and complex world, those employing visual displays to investigate the world around them must bring together multiple relevant factors (Tufte, 1983). Displays of proficiency test results must bring together factors like numbers, pictures, words, colors, and graphs in ways that allow teachers to develop meaningful and complex understandings of their own classroom realities. Well-organized visual displays of data can function as quantitative arrows that point the way to the qualitative needs of learners. With this as a foundation, for instance, one district has used graphical representations of data to provide early identification of at-risk students at different levels of need for intervention (Snodgrass and Salzman, 1998). This current study was designed to follow that district as it created visual resources to be used within the district to further improve students' performance on state proficiency tests.

Creating Resources to Help Teachers Make Sense of Data

In order to unify the multiple sources of information that was swirling about the state, a group of district educators pulled these multiple sources into a set of visually-meaningful resources that became Proficiency Test Notebooks for teachers and principals. The purpose in creating these notebooks was to help teachers develop a more meaningful and complex understanding of the multiple variables that can impact their classroom reality. The notebooks were placed into the hands of teachers and unified the following information:

1. Multiple examples of past proficiency test items.
2. Information written beside each item that identified the strand and outcome the item is intended to measure.
3. State definitions of strands and outcomes.
4. Data about the proportion of students in the state, district and building that passed each item included in the notebook.
5. The correct response to each item.
6. A visual organizer that allows the user to quickly locate any item in the notebook based on any of the following criteria: **strand, outcome, item type** and more.
7. A list of content-specific and general vocabulary words from three years of past tests.

These notebooks were created with the intent to unify, organize and highlight proficiency test information in a way that provided educators with the keys for making instructional sense of the proficiency test. The *Fourth Grade Mathematics Notebook* was the first to be designed, followed by a *Fourth Grade Science Notebook*. This study evaluates

how the mathematics and science proficiency notebooks played a role in helping to impact a district's fourth grade proficiency test scores.

Organization of the Teacher Resource Notebook

The *Fourth Grade Mathematics Notebook* includes over 100 items taken directly from three years of past fourth grade mathematics proficiency tests. The test items were cut apart, reorganized and compiled within the notebook by the 25 outcomes, which are organized under the 8 strands that make-up the *Ohio Fourth Grade Mathematics Proficiency Tests*. These fourth grade mathematics strands and outcomes are listed in Table 2.

Table 2
The 8 Strands and 25 Outcomes that Are Measured on the Ohio Fourth Grade Proficiency Tests

Strands	Outcomes
1. Patterns Relations and Functions	1, 2
2. Problem Solving	3, 4, 5
3. Numbers and Number Relations	6, 7, 8, 9, 10, 11
4. Geometry	12, 13, 14
5. Algebra	15, 16
6. Measurement	17, 18, 19, 20, 21
7. Estimation and Mental Computations	22, 23
8. Data Analysis and Probability	24, 25

Each test item is placed on a separate page in the notebook (See Appendix A for example). A number of other pertinent pieces of test information are placed on the page next to each item. First, the *strand name* and *outcome number* the item is intended to measure are placed on the page beside the item (The state's definitions of each strand and outcome are inserted in separate section at the beginning of the notebook). Second, the *proportion of Ohio students, district students and building students* who were able to pass the item is also written on the page. Third, the correct answer to the item is identified. In the case of each multiple-choice item the correct response is placed directly on the item. In the case of a constructed-response item the state's scoring guideline for that item is located on the page(s) directly following the item.

The notebooks also include a visual organizer, which functions as a *table of contents* (see Appendix B). This visual organizer is located in the first section of the notebook. Using this organizer, teachers can quickly locate all items in the notebook by *strand, outcome, item type* (multiple-choice or constructed response), *date of administration* of the test item and more.

Like the notebook, the visual organizer is organized by strand and outcome, and can reveal unexpected surprises within the tests. A quick analysis of the visual organizer reveals interesting patterns in the mathematics proficiency tests that are not readily seen when looking through past test booklets themselves. For example, the visual organizer

shows that 3 of the 6 items over the past three years that are the test's most challenging items, *extended response items*, come from the *Measurement* strands. By organizing numbers and words in this way, teachers can discover *patterns and surprises* that are not readily seen when the numbers and words are in their unorganized form.

The notebook also includes a section that lists *content-specific* and *general vocabulary words* from the past 3 years of the *Fourth Grade Mathematics OPT*. The vocabulary words make-up the last three pages of the book. Understanding and accurately employing these vocabulary words is essential to academic success in the content areas. Unlike literature-based reading, where readers can glean the meaning of the text even with a general or colloquial understanding of words, to succeed in content-specific, non-fiction reading, learners must have explicit understanding of each technical word. Content area vocabulary words by their very nature have specific meanings and, to understand content area reading, students must have mastered and internalized those meanings. In this case, to succeed, both on the mathematics proficiency test and in mathematics in the world at large, students must employ mathematics vocabulary words in their most specific and literal sense.

Seemingly everyday words also can present a problem for students engaging in content area reading, when these common words form phrases that are unclear or unfamiliar to the students. For example, a child may have a problem understanding a test item when the child is accustomed to hearing his teacher use the phrase "each week," but the test employs the phrase "per week." The vocabulary list in the notebook provided a ready source of technical and non-technical vocabulary words from the *Mathematics OPTs*, which could be systematically woven into daily lessons.

