Volume 18, No. 4–Fall 2005

MID-WESTERN EDUCATIONAL RESEARCHER

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

Pulliam Hall, Southern Illinois University Carbondale

On The Cover

Pulliam Hall

Pulliam Hall is located at Southern Illinois University Carbondale, the flagship campus of the Southern Illinois University system. Founded in 1869, SIUC has a Carnegie distinction of Doctoral/Extensive. Guiding the University's development is ***Southern at 150: Building Excellence Through Commitment***. The primary goal of this strategic plan is to have SIUC rank among the nation's top 75 public research institutions by the time the University celebrates its 150th anniversary in 2019.

Pulliam Hall was completed in 1951 and housed the University School, offering early childhood education through high school. Originally called the Model School, University School was launched when Southern Illinois Normal University was established in 1869. University School closed in 1971.

From its inception as a small teachers college, Southern Illinois University Carbondale has been dedicated to the preparation of teachers for the schools and classrooms of Illinois. This tradition continued with the creation of the College of Education in 1944. Today, the College of Education and Human Services prepares not only teachers and administrators for the public schools, but also professionals for a wide array of careers in such areas as athletic training, behavior analysis and therapy, communication disorders and sciences, counselor education, educational measurement and statistics, public health, recreation administration, therapeutic recreation, rehabilitation services, and workforce development.

The College of Education and Human Services includes the Departments of Curriculum and Instruction, Educational Administration and Higher Education, Educational Psychology and Special Education, Health Education and Recreation, Physical Education, Workforce Education and Development, the Rehabilitation Institute, and the School of Social Work. In 2019 the College will celebrate 150 years of service to the people of Illinois. The College is proud of its talented and dedicated faculty and students as well as more than 50,000 alumni across the state, nation, and world.

Call for Manuscripts

The ***Mid-Western Educational Researcher*** is a scholarly journal that publishes research-based articles addressing a full range of educational issues. The journal also publishes literature reviews, theoretical and methodological discussions that make an original contribution to the research literature, and feature columns. There are four issues of the journal published annually.

The journal is accepting manuscripts for review and possible publication. Manuscripts are submitted to blind reviews by three researchers with knowledge of the literature in the appropriate area. The editors will review the manuscript and make the final decision. The review process requires approximately four months.

Manuscripts are accepted from faculty, students, and professionals working in educational or non-educational settings. Membership in the MWERA is not required in order to submit a manuscript for review. The editors encourage the submission of revised papers that have been presented at the annual meetings of the MWERA, AERA, and other professional organizations.

Manuscripts may be submitted for review as hard copy or electronically.

Hard Copy Submission. Submit four (4) copies of the manuscript with a cover letter to Deborah Bainer Jenkins, Co-Editor. Manuscripts should conform to the style and format described in the *Publication Manual of the American Psychological Association, 5th edition*. All manuscripts should be typed, double-spaced, and on 8½ x 11 paper with 1½ inch margins on all sides. An abstract of less than 100 words should accompany the manuscript. The author's name, contact information, and affiliation should appear on the title page only. Submissions typically are less than 20 pages in length. A disk file (3½ inch diskette, MS Word) is also required with the submission.

Electronic Submission. Submit the manuscript to Deborah Bainer Jenkins, Co-Editor, at mer@westga.edu as an e-mail attachment. Indicate in the subject line that this is a MWERJ manuscript. As with hard copy, the manuscript should conform to APA style, be produced in MS Word, and be limited to 20 pages, including abstract and references, and contain full contact information for the author(s).

All manuscripts, whether submitted in hard copy or electronically, will be acknowledged upon receipt. Please note that authors are responsible to submit manuscripts that are free of grammatical and mechanical errors. The editors reserve the right to make minor modifications in order to produce a more concise and clear article.

Questions regarding the journal or the submission of feature columns should be directed to the co-editors listed below.

Deborah Bainer Jenkins, Ph.D.
Professor and Director
University of West Georgia
228 Education Annex
Carrollton, GA 30118
(678) 839-6078
mer@westga.edu

Adrian Rodgers, Ph.D.
Assistant Professor
The Ohio State University
1179 University Drive
Newark, OH 43055
(740) 366-9261
rodgers.50@osu.edu

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, Donna Evans, Dean.
POSTMASTER: Send address change to Jean W. Pierce, Dept. EPCSE, Northern Illinois University, DeKalb, IL 60115.

MID-WESTERN EDUCATIONAL RESEARCHER

Volume 18, Number 4 · Fall 2005 ISSN 1056-3997

MWER Publication Address

Deborah Bainer Jenkins
University of West Georgia
228 Education Annex
Carrollton, GA 30118
Phone: (678) 839-6078
e-mail: mer@westga.edu

MWER Subscription & MWERA Membership Information

Jean W. Pierce
LEPF Department
Northern Illinois University
DeKalb, IL 60115
Phone: (815) 753-8470
Fax: (815) 753-8750
e-mail: jpierce@niu.edu

A Resource Inputs Model Examining NAEP Reading Score Change from 1994-1998: State Median Household Income and Free and Reduced-Price Lunch Percentages Matter[1]

David A. Walker
Northern Illinois University

Shereeza F. Mohammed
Florida Atlantic University

Abstract

This analysis indicated that from variables theorized to influence score change in NAEP reading scores from 1994-1998, two were the most consistent with the pattern of correlations found in the data. Together, both median household income (AVGINC) and the percentage of students eligible for free and reduced-price lunch (FRELCH) had fairly large impacts on reading score change. Additionally, the rate of change in reading scores varied across states. AVGINC had constant growth over time, while FRELCH's development over time was quite slow in both the initial model and an alternative model.