District Procedures for Employing the Notebooks

The mathematics and science notebooks, their contents and their potential usage were presented to the district's six elementary principals by the notebooks' authors at an administrative retreat prior to the opening of the 2000/2001 school year. Principals were asked to share the notebooks and the information about the notebooks with their fourth grade teachers. The two notebooks were in the hands of all the district's fourth grade teachers in the fall of the 2000/2001 school year.

The notebooks were intended to be a tool for teachers to use and were employed in a variety of ways within the district. For example, one principal suggested to his teachers that they review only those multiple-choice items and their related concepts on which 50% or more of the district students were able to pass. After careful analysis of the available information from reviewing the notebook items, he hypothesized that multiple-choice items with less than a 50% district passage rate were either poorly constructed or developmentally inappropriate. In either case, he chose not to have his teachers waste valuable instructional time on these

types of concerns. Other principals empowered their teachers to determine to how best to use the information provided in the notebook. Though there was no one district-wide model for rolling out the information contained in the notebooks, all of the district fourth grade teachers used both the mathematics and science notebooks during the 2000/2001 school year.

These two notebooks provided teachers with enough exemplars of each strand and outcome that teachers could develop a conceptual understanding of how the state operationally defined the strands and outcomes. Providing teachers with the proportion of student passage for each item allowed them to identify the specific items that their students had mastered and target those that their students had failed to master. This information led teachers to more defensible decisions about the use of instructional time.

A principal of one of the district's traditionally lowest scoring elementary buildings commented to one of the authors of this paper that her fourth grade teachers told her that it was the sample items from the notebooks that pushed their students "over the top" on the mathematics proficiency test.

Visually Presenting Data to Help Teachers Decipher the Tests

Eight sets of bar graphs were later produced as companion tools to the *Fourth Grade Mathematics Notebook*, each representing one of the 8 different strands measured on the Fourth Grade Mathematics OPT (See Table 2 for a list of fourth grade mathematics strands). Figure 1 shows the bar graph that represented the items from the *Problem Solving* strand.

Each bar on Figure 1 represents a single multiple-choice item from the Fourth Grade Mathematics Notebook. The bars are organized in the same order as the items are organized in the *Fourth Grade Mathematics Notebook*. The height of the individual bars illustrates the proportion of Ohio students who passed each multiple-choice item from the strand entitled *Problem Solving*. Individual bars within a given strand are grouped and shaded by outcome. The first four darkly shaded bars represent student performance on outcome #3. The next three lightly shaded bars represent student performance on outcome #4 and the final four bars represent student performance on outcome #5.

The letters and numbers in the graph's legend are page numbers from the *Fourth Grade Mathematics Notebook* and relate to the bar's corresponding notebook item. The legend reads as follows: **strand-outcome-sample item**. Therefore,

PR-3-1 reads as *Problem Solving Strand-Outcome #3-Sample Item 1*;

PR-3-2 reads as *Problem Solving Strand-Outcome #3-Sample item 2*, etc.

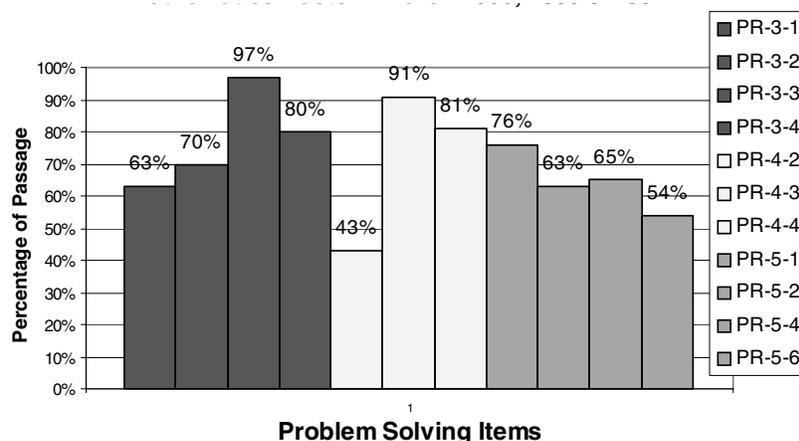


Figure 1. Percentage of Ohio Students Passing the Multiple-Choice Items on the Ohio Fourth Grade Mathematics Tests in March 1999, 2000 and 2001

An analysis of the bars in Figure 1 highlight how visual displays of data can be used to quickly identify patterns that can lead to meaningful instructional practices. For example, the three lightly shaded bars in the center of Figure 3 representing multiple-choice items are, in theory, designed to measure the same outcome (#4). However, there is a noticeable variation in the proportion of Ohio students—43%, 91% and 81, respectfully—who are able to pass each of the three items. Though ostensibly designed to measure the same domain of knowledge, the differences in the proportion of students able to pass the items lead one to question that assumption. Why, one might ask, is one item obviously so much more difficult for students to answer (43% rate of passage) than another item (91% passage rate)? These two test items are presented in appendices C and D as they appear in the proficiency test. An analysis of the actual two items reveals a surprise that might otherwise have remained hidden, providing teachers with instructionally useful information about the similarities and differences between the items.