The intent of this research was to use the National Assessment of Educational Progress (NAEP) data to examine score changes in reading from 4th to 8th grade from 1994-1998. To help parcel out the sources of variance, a model based on state and federal-level resource inputs provided estimates of the magnitude of hypothesized connections between sets of variables indicating score change in NAEP reading from 1994 to 1998. As separate growth curve models of change of the group as a whole, a more in-depth analysis was examined of the final variables that were chosen for the regression model.

Review of the Literature

Achievement testing is not a new idea in the American educational system. In fact, standardized achievement testing has been a part of the educational system since the mid-1800s, with extensive testing of students starting in the 1920s in association with the administration of the Stanford Achievement Test (Haladyna, 2002). However, the notion of a single, national assessment is relatively new in the United States. Since 1969, NAEP has assessed the condition of learning in particular subject areas and in specific grades and/or ages. As a congressionally decreed testing scheme, NAEP is known as the Nation's report card in the sense that it indicates student achievement score trends in academic areas such as mathematics, reading, science, and history via a nationally representative sample (National Center for Education Statistics (NCES), 2002). Since NAEP is a national assessment, and acknowledged as a dependable measure of average student performance (Hedges & Greenwald, 1996), states can use testing results as a benchmark to compare their score gains against national score gains, for example in reading from the 4th to 8th grade, to determine acceptable levels of achievement growth.

Numerous factors, often beyond the control of schools and classrooms, influence student achievement. Examples of these "external" conditions include variables such as parental education, poverty, mobility, Limited English Proficiency (LEP), and community type (Chavkin, 1996; Coleman et al., 1966; Coleman, 1987; Schuler, 1990). In fact, investigating these factors appears to be the movement in student achievement research (Steinberg, 1996; Wenglinsky, 2002). Of the various factors that influence student learning outside of the classroom, low socio-economic status (SES) may have one of the stronger relationships with student achievement on standardized tests (Educational Testing Service, 1980; Ferguson, 1991).

The scope of this research involved independent and dependent variables that were measured at the state and/or federal level and often outside the control domain of schools and classrooms. The numerous independent variables considered for the study were categorized as "resource inputs," usually derived from state and/or federal sources. Some of these resource inputs have been noted for their positive, albeit at times modest, relationship with student test score outcomes (Card & Krueger, 1996; Ferguson, 1991; Hedges & Greenwald, 1996; Hedges, Laine, & Greenwald, 1994; Wenglinsky, 1997).

Among the major economic policy objectives of the federal government are economic growth, full employment, stable prices, balancing the fund flow into and out of the economy, competitiveness with other governments, budget deficits, and overall debt (Lee & Johnson, 1998). Federal

This paper is based upon work supported by the National Center for Education Statistics through a fellowship grant awarded to the lead author to participate in the 2002 NAEP Analysis Institute.

[1] This article was accepted for publication by the previous editorial team.

government policy involves economic stabilization to "maintain a high level of resource utilization and a stable value of money" (Musgrave, 1959, p. 22). The goal of stabilization is to maintain price-level stability and full employment, while using public policy instead of market forces to influence the rate of growth (Musgrave, 1959).

Both fiscal and monetary policies are used to affect the aforementioned variables in an attempt to impact macroeconomic performance. Hence, regulating revenues and expenditures can result in changes in employment, inflation, economic dampening, or stimulation. Similarly, monetary policies, which change interest rates along with the Federal Reserve, influence the pace of economic activity (Lee & Johnson, 1998).

Unlike the federal government, state and local governments operate by providing the basic infrastructure for the private sector to attract business activity to enhance economic performance. This differential of purpose results from the vertical imbalance of generating revenue among the three levels of government. While the federal government has the greatest revenue capacity derived from personal and corporate income taxes, state and local governments rely on sales and wealth taxes along with user fees. Further, states experience horizontal fiscal differences "in income and wealth [that] lead to differences in revenue-generating abilities, tax burdens and levels of public services" (Lee & Johnson, 1998, p. 433). Consequently, the federal government redistributes revenue to states and local governments in the form of transfer programs, grants, tax expenditures, and infrastructural investment.

However, with such a complex economy, changes in stabilization can affect redistribution. For instance, in times of high inflation, those with fixed incomes lose purchasing power and income is redistributed to the wealthy that benefit from high interest investments. In fact, since the 1960s, and especially in the 1980s, the cutting of corporate and personal income taxes has resulted in an increasing degree of income inequality. A number of income stabilization and redistributive policies exist in terms of aid or transfer programs to less fortunate individuals and changes in the tax structure to redistribute income to groups in need (Lee & Johnson, 1998). However, these must work in concert with other factors such as creating a skilled and knowledge-based workforce. Hence, expenditures on both academic and vocational education would increase productivity and international competitiveness. Therefore, while the federal government concentrates its state aid on areas of education, income security, health and transportation; state education aid to local government is larger and is allocated on a formula basis (Lee & Johnson, 1998).