A Closer Look at Surprises Revealed

Both items are short, story problems that require that students recognize information necessary to solve the problem. A quick review of the two items highlights differences between the two items. For instance, the “easier” item (Appendix C) presented students with familiar concepts, like purchasing school supplies (pencils, notebooks, etc.) and receiving change. The more difficult item (Appendix D) also presented students with a short, story problem, but this time about hourly wages and the amount of money earned per week. *Hourly wages* is a concept that is certainly less familiar to fourth graders than purchasing school supplies and receiving change. Moreover, the easier item has a single, correct answer. The harder item requires that a child make a choice of which answer is the *best answer*. The typical fourth grader generally does not consider that there might be a *best answer*. Many fourth graders will likely choose the first answer that seems correct, even if that answer is only partially correct. Both A and B, for instance, are pieces of in-

formation that can help you solve this problem, though C is the only answer that provides **all** the information that students need. This is further complicated by the fact that students were never explicitly told **within the prompt** that there is, or that they should be looking for, a *best* answer. Students must infer that information from the prompt.

In this case, the differences in the difficulty of the items may be less related to the mathematical thinking the child is asked to perform than the story that the problem is embedded in and the structure of the item itself. The discrepancy between the scores on the two items (43% rate of passage versus 91% rate of passage) is a good example of how noticing a *quantitative* difference between the passage rates on two items can point to *qualitatively* different instructions need of children.

Engaging teachers in detailed analysis of items in this way leads to another instructionally useful situation. Two different test items, presented in Appendices E and F, are categorized under the strand *Numbers and Number Relations*. These items present students with virtually the same task. Curiously, however, there is an 11% difference in the proportion of district students who correctly answered one item over the other item. Both items require that students identify the correct symbol “<, > or =” that shows the relationship between two simple fractions with unlike denominators. Further review of the two items shows that the item with the higher rate of passage, 72%, (Appendix E) includes two pie graphs. The more difficult item does not include the pie graphs. Students are instructed in the easier problem that they can use the pie graphs to help them solve the problem. The pie graphs, it appears, are designed to function as visual aids for the students. As the difference in the percentage of students who were able to pass the two items suggests, they do appear to assist the students. The instructional implications resulting from the differential rate of passage of these two problems are that visual aids, like pie graphs, assist students in solving problem or checking their work. Again, the *quantitative* differences in the magnitude

of the two bar graphs help teachers to identify the *qualitative* instructional needs of learners.

Professional development in the district was built on these types of analyses of individual and groups of items purporting to represent the same strands and outcomes. Ultimately, creating the notebooks and graphs, and engaging educators in using them, resulted in their asking questions like:

- Why do two items which purport to measure the same strand and outcome result in large differences in the percentage of students able to pass each of the items?
- Is there a relationship between the proportion of fourth grade students who pass an item and the developmental appropriateness of the item, the quality of the item or the way we teach the item?
- How does our students' performance compare to student performance across the state on each item?
- Do our children have the same level of mastery on constructed-responses as multiple-choice items?

The notebooks and graphs cannot in and of themselves change student performance. Nor can the notebooks tell educators which items on the proficiency tests are developmentally appropriate and critical to students' understanding of mathematics and which items are not. These decisions are judgments. Good judgments do not exist in a vacuum or by playing a guessing game with the state. The graphs and notebooks can provide a way to decipher the contents of the test and help teachers make defensible instructional decisions. Only when resources, like these notebooks, are available and used as part of a systematic and integrated curricular initiative, will district personnel, both teachers and principals, maximize teachers' abilities to help students reach their potential as measured on high stakes tests.

Results

The data in Table 3 show that there was an increase of 12% in the proportion of students able to pass the mathematics proficiency test in March 2001, when teachers employed the mathematics notebook, over the March 2000 administration of the test, when the notebook was not employed. Table 3 also shows a parallel increase in science (+8%) from 2000 to 2001. Chi-square tests indicated that these gains were statistically significant.

The question of whether or not the increased proportion of student passing the Mathematics and Science tests might be attributed to factors other than the notebooks exists. For example, the improvement in the scores could be the result of an inherent difference in the ability of the two sets of students who participated in the March 2000 and March 2001 tests. One way to investigate the likelihood that the difference in the scores is a function of an inherent difference in the abilities of the two sets of students is to compare the two sets of fourth graders' scores on the other 3 proficiency tests: *Writing*, *Reading* and *Citizenship*, since

notebooks were neither used nor developed in those areas. The percentage of passage in *writing* between March 2000 and March 2001 increased by two percent (+2%). In *reading* the percentage of passage remained exactly the same (0% difference) both years (68%). In *citizenship* the percentage of passage increased by 1%. None of the gains were statistically significant. While there are clearly other plausible reasons for the differences in scores, it appears that the notebooks and adjunct materials may provide a feasible explanation for the improvement in mathematics and science proficiency scores between March 2000 and March 2001 and not merely reflect any inherent difference in ability of the two different sets of fourth grade students.

Table 3
Comparison of the Percentage of District Students Who Passed the Ohio Fourth Grade Mathematics and Science Proficiency Tests between March 2000 and March 2001

Test	March 2000		March 2001		% Diff
	N	% Pass	N	% Pass	
1. Mathematics	284	65%	256	77%	+12%*
2. Science	285	59%	255	67%	+8%*
3. Writing	283	88%	250	90%	+2%
4. Reading	281	68%	250	68%	0%
5. Citizenship	285	75%	249	76%	+1%

Note: * indicates chi-square test significant at $p < .05$

Educational Importance

The development of the workbooks seems to have had a three-fold benefit. First, large amounts of conceptual and numerical information, flowing from the measures of the accountability movement have been unified in a way that makes the intended outcomes of the *Ohio Fourth Grade Proficiency Test* more accessible to the major stakeholders. Second, this model has unified numbers, words and graphical information in ways that help teachers drive instructional practice in their classrooms. Finally, the efficacy of this model appears to be supported. A statistically- and practically-significant increased proportion of students did pass the mathematics and science tests during the school year when the notebooks were employed (2000/01) as compared to the proportion of students passing the mathematics and science tests during the school year prior to their development (1999/2000). In areas where notebooks did not exist (*writing*, *reading* and *citizenship*) increases in performance were not evident at a district level.