Previous studies by Coleman et al. (1966) and Hanushek (1986) found that student performance was unrelated to school expenditures. However, the idea that strategically targeted funding can impact student achievement has been promoted by authors such as Hill, Cohen, and Moffitt (1999) and Ladd and Hansen (1999). In fact, two separate literature reviews (Hedges et al., 1994; Krueger, 1999) who used the same studies as Hanushek (1986) "supported a positive relation between resources and outcomes" (Grissmer, Flanagan, Kawata, & Williamson, 2000, p. 27). Further, Greenwald, Hedges, and Laine (1996) had similar findings when they conducted a comprehensive review of the literature using meta-analytic techniques. The resources in each case related to various measures of per pupil expenditures.

Moreover, in a recent RAND publication, Grissmer et al. (2000) investigated the effect of both family capital and social capital factors on NAEP score change between 1990 and 1996. They confirmed earlier findings, which indicated that with other factors being equal, higher per-pupil expenditures among other classroom related factors showed "positive, statistically significant effects on achievement" (p. 98). Additionally, they affirmed several positions. First, states varied significantly in the manner by which they allocated per pupil expenditures. Secondly, the wide variation between and within states were due to the different profiles of SES of their students. Third, the score gains among minorities and the disadvantaged in the 1990s were attributed to the moderate, additional resources in the 1970s and 1980s. In fact, they held that these resources were "most effective and efficient when spent in states with higher proportions of minority and disadvantaged students" (Grissmer et al., 2000, p. 102). Moreover, their analysis pointed to "score gains of 12 to15 percentile points from additional targeted expenditures of less than $1000 dollars a pupil in states with the lowest SES" (Grissmer et al., 2000, p. 101). Hence, the literature indicates that targeted expenditure can impact positively on score gains.

Methods

Research Questions

The research questions were:

1. Which resource inputs were the most influential in determining score gains on NAEP reading from 1994 to 1998?
2. Did the rate of change in reading scores vary across states? If so, did this deviation relate to the resource inputs that were chosen as the most influential?
3. Did the initial models proposed for the resource input variables correspond strongly to the data or were there alternative models that coincided stronger with the data?

To assist in answering these research questions, various types of modeling will be used in this study. Often, the gist of many models, including this study's models, is to allow for a good assessment of the structure of the variables chosen and the degree of the correspondence between them both within preliminary and final models. As well, modeling contributes theoretically by assisting researchers in the determination of simpler or parsimonious models. Also, modeling in this study is used to increase cross-validation accuracy,

examine change trajectories over time for a particular outcome, and determine which of two competing models fit the data better for a specific situation.

Variable Selection and Diagnostics

Previous research has used many of the subsequent variables in studies that looked at the relationship between measures of resource inputs and student achievement (cf. Chubb & Hanushek, 1990; Ferguson, 1991; Hanushek, 1996; Hedges et al., 1994). Initially, in this theorized model, there were 17 independent variables per the 30 states in the study that had both NAEP reading scores from 1994 and 1998: state revenue, instructional expenditures, support services expenditures, non-instructional expenditures, per pupil total expenditures, per pupil instructional expenditures, per pupil support services expenditures, per pupil non-instructional expenditures, percentage of student with IEPs (Individual Education Plan), number of Title I schools, number of migrant students, percentage of students receiving LEP services, Title VII state grant program funding for LEP students, state taxes per $1000 personal income, property taxes per capita, percentage of students eligible for free and reduced-price lunch, and state median household income.

Upon initial diagnostic review to find the most parsimonious model and estimate the parameters of the regression model, dimensionality reduction techniques were implemented via use of SPSS (Statistical Package for the Social Sciences v. 12.0). Areas such as VIF (variance inflation factor) and tolerance were observed. Of the initial 17 instructionally-related and non-instructional variables considered, 11 were discarded from the original model due to very high VIF values > 10, which is an indicator of collinearity that can cause problems with the interpretation of regression weights and the stability of regression weights from sample to sample. Thus, six variables were retained as components of a parsimonious and more credible model in terms of the population (i.e., per pupil non-instructional expenditures, percentage of students with IEPs, percentage of students receiving LEP services, state taxes per $1000 personal income, percentage of students eligible for free and reduced-price lunch, and state median household income).

To determine variable selection for the final model, the 6 variables were entered into a best subsets regression model, written in SPSS syntax, which meant that there were 63 possible subset analyses (i.e., $2^6 - 1$). The best subsets analysis examined multiple, comprehensive variable selection in three areas: the best fit of *n* models based on the maximum adjusted R^2 (max. adj. R^2) or variance explained, the lowest root mean square error (RMSE) or lack of fit, and the lowest Mallow's Cp or the smallest amount of total square errors. Of the 63 regressions conducted, the model containing the percentage of students eligible for free and reduced-price lunch per state (FRELCH) and median household income per state (AVGINC) was chosen because it consistently bested the other models in the three areas of variable selection, where max. adj. $R^2 = .219$, RMSE = 3.364, and Mallow's Cp = 64.825. That is, free and reduced-price lunch (FRELCH) and median household income (AVGINC) were deemed variables essential and ubiquitous in terms of the best equation fitted to the data. For the final variables of study, FRELCH data came from the NCES and AVGINC came from the U.S. Census Bureau. Of note, is the fact that all of the states in the study had access to both of these variables. Finally, the dependent variable was the difference in state reading scores from 1994 to 1998, which came from the NCES.