Based on these experiences, in order for statewide testing programs to be useful for instructional purposes educators must be provided with at least the following information:

1. Specific content and performance standards around which items are developed. Currently there are no apparent content or performance standards around which the Ohio Proficiency Tests are built. Though that may change with the development of new content standards

in 2002, the state's current practices do not inspire confidence.

2. An adequate number and variety of specific concrete examples of the types of items on the proficiency test.
3. Item analyses reports that present the rate of student passage of each *item* by state, district, building and classroom results. Item analyses reports should be a standard contracted part of the test package negotiated by the state with the test vendor and, therefore, free and readily available to all districts. While there is great value for teachers and other district personnel to engage in the inquiry and analysis necessary to create resources like these, it is our belief that, if a state is going to hold schools accountable for meeting high-stakes accountability measures, they have a concomitant responsibility to invest the necessary monies to ensure that teachers are provided with the tools to help their students succeed.
4. Easy access to graphically organized student data presented in a form that is readily understandable to the statistically less inclined user.

A Caveat

Currently in Ohio, rather than creating new test items after the old test items have been released to the public, many test items from all the tests (reading, writing, mathematics, science and citizenship) are being reused by the state in subsequent versions of the proficiency tests. This is arguably a questionable decision. For, there is little doubt that the reuse of test items after they have been released contributes to an increase in the percentage of students who are able to meet Ohio standards on the Ohio Proficiency Tests. Moreover, the more savvy educator immediately realizes that as items are being released and reused less and less can be inferred from student test results about students' ability to perform items in the greater domain that the test purports to be measuring. This is an unfortunate situation.

The intent of the authors of the notebooks was not to provide items to teachers so teachers could teach *the test*. Rather, the purpose of the authors was to provide a rudimentary model of an *educational Rosetta Stone* that would help teachers decipher the contents of the proficiency test at a deep and complex enough level to identify the basic and fundamental instructional needs of students. These needs are based on skills and knowledge that always exist independent of any test that attempts to measure them.

Epilogue

The intention of this paper was to introduce readers to the work done in trying to make sense of high-stakes testing information. The author is part of a grass roots network (Support Team for Assessment, Testing, and Statistics or STATS) of testing supervisors, curriculum directors, university professors and other interested educators who meet monthly to

discuss testing and assessment issues. Since these notebooks were created, other STATS-member districts, having heard about them and seen them, have used the template we've described and created other notebooks to address aspects of the test that had yet to be organized in this way. In 2003, all of these notebooks were compiled, burned onto a CD and sent to every school district in the state of Ohio.

References

- Commission on Instructional Support Assessment (2001, October). *Illustrative language for an RFP to build tests to support instruction and accountability*. Chair: W. J. Popham.
- Kohn, A. (2000). *The case against standardized testing: Raising the scores, ruining the schools*. Portsmouth, NH: Heinemann.
- Landwehr, J. M., and Watkins, A. E. (1986). *Exploring data*. Palo Alto, CA: Dale Seymour Publications.
- Newman, I., and Benz, C. (1996 October). *Standardized testing*. Paper presented at the Mid-Western Educational Research Association, Chicago, IL.
- Ohio Department of Education. (1994). *Language Arts: Ohio's Model Competency-Based Program*. Columbus, OH: Author.
- Popham, W. J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus.
- Smith, F. (1998). *The book of learning and forgetting*. New York: Teachers College Press.
- Snodgrass, D. G., and Salzman, J. A. (1998). *Graphical representations of data: Iowa Tests of Basic Skills, Ohio sixth grade proficiency tests, and SB 55*. Paper presented at the Mid-Western Educational Research Association, Chicago, IL. (ERIC Document Reproduction Services No.)
- Tucker, M. S., and Coddling, J. B. (1998). *Standards for our schools: How to set them, measure them, and reach them*. San Francisco: Jossey-Bass.
- Tufte, E. R. (1983). *Visual display of quantitative information*. Cheshire, CT: Graphics Press.
- Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison Wesley.
- Weiss, C. H., and Bucuvalas, M. J. (1980). Tests and utility tests: Decision makers' frames of reference for social science research. *American Sociological Review*, 45, 302-313.
- Worthen, B. R., and Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. New York: Longman.

Fourth Grade Mathematics Proficiency Test

Test Date: 99 **Percentage of Students Passing**
Item #: 31 State 43 % District _____ %
Strand: PR School _____ % School _____ %
Objective: 4 School _____ % School _____ %
Item Type: MC SA ER

31. Mavis works at the hardware store. Her hourly wage is \$4.50.
How much money is Mavis paid for one week's work?

Which piece of information will help you solve this problem?

- A. The number of hours she works each day
- B. The number of days she works each week.
- C. The number of hours she works each week.

State, district and building passage rates based on individual Strands, Objectives and Item Types
Prepared by: Innocenzi and Snodgrass, July 2000

PR 4-2

Fourth Grade Science Proficiency Notebook

Table of Contents

Pages 49-63

Section I

State definitions of *Strands, Objectives* and *Levels of Understanding in science*.

Section II

Examples of *Strands, Objectives* and *Levels of Understanding* taken from actual 1999 and 2000 Fourth Grade Proficiency Tests, as follows:

Strand	Objective	Example	Type	Levels of Understanding	Strand/ Objective/Example	Year/ Item No.
1. Nature of Science	1	1	SA	Extending Scientific Knowledge	NS-1-1	99-15
1. Nature of Science	1	2	MC	Processing Scientific Knowledge	NS-1-2	99-38
1. Nature of Science	1	3	MC	Acquiring Scientific Knowledge	NS-1-3	00-1
1. Nature of Science	1	4	MC	Acquiring Scientific Knowledge	NS-1-4	00-5
1. Nature of Science	1	5	SA	Extending Scientific Knowledge	NS-1-5	00-12
1. Nature of Science	2	1	MC	Acquiring Scientific Knowledge	NS-2-1	99-4
1. Nature of Science	2	2	MC	Processing Scientific Knowledge	NS-2-2	99-11
1. Nature of Science	2	3	MC	Processing Scientific Knowledge	NS-2-3	99-19
1. Nature of Science	2	4	MC	Processing Scientific Knowledge	NS-2-4	00-16
1. Nature of Science	2	5	SA	Processing Scientific Knowledge	NS-2-5	00-31
1. Nature of Science	3	1	MC	Processing Scientific Knowledge	NS-3-1	99-27
1. Nature of Science	3	2	MC	Extending Scientific Knowledge	NS-3-2	99-35
1. Nature of Science	3	3	MC	Acquiring Scientific Knowledge	NS-3-3	00-7
1. Nature of Science	3	4	MC	Processing Scientific Knowledge	NS-3-4	00-38

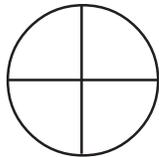
Appendix C

24. At the school store, José bought 2 pencils for 10¢ each, a notebook for 65¢, and a candy bar for 40¢. To find out how much change he will get, you need to know
- A. How much 2 notebooks would cost
 - B. How much money he gave the salesperson.
 - C. How much he saved by buying one notebook.

Appendix D

31. Mavis works at the hardware store. Her hourly wage is \$4.50. How much money is Mavis paid for one week's work?
- Which piece of information will help you solve this problem?
- A. The number of hours she works each day
 - B. The number of days she works each week
 - C. The number of hours she works each week

Appendix E



Compare the fractions. Then chose the correct symbol.
You may use the pictures to help you solve the problem.

$$\frac{2}{4} \square \frac{3}{6}$$

- A. =
- B. <
- C. >

Appendix F

32. Compare the two fractions. Then choose the correct symbol.

$$\frac{2}{3} \square \frac{1}{2}$$

- A. >
- B. <
- C. =

Two Methods of Estimating a Study's Replicability

Isadore Newman
University of Akron

Keith McNeil
New Mexico State University

John Fraas
Ashland University

Abstract

This article presents two methods of estimating a study's replicability that researchers should consider reporting along with their statistical significant and effect size findings. One method of estimating the replicability of the findings deals with replication in the exact same system. The second method, which may contain subjective probability values, is used to estimate the replicability of a study's findings in a system that may differ from the initial system with respect to salient variables. The incorporation of the replicability estimates delineated in this paper would provide critical information to decision makers about the likelihood that the implementation of a particular method or treatment would produce similar results in their systems.

Most researchers would agree that the statistical significance levels and effects sizes reported in a study are important pieces of information (Fraas and Newman, 2000; Levin and Robinson, 2000; Robinson and Levin, 1997; Thompson, 1996, 1997, 1999a, 1999b). We take the position that what may be the most relevant piece of information practitioners and policy makers need to glean from a study is the ability of the study's findings to replicate (assuming the intent of the study is inferential not descriptive). If a study's findings are unlikely to replicate, the study's significance levels and effect sizes are virtually meaningless to interested practitioners and policy makers. Thus we as applied statisticians have the responsibility not only to provide estimates of a study's replicability but also delineate the assumptions on which these estimates are based.

In this article we define two types of replications and present methods by which researchers can provide estimates of each type. The definition of replicability that we are developing is the extent to which a curriculum, treatment, etc., can be successfully implemented in two types of systems. One system is assumed to be an exact replication of the one used in the initial study. The replicability estimate that deals with this exact same system criterion is based on the same underlying assumptions of calculating the p value on a random sample of a known population. The other replicability estimate assumes the system of interest is not an exact replication of the system used in the original study. This replicability estimate is based not only on random sampling assumptions but also on probability estimates, which will be somewhat subjective, that certain key variables will differ between the system used in the initial study and the system of interest. It is important to note that this second type of replication estimate can be calculated before as well as after the initial study is implemented. If it is calculated before the implementation of the study, it could assist in the re-design of the study before it is actually implemented. If

calculated after the study is implemented, it will have implications for practitioners and decision makers.

Statistically Significant Exact Replications of a Study

One value we believe should be contained in research reports is the likelihood that the study's findings are replicable in the same system. Such a value should not take the place of statistical significance tests but rather should be reported along with them. We agree with Robinson and Levin (1997) who expressed the position that the probability value (p value) produced by a statistical test is an important piece of information to report in a quantitative study. Robinson and Levin stated that "authors should *first* indicate whether the observed effect is a statistically improbable one (e.g., is the difference greater than what would be expected by chance?)" (p. 22).