Final Variables of Study

In an examination of median household income, Camilli (2000) illustrated that wealth "is one of the most reliable predictors of achievement" (n.p.). Derryberry confirmed this finding in Washington, where scores on the state test, the WASL (Washington Assessment of Student Learning), varied directly with family income (as cited by Abe & Sullivan, 2002). Similarly, Sauer's (2003) analysis of Oregon's report card showed a significant difference in the poverty level between high-performing and low-performing schools. These findings corroborate and show little change after 30 years since the study conducted by Jencks et al. (1972) in which it was estimated "that a family's economic status probably correlates about 0.35 with the children's test scores" (p. 78). Furthermore, a frequently used indicator of poverty is the number of students participating in free and reduced-priced lunch programs, which the literature reveals, tends to have a negative influence on achieving higher gain scores in many academic subjects (NCES, 1998; NCES, 2003a).

Data

Using EDA (Exploratory Data Analysis), boxplots, histograms, normal-probability plots, and numerical data for kurtosis and skewness were observed to look at possible anomalies and curvilinearity within the data. Outliers were reviewed via z standardized values of the data and none of the data were found to have z scores smaller than –2.5 or larger than +2.5. The variables of study were found to have approximately normal, bell-shaped distributions. Further, the data had homogeneous variances. Both AVGINC and FRELCH were averaged over three years to control for fluctuations and reduce measurement error. In terms of missing data, there were only a few cases for the variable FRELCH. Thus, the series mean for this variable was imputed to correct for this issue. A maximum likelihood method was used in the growth curve analyses for the parameters in the model to counter against heteroscedasticity, which can yield biased estimates of slope and intercept coefficients. The study's sample size, n = 30 states, appeared small for the proposed model. However, using Green's (1991) formula to determine an appropriate sample size, showed that an n = 21 was suitable in this case.

$(8 / F^2) + (I - 1)$ (1)

$F^2 = R^2 / (1 - R^2)$

I = number of predictors

Where, $F^2 = .282 / .718 = .39275766$

$= 8 / .39275766 = 20.37 + (2 - 1) = 21.37.$

Results

Regression Analysis

Using the software AMOS (Analysis of Moment Structures) (Arbuckle, 1999), the final regression model was created. To obtain 90% lower and upper confidence intervals for the relationships between the two predictors and the criterion, based on 5000 bootstrapped samples, indications were that estimation stability for the distribution of the *r*-related statistics had been achieved and were strong and, ultimately, more likely to be stable upon replication. An analysis of this model indicated that the partial regression coefficients were somewhat valuable at explicating the variance in reading score change (Adj. $R^2 = .219$; $R^2 = .273$; 90% CI $R^2 = .073, .555$; M = .299, SE = .147). An examination of the coefficients showed that AVGINC (Beta = .451; 90% CI = .113, .757; p =.035; M = .442, SE = .195) and FRELCH (Beta = .658; 90% CI = .322, .920; p =.007; M = .646, SE = .187) both had effects on score change, with FRELCH having the stronger effect.

Regression Cross-Validation

To measure how well the sample regression equation would perform when predicting future observations from the same population, cross-validation was conducted. The total sample was divided into equal size sub-samples of 15, where odd-numbered cases were assigned a value of 1 and even-numbered cases were assigned a value of 2. The root mean square error (RMSE) for the calibration sample was 3.954 and for the validation sample was 3.105, which are internal measures of performance that indicate how well the equation predicts reading score change. To estimate the prediction error or the future, the prediction sum of squares (PRESS) was calculated via a SPSS syntax program to specify the error in the model in terms of predictive variance. The PRESS value for the validation sample was 154.014. The mean squared error of prediction (MSEP), or the mean value of PRESS, was determined and used as a summary measure of predictive ability. The validation sample had an MSEP of 10.268. Finally, the root mean squared error of prediction (RMSEP) for the validation sample was 3.204, which is an external measure of predictive performance and was very similar to its RMSE value noted above, and nearly 1 NAEP point less of error on the dependent variable than the calibration sample's RMSE of 3.954. Thus, the validation sample's RMSEP indicated that the final model chosen, containing free and reduced-price lunch (FRELCH) and median household income (AVGINC), should be quite accurate in the future when using the equation for the straight line selected to predict NAEP score changes in reading from 4th to 8th grade from 1994-1998. That is, the cross-validation results showed that error could occur in this prediction by approximately three NAEP reading points.

Growth Curve Models: FRELCH

Growth curve modeling examines the change trajectories over time of an outcome and can present this change as either individual or group movement so that an estimation of future performance may be determined. For this study, change will be represented as:

$Y_{it} = \pi_{0i} + \pi_{1i} x_{it} + \varepsilon_{it}$

$\pi_{0i} = \beta_{00} + u_{0i}$

$\pi_{1i} = \beta_{10} + u_{1i}$

Y_{it} = dependent variable measured for a participant at a specific time

π_{0i} = intercept = average Y at t = 0

π_{1i} = slope = linear rate of change

x_{it} = time for a linear curve at one-year intervals, where t_1, t_2, t_3, etc. = 0, 1, 2... for participant i

ε_{it} = error or disturbance term for participant i at time t

Intercept and Slope

To look at the two variables' change over time, this research used slope and intercept parameters of group growth trajectories across the sample. Each state's intercept and slope were combined together and a group slope and intercept, with accompanying variance measures, were analyzed, along with a total error variance at each of the points in time for the two variables. For FRELCH, the first time point's slope parameter was set to 0 and the last time point's slope was set to 1.00. This allowed the researchers to conceptualize growth as being 0% complete at time one and 100% complete at time six. Therefore, the first time point was treated as the starting point for the growth curve. There were six time points in the analysis.

The mean intercept, which is the average initial value across states of the percentage of students eligible for free and reduced-price lunch was 30.57 or about 31% and was statistically significant from zero. The variance of the intercept, which was 10.39 or about 10%, indicated the variability of the average value across states and was also statistically significant. The mean slope, which is the average change per year, was 3.17 or about 3% and was statistically significant. Because the variance of the intercept was statistically significant, this indicated that there may be consequential variation in the percentage of free and reduced-price lunch students by state during this time frame. These data indicated that there were likely interindividual differences in that some states had increased percentages eligible for free and reduced-price lunch, while other states had decreases in this area.

With these interindividual differences aside, the percentage of students eligible for free and reduced-price lunch appeared to stagnate over time as evidenced by the small mean slope of 3.17. As noted in figure 1, the parameter estimates for the observed slope-variable regression coefficients indicated slow growth also when compared to expected values: time one was fixed, time two = .05 (.20 expected value), time three = .15 (.40), time four = .64 (.60), time five = .57 (.80), and time six was fixed. Looking at the observed values indicated slow growth through the first three time periods, with acceleration at time four, and then a decrease at time five. The point of inflexion, or maximum point of growth, was between 1995 and 1996.

When examining the correlation between the intercept and slope (–.24), its non-statistically significant, negative relationship indicated the paucity between the trajectories of FRELCH and score change, where the initial quantity of FRELCH was not associated with change over time. Thus, there does not appear to be much meaningful change in the percentage of students eligible for free and reduced-price lunch. In this scenario, it seems to be a variable that should be coupled with other variables in order to be used as an estimator of future performance. To illustrate, Jencks et al. (1972) stated that evaluations done on Title I programs produced random results so that the students served may or may not perform as expected on tests. In two related studies, in California and West Virginia, Friedkin and Necochea (1988) and Howley (1995) found that small school size assisted in reducing the negative effects on achievement of students in the free and reduced-price lunch program.

Model Fit

In terms of the goodness-of-fit of the model to the sample data, large values of chi square (χ^2) mean that the model is a bad fit for the data, where the implied and sample covariance matrices differ greatly and do not approximate the population discrepancy well. Small values of χ^2 signify that the data is a good fit. This model's statistically significant chi-square test ($\chi^2 = 54.781$ (12), $p < .001$) indicated initially that this model was possibly not an appropriate fit for the data and there was evidence not to accept the null hypothesis that states shared the same variances and covariances. Yet, the use of only the χ^2 statistic as a measure of fit may render uncertainty concerning the overall appropriateness of the study's model. As advocated by Tanaka (1993), various indicators of fit were utilized beyond the χ^2 criterion of fit or no fit. As relative fit measures, the incremental fix index (IFI = .943), the comparative fit index (CFI = .942), and the normed fit index (NFI = .928) all indicated that the proposed model compared to, but did not exceed, a null model per the cut point fixed at $\geq$.95.

A second model was created and tested against the above model via a nested chi-square test, where parameter constraints were the expected values for the linear constraints (i.e., .20, .40, .60, .80, and 1.00 for times one to five, with time one calculated as 1/5, time two as 2/5, etc.). This second model yielded ($\chi^2 = 96.476$ (16), $p < .001$), IFI = .892, CFI = .892, and the NFI = .874. Thus, the first model fit the data better then the second, rival model. However, neither model presented a very small χ^2 value or non-statistical significance and did not contribute evidence that a significant rate of change was existent in the population pertaining to this variable.

Growth Curve Models: AVGINC

For median household income (AVGINC), it appeared as though this variable's ability within the model was due to the fact that over a five-year period from 1993, 1995-1998 (note: per the U.S. Census Bureau, 1994 data is not available), the median household income by state had increased, which the literature says should have a positive effect on achieving higher gain scores. Camilli (2000) illustrated this by averaging median state household income between 1995