It is important, however, not to misinterpret a p value with respect to the likelihood of the replication of results (Nickerson, 2000). This point was addressed by Posavac (2002) who stated that "some believe [incorrectly] that rejecting a null hypothesis means that at least 95% of replications would be statistically significant" (p. 102). Posavac does take the position, however, that rejecting a null hypothesis should increase the researcher's expectation that replications of the research would yield similar results.

Using p Values to Estimate Statistically Significant Exact Probabilities

Greenwald, Gonzalez, Harris, and Guthrie (1996) presented an analytic method by which a p value can be converted into a probability estimate that an exact replication of the research would produce a statistically significant result. Posavac (2002), who elaborated on the method proposed by Greenwald, et al., noted that an "*exact replication*

means that the initial experiment is repeated using the same independent and dependent variables with the same number of participants selected in the same way from the same population” (p. 102). In this type of replication the difference between the replication and the original study is due to random variation. We believe that this is one type of replication that should be addressed by researchers.

Green et al. (1996) and Posavac (2002) suggest that the probability of a statistically significant exact replication (SSER) can be estimated from the probability of the statistical test. As a means of demonstrating how a researcher could convert a p value from a statistical test to an estimate of the SSER probability, we present a brief discuss of the procedure. It is beyond the scope of this paper to present the rationale on which this procedure is based. We encourage interested readers to review the works published by Greenwald et al. and Posavac for a more in-depth discussion of this concept.

An Illustration

To illustrate the calculation of the SSER probability value, assume researchers are testing the difference between sample means of two independent groups consisting of 20 subjects each. Further assume that the t value produced by the difference between the two means recorded for their study was 2.150. Since this observed t value (t_{obs}) is greater than the two-tailed critical t value (t_{crit}) of 2.024 for an alpha level of .05, the researchers would declare the difference between the two group means to be statistically significant. The question we believe is important for these researchers to address is: What is the chance that the difference between the two group means recorded for an exact replication of the study would be declared statistically significant?

Calculation of the SSER probability value. As noted by Posavac (2002), the probability of obtaining a SSER can be obtained by executing three steps. First, the replication t value (t_{rep}) is calculated by subtracting the critical t value used in the initial study from the study’s observed t value. Thus the t_{rep} value is calculated as follows for our hypothetical example:

$$\begin{aligned}t_{rep} &= t_{obs} - t_{crit} \\t_{rep} &= 2.150 - 2.025 \\t_{rep} &= .125\end{aligned}$$

Second, the researchers would obtain the one-tailed probability for this t_{rep} value of .125 with 38 degrees of freedom. With respect to the procedure used in this step, Posavac (2002) stated that “a one-tailed test is used because one would want a replication to produce means of the same relative magnitudes as found in the first study (p. 108).” The one-tailed probability for the t_{rep} value of .125 with 38 degrees of freedom is .45.

Third, the researchers subtract the .45 probability value from 1.00, which produces a value of .55. This value indicates that the chance that an exact replication will be statistically significant is .55.

Points to note regarding the SSER probability value.

Three points should be noted regarding this SSER probability value of .55. First, the SSER probability value is a function of the p value. However, practitioners need to be careful not to directly interpret the p value as a replicability value. Second, Greenwald et al. (1996) and Posavac (2002) recommended that SSER probability values should be considered upper limits. The reason for this recommendation is based on the fact that “even in a careful replication the participants would be a different sample from the population, the calendar date would be different, the weather would be different and so forth” (Posavac, p. 111). Third, researchers may be surprised that for a study, such as the one used in our example, which had 38 degrees of freedom and an observed t value of 2.150 ($p = .038$), the chance that an exact replication will be statistically significant (SSER probability level = .55) is only slightly above the 50-50 level. In fact an observed t value for this hypothetical study would need to be 2.874, which produces a p value of .01, in order for the a SSER probability level to reach the .80 level.

To further emphasize this third point, a review of values produced by Posavac (2002) reveals that when degrees of freedom value is at least eight and the p value is .05, the SSER probability value will be .50. That is, there is a 50-50 chance of replicating significant findings. If the degrees of freedom value is at least eight and the p value is .01 for a two-tailed test, the SSER probability value will not be less than .73 or greater than .84. And if the degrees of freedom value is at least eight and the p value is .005, the SSER probability value will not be less than .80 and not greater than .92. (It is interesting to note that these replicability values are less for corresponding p values for one-tailed tests.) Thus researchers need to be careful not to assume that statistically significant findings automatically mean that the chance of obtaining statistically significant exact replications for the study will be high. For this reason we believe that researchers should report the SSER probability value along with the probability of the observed t test.

Replication in a Different System

We believe that a second type of replication of findings is important for researchers to address. That is, the type of replication that deals with the question: Would the study’s findings replicate in a system different from the one used in the initial study? It should be noted that we consider this type of replication of findings important even if an individual is interested in the same system in which the study was conducted, assuming the system is a dynamic one. That is, the system experiences considerable change with respect to the variables that may influence the replicability of the findings. Since most people attempt to relate research findings to systems that are different from their own or, at least, relate findings to systems that are similar but dynamic, we believe obtaining a likelihood estimate for this type of replication would be most valuable for them. The remaining

portion of this section of the article presents our *preliminary* attempt to develop a procedure for calculating such a likelihood estimate.

Estimate Procedure

The procedure we are proposing for the estimate of the likelihood of replication of findings for a system different from the one in which the study was conducted can best be presented through an example modeled on a study conducted by Benson, Aronson, Desmett, Shaheen, and Showalter (2002), which presented an evaluation of a multiage classroom educational program. In our example we have children in grades 1-3 who were grouped in the same classroom and their teacher stayed with them for the three years. The evaluation indicated that the teachers volunteered for the project and were enthusiastic about the concept of multiage education. The project was supported fully by the principals and was enthusiastically supported by the parents. Achievement scores indicated moderate success of the project as compared to national norms and comparison students in the same school.

Internal validity issues are apparent, since parents voluntarily allowed their children into the project (see Campbell and Stanley, 1963, for a discussion of internal validity issues). Enthusiastic teachers might generate better results, no matter what the curriculum. In addition, a supportive principal might be partly (or entirely) responsible for the achievement results. Other internal validity concerns could also be raised.

External validity issues are also of concern with this study. Would the same effects be observed with less enthusiastic teachers? Will the same effects occur after the novelty of the multiage grouping wears off? Other external validity concerns could be raised (see Campbell and Stanley, 1963, for a discussion of external validity issues). The concept of replicability, though, is different from internal validity and external validity. It is based on the realization that any implementation is accomplished in a system and the realization that that system is likely to be dynamic. We believe that the likelihood of replicating a study's findings in a different system or even the same dynamic system is crucial to estimate.

Important variables. The first step in the estimation process is to identify key variables that influenced the findings but may be different in the new system. Let us assume that for our multiage project example four such variables were identified:

1. Twelve volunteer teachers were used.
2. The study involved supportive principals.
3. The study used 240 volunteer (supportive) parents.
4. A total of 5 days of in-service training was given to the teachers on the multiage project.

As an illustration of how these variables could influence the replicability of the findings of the original study,

consider the principals variable. If a principal leaves, the project will, in all likelihood, be supported less by the new principal. The new principal may even kill the project, not because the project is ineffective, not because the concept of multiage education is bad, but because the crucial component of the system (the principal) does not believe in or want the project.

The likelihood of each crucial component changing should be taken into account when the project is envisioned. If a particular component is likely to change, then the project should be devised so it is immune to that change—in the case of principal change—the project should be made “principal proof.”

Once the variables are identified, the second step is to estimate the proportion of the R^2 value accounted for by each variable, the probability of that variable changing, and the probability of the changed variable being negatively influential on the original findings. Table 1 contains such hypothetical values of these estimates for our example.

Table 1
Proportion of the R^2 Accounted for in the Dependent Variable and the Probability Values for Each Variable

Variable	Proportion of R^2	Estimate of the Probability of Change	Estimate of the Probability of Negative Impact
Teacher	.50	.30	.03
Principals	.20	.70	.60
Parents	.10	.33	.02
Staff Development	.20	.02	.02

The proportion of the R^2 value accounted for by each component is determined. This could be accomplished with GLM if enough implementation sites were available (similar to meta analysis), or conceptualized either before the study started or afterwards. In the example, here we provide “educational guesses.” For instance, it is likely that some teachers will leave the project. Some may become disillusioned with the project or with education in general. Others may find a more lucrative job in another district or another profession. Nevertheless, other enthusiastic teachers are likely available, so the systemic effect on the project of teacher change would be minimal.

On-the-other hand, the likelihood of a principal leaving the system is high (estimated to be .70 in a three-year period) and the likelihood of the replacement being equally enthusiastic is low (.40). Indeed, most replacement principals may gut the project, leading to absolutely no replicability from the component of the principal. Therefore, because of the high probability of principal change, and high probability of a different (lower) level of support, the overall replicability is lowered.

Parent turnover will be at least 33% every year, with third graders moving to fourth grade. But we suspect that the parents of the incoming first graders will be just as enthusiastic (maybe even more so if the project is a success).

Thus, the high turnover rate (large system change) of parents will have little effect on replicability—the project is “parent proof.” If the staff development is “packaged” then it could easily stay the same from year to year. This part of the system would likely be stable.

Actual probabilities may be quite difficult to determine. To deal with this problem, one might rate the stability of each component on a 1 to 5 scale, with 5 being the most stable. Such estimates and the calculation of the reliability value for the multiage example are listed in Table 2. It should be noted that a replicability value calculated in this manner would produce higher values the more stable key variables are from the system used in the initial study and the system of interest especially for the variables that account for the higher proportion of the R^2 value.

Table 2
Calculation of the Replicability Value

Variable	Proportion of R^2	Stability	(Proportion of R^2) * (Stability)
Teacher	.50	4	.50*4 = 2.00
Principals	.20	1	.20*1 = 0.20
Parents	.10	5	.10*5 = 0.50
Staff development	.20	5	.20*5 = 1.00
			Replicability = 3.70 / 5 = .74

Estimating replicability before implementing. If a researcher calculated replicability before first implementing a new project, and obtained a low replicability value, the researcher might try to re-conceptualize the project by either doing something to minimize system change or to minimize the effects of the change within any one component. One could minimize system change in the multiage project by getting the school board to mandate multiage in all elementary schools or find another district where all the principals are supportive of multiage programs. It should be noted that a replicability value calculated in this manner would produce higher values the more stable key variables are from the system used in the initial study and the system of interest, especially for the variables that account for the higher proportion of the R^2 value.

Minimizing the effects of change in teachers could be accomplished by each project teacher identifying a non project teacher who would like to be in the project and then keeping that teacher informed about multiage grouping during the year. This “information partnership” actually becomes a new component of the project (or at least modifies the teacher component.) Curricula that purport to be “teacher proof,” such as highly prescriptive direct instructional methods, are another example of minimizing the effects of teacher change.