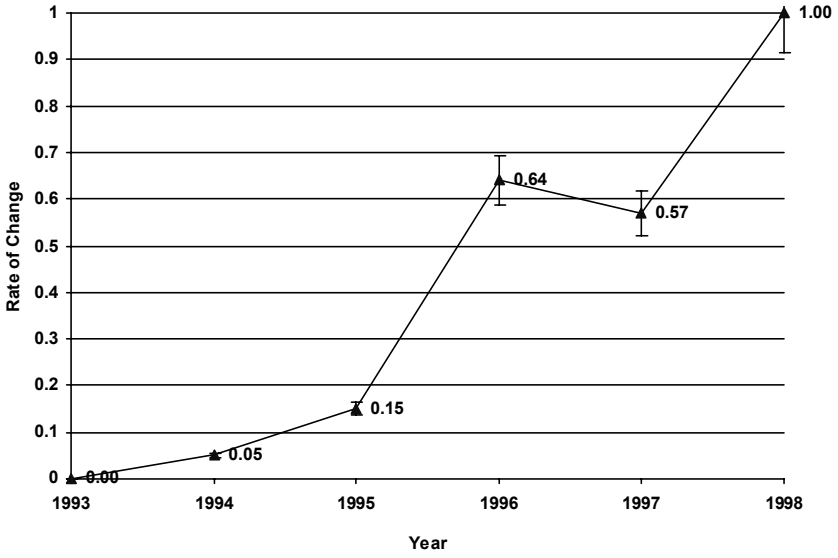

Figure 1. Change in Free and Reduced-Price Lunch Variable. *Note*: Bars reflect the standard error for each data point.

and 1997 and plotting them against NAEP cohort math growth scores from 4th to 8th grade. Further, *The Condition of Education 2003* (NCES, 2003a) stated that poverty related to annual income presented serious challenges to student achievement and that 16% of children lived in households that were below the poverty level. The poor and non-poor threshold levels are determined annually using the Consumer Price Index (CPI) household income. In fact, Jencks et al. (1972) related that school reform could not single-handedly negate the effects of the income gap in society.

Intercept and Slope

There were five time points in the analysis. The mean intercept was $30,723 and statistically significant. The variance of the intercept of $4,888 was also statistically significant. The mean slope was $7,120 and statistically significant. Because the variance of the intercept was statistically significant, this indicated that there was heterogeneity and interindividual differences existed over time within the sample, where some states had increased levels of median household income and some had decreased levels.

The amount of median household income appeared to increase over time as indicated by the mean slope of $7,120. Analysis of the estimated values indicated an appropriate growth curve through the time periods for median household income, with an accelerated growth component at time point one and normalcy for the rest of the time points. That is, for constant growth across time, one would expect a value near .50 at time point three (i.e., the half-way point), which was found (observed = .54). As can be seen in figure 2, the parameter estimates for the observed slope-variable regression coefficients indicated constant growth when compared to expected values: time one was fixed, time two = .35 (.25 expected value), time three = .54 (.50), time four = .75 (.75), and time five was fixed. The time points all provided a good approximation of a constant change function. Finally, when looking at the correlation between the intercept and slope (.16), its relationship indicated a more prominent extent between the trajectories of AVGINC and score change in comparison to FRELCH and score change. The positive relationship between the initial value of AVGINC and growth over time appeared to be somewhat related. AVGINC, in this situation, appears to be a fair estimator of future performance.

In addition, the effects of inflation and how it compared to the rates of change for median household income (AVGINC) were regarded. When factoring in the Consumer Price Index (CPI) and the Higher Education Price Index (HEPI) from 1993, 1995-1998, both of which are considered inflationary measures of changes in prices over time, AVGINC's rate of change nevertheless outpaced these inflation indices in every instance (American Library Association, 2002; University of San Francisco, 2001; U.S. Census Bureau, 1999). That is, over the years of study, the CPI average was 2.20% and the HEPI average was 3.08%, while the rate of change for AVGINC was 8.30% in 1995, 4.10% in 1996, 4.20% in 1997, and 4.90% in 1998. The average rate of change for AVGINC was 5.38% or twice as much as the CPI and nearly that amount for the HEPI.

Model Fit

The results indicated that the rate of change in median household income was constant growth across time. This model fit the data well with a non-statistically significant chi-square test ($\chi^2 = 11.108$ (7), $p = .134$), IFI = .996, CFI = .996, and the NFI = .989. Further, a second, more complex model was created and tested against the above model, where parameter constraints were the expected values for the linear constraints (i.e., .25, .50, .75, and 1.00 for times one to four). This second model yielded ($\chi^2 = 30.633$ (10), $p < .001$), IFI = .978, CFI = .978, and the NFI = .968. Looking at a model comparison, the difference between the two models' χ^2 is 19.525, with 3 degrees of freedom, and $p = .001$.

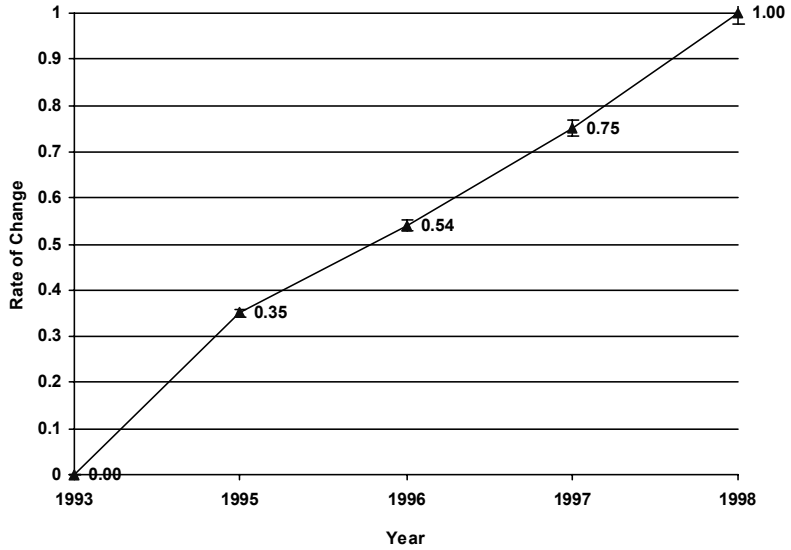

Figure 2. Change in Median Household Income Variable

That is, the probability of having a χ^2 this large is .001, which is very remote. Thus, the first, simpler model fit the data much better then the second model, which indicated that a significant rate of change was existent in the population pertaining to AVGINC.

Limitations

Obviously, the current model only looked at variables considered to be of the ilk of state and federal-level resource inputs. There may be present some aggregation bias, such as certain state characteristics that were not controlled, or omitted variable bias that may influence results (cf. Betts, 1996). Further, one could argue that this model did not encompass enough assorted variables that influence student achievement. However, a comprehensive model was not the purpose of the research. A more discrete, resources-based model was the intent. Further, some may contend that an initial model comprised of 17 independent variables, when compared to the final model consisting of two of the original variables, is objectionable. However, as was noted previously, a parsimonious model was a goal of the research and, thus, many of these competing independent variables did not add to the explication power for the variance in the dependent variable. Instead, their inclusion in early models only caused some parameters to be unidentified. The final model used observational data aggregated at the state-level, which does contain caution, for example, with a variable such as FRELCH that may yield different consequences at the state and school level.

Implications and Conclusions

This analysis indicated that from 17 variables theorized to influence score change in NAEP reading scores from 1994-1998, two were the most consistent with the pattern of correlations found in the data. To answer the previously-posed research questions, it was found that, together, both median household income and the percentage of students eligible for free and reduced-price lunch had fairly large impacts on reading score change. As distinct variables, AVGINC appeared to be the better estimator of score change. This research was able to ascertain the effects of free and reduced-price lunch (FRELCH) and median household income (AVGINC) on score change in NAEP reading from 1994-1998. The rate of change in reading scores did vary across states and the research found that AVGINC had constant growth over time, while FRELCH's development over time was quite languid in both the initial model and an alternative model and did not appear to be an ideal estimator of future performance when examined in isolation, but was quite robust when coupled with AVGINC.

When the NAEP 1994 and 1998 reading scores for low-performing 4th graders were examined, it appeared that students in the 10th percentile increased by eight points, while those in the 25th percentile rose four points. However, higher performing students at the 50th and 75th percentile remained unchanged (NCES, 2000). Similarly, scores at the 10th, 25th, and 50th percentiles were higher in 2002 than in 1998 and 2000. In fact, average scores increased between 1998 and 2002 for 4th graders eligible for free and reduced-price lunch (Grigg, Daane, Jin, and Campbell, 2003). Correspondingly, since 1990, NAEP mathematics performance for 4th, 8th, and 12th grades showed a direct and inverse relationship between scores and the percentage of students eligible for free and reduced-price lunch. Such findings re-appeared in the 1994 and 2001 NAEP geography and U.S. history assessments (NCES, 2001). These patterns seem to imply that many groups who have been served by Title I programs, such as free and reduced-price lunch, showed incremental gains in achievement. In a study conducted by RAND's Institute on Education and Training, similar gains in student performance between 1970 and 1990 can be attributed to "public investment in schools and families and equal educational opportunity policies" (Grissmer, Kirby, Berends, & Williamson, 1994, n.p.).

It is indeed interesting that out of the 17 independent variables investigated, that 15 appear unrelated to score gains. To understand this better, it is necessary to look at the degree of representative sampling done during the administration of NAEP tests. Participating in the NAEP is voluntary and when schools opt out, administrators use NCES guidelines to substitute other schools in their place. These guidelines have been created to remove any bias caused by the substitution. However, in a recent RAND study, Grissmer et al. (2000) found that participation is higher in states with higher SES and that "schools choosing not to take the tests are probably more often schools with lower-scoring students" (p.150). This is noteworthy since many low performing schools who serve students on free and reduced-price lunch are more focused on achievement of their state tests and may be more disinclined to participate in NAEP.

Further, Grissmer et al. (2000) suggested that students from more resource-rich backgrounds are relatively unaffected by the level of school resources as are those from impoverished homes. This implies that student scores from high SES areas can be masking the effects of the per pupil expenditures in poorer areas. This could explain why the model in this study did not support stronger relationships with variables such as: instructional expenditures, support services expenditures and the per pupil equivalent of each, as well as the number of Title I schools and migrant students.

To improve NAEP data, Grissmer et al. (2000) suggested changing NAEP administration to school districts and using census data to distinguish between urban and suburban trends. This would then produce a more representative sample of income levels, which would be more reflective of AVGINC and FRELCH and possibly detect underlying relationships of per pupil expenditures now masked by the overwhelming presence of AVGINC and FRELCH.