If the replicability index is low, and the researcher cannot identify changes or strategies that would make it higher, then the project should not be implemented. The time of teachers, principals, parents, staff developers, and especially students should not be wasted. If there is very little hope for replication of a particular project, then we have no business investigating the effectiveness of that project.

How about changes in system components not relevant to the project? Changes in components that are not relevant to the project will not affect the replicability, by definition. Nor will these changes affect the index, as the percent of variance accounted for is 0 and the contribution of that component would be 0. Unfortunately, in most educational systems, many components can influence the success of a project.

Implications

The implications of this article relate our position that statistical significance and effect size are important concepts, but they must be examined in light of replicability. Replicability is, in and of itself, not a one-dimensional concept but a multi-dimensional one. In this paper we identified two types of replication estimates. The first type is the SSER probability estimate, which is based on traditional statistical assumptions and probability concepts.

The second type is related to design and subjective probability issues. This approach provides a number of advantages. First, it can assist in the teaching of research design. That is, teaching this replication estimate emphasizes the need for researchers to attempt to identify the relevant variables in a study. Second, it can improve communication among researchers regarding relevant variables in a study in order to improve the design of such studies. Third, it encourages the use of meta-analysis to identify relevant variables. Fourth, it provides a method of simulating the effects of the relevant variables on replicability of findings. One can simulate small changes or large changes on relevant variables and the impact of these changes on replicability. As one can see, this second estimate is not a static approach but a dynamic one and may only be limited by the investigators’ creativity and insight.

An emphasis on replications has implications for researchers regarding the research methodology they use. That is, researchers should consider conducting partial replications. Partial replication can be conducted by one of two approaches. First, half of the study could be an exact replication, and the other half could be an extension (into another grade level, using different in-service materials, or checking on efficacy in another bureaucratic situation). Second, the researcher could put a slight twist on the implementation, by reducing or eliminating a component, shortening the period, streamlining in-service, or monitoring more closely the actual implementation.

We believe that an emphasis on replicability estimates are as important to analyzing the data contained in a study as are the statistical test results and effect size estimates. The value of a study’s results can be better assessed by researchers and practitioners when all three types of information (i.e., replicability estimates, statistical test results, and effect size estimates) are reported.

References

- Benson, S. N. K., Aronson, E., Desmett, P., Shaheen, M., and Showalter, J. (2002, October). *Multiage education: A process and product evaluation*. Paper presented at the annual meeting of the Midwestern Educational Research Association, Columbus, OH.
- Campbell, D. T. and Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally College Publishing Co.
- Fraas, J. W., and Newman, I. (2000, October). *Testing for statistical and practical significance: A suggested technique using a randomization test*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago, IL.
- Greenwald, A. G., Gonzalez, R., Harris, R.J., and Guthrie, D. (1996). Effect sizes and p values: What should be reported and what should be replicated? *Psychophysiology*, 33, 175-183.
- Levin, J. R., and Robinson, D. H. (2000). Rejoinder: Statistical hypothesis testing, effect-size estimate, and the conclusion coherence of primary research studies. *Educational Researcher*, 29(1), 34-36.
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5, 241-301.
- Posavac, E. J. (2002). Using p values to estimate the probability of a statistically significant Replications. *Understanding Statistics*, 1(2), 101-112.
- Robinson, D. H., and Levin, J. R. (1997). Reflections on statistical and substantive significance, with a slice of replication. *Educational Researcher*, 26(5), 21-26.
- Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, 25(2), 26-30.
- Thompson, B. (1997). Editorial policies regarding statistical significance tests: Further comments. *Educational Researcher*, 26(5), 29-32.
- Thompson, B. (1999a). Statistical significance tests, effect size reporting and the vain pursuit of pseudo-objectivity. *Theory and Psychology*, 9(2), 191-196.
- Thompson, B. (1999b). Journal editorial policies regarding statistical significance tests: Heat is to fire as p is to importance. *Educational Psychology Review*, 11(2), 157-169.

The Mid-Western Educational Research Association *Gift Membership*

A gift membership has been given to you, _____

by _____

Your name is now included as a member in one of the most recognized, well respected, educational research groups in the United States and Canada. Your **one year membership** includes a subscription to the ***Mid-Western Educational Researcher***, the Association's journal that highlights research articles, features, interviews, and Association news. Members pay reduced registration fees for the annual meeting held in Columbus, Ohio, in October. This conference attracts many nationally recognized leaders in educational research. Enjoy your membership.



Thank you for providing your colleague, student, or friend with a special one year gift membership to the Mid-Western Educational Research Association. It is a gift of professional involvement that is sure to be appreciated throughout the year. To give your gift membership fill out the top portion of this card and use it to inform the recipient of the gift membership; then fill out the bottom portion of this card and mail it with your check to: **Jean W. Pierce—Dept EDCSE—Northern Illinois Univ.—DeKalb, IL 60115**

	Person Receiving Gift Membership	Person Giving Gift Membership
Name	_____	_____
Address	_____	_____
	_____	_____
	_____	_____
Affiliation	_____	_____
Work Phone	_____	_____
Home Phone	_____	_____
E-mail	_____	_____
Fax	_____	_____
Division	_____	_____
Preference (optional)		

Check one below and make check payable to Mid-Western Educational Research Association.

Professional Membership—\$45

Student Membership—\$25
Student must be currently enrolled.

Lifetime Membership—\$450



The Ohio State University
2500 Kenny Road
Columbus, Ohio 43210

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