Another factor that may explain why variables relating to limited English proficiency (LEP) and students on individualized education plans (IEP) did not enter our model is that certain weighted percentages of these students are excluded from NAEP assessments. As the Grissmer et al. (2000) study indicated, the IEP percentages varied "from 2.6 in North Dakota to 6.7 in Florida" and for LEP they ranged from "0.0 percent in Wyoming to 8.3 percent in California" (p.13). Moreover, if NAEP is to be a more precise national report card and act as an instrument to direct policy and efficient funding, then it would be effective to use a more representative sample at the district level. This may improve the models that can be constructed to see the full effects of using per pupil expenditures and of directing them more strategically.

Under the NCLB (No Child Left Behind) Act of 2001, states must participate in NAEP reading and mathematics assessments at grades 4 and 8 every two years or risk losing Title I funding (NCES, 2003b). The underlying purpose is for NAEP to confirm the annual yearly progress, which states have selected as targets, to the ultimate goal of having all students become proficient in specified subject areas by 2014. Hence, it is hoped that NAEP will serve as an independent measure to inform education policy in its data collection and, thus, assist educators and legislators in attaining this goal (National Education Association, 2002). This is particularly important in times of budgetary restraints when maximizing achievement will require funding for those variables that can overcome the negative effects of income and tend to affect positively students enrolled in free and reduced-priced lunch programs.

Ultimately, if a policy goal is to review the U.S. educational system in the area of reading to conclude which variables enable educators to determine what students from 4th to 8th grade are able to do and the score gains that may have occurred, then AVGINC and FRELCH collectively should be regarded as indicators of this objective.

Acknowledgement

The lead author would like to thank Janet Holt for her assistance with the equation pertaining to growth curve modeling.

References

Abe, D., & Sullivan, P. (2002, December 8). Poverty vs. achievement: WASL test score analysis shows low-income students face greater hurdles. *The News Tribune.*

American Library Association. (2002). *U.S. periodical prices-2002*. Retrieved January 5, 2004, from http://www.ala.org/Content/NavigationMenu/Products_and_Publications/Periodicals/American_Libraries/Selected_articles/charta.htm

Arbuckle, J. L. (1999). AMOS (Version 4.01). [Computer software]. Chicago: SmallWaters Corporation.

Betts, J. R. (1996). Is there a link between school inputs and earning? Fresh scrutiny of an old literature. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 141-191). Washington, DC: Brookings Institute Press.

Camilli, G. (2000, August 21). Texas gains on NAEP: Points of light? *Education Policy Analysis Archives*, 8(42). Retrieved November 5, 2002, from http://epaa.asu.edu/epaa/v8n42.html

Card, D., & Krueger, A. B. (1996). Labor market effects of school quality: Theory and evidence. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 97-140). Washington, DC: Brookings Institute Press.

Chavkin, N. F. (1996). Involving immigrant families in their children's education: Challenges and opportunities for schools. In J. L. Flores (Ed.), *Children of la frontera: Binational efforts to serve Mexican migrant and immigrant students* (pp. 325-339). Charleston, WV: ERIC Clearinghouse on Rural Education and Small Schools.

Chubb, J. E., & Hanushek, E. A. (1990). Reforming educational reform. In H. J. Aaron (Ed.), *Setting national priorities: Policy for the nineties* (pp. 213-248). Washington, DC: Brookings Institute Press.

Coleman, J. S. (1987). Families and schools. *Educational Researcher, 16*, 32-38.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office. Educational Testing Service. (1980). *Test scores and family income: A response to charges in the Nader/Nairn report on ETS*. Princeton, NJ: Author.

Ferguson, R. F. (1991). Paying for public education: New evidence on how and why money matters. *Harvard Journal of Legislation, 28*(2), 465-498.

Friedkin, N., & Necochea, J. (1988). School system size and performance: A contingency perspective. *Educational Evaluation and Policy Analysis, 10*(3), 237-249.

Green, S. B. (1991). How many subjects does it take to do a regression analysis? *Multivariate Behavioral Research, 26*, 499-510.

Greenwald, R. L., Hedges, L. V., & Laine, R. (1996). The effect of school resources on student achievement. *Review of Educational Research, 66*(3) 361-396.

Grigg, W., Daane, C., Jin, Y., & Campbell, J. (2003, June). *The nation's report card: Reading 2002*. Retrieved June 6, 2003, from http://nces.ed.gov/nationsreportcard/pubs/main2002/2003521.asp

Grissmer, D. W., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP test scores tell us.* Retrieved December 1, 2004, from http://www.rand.org/publications/MR/MR924/

Grissmer, D., Kirby, S., Berends, M., & Williamson, S. (1994). *Student achievement and the changing American family.* Retrieved January 5, 2003, from http://www.rand.org/publications/RB?RB8009/

Haladyna, T. M. (2002). *Essentials of standardized achievement testing: Validity and accountability*. Boston: Allyn & Bacon.

Hanushek, E. A. (1986). The economics of schooling: Production and efficiency in public schools. In D. W. Grissmer, A. Flanagan, J. Kawata, & S. Williamson (Eds.), *Improving student achievement: What state NAEP test scores tell us.* Retrieved December 1, 2004, from http://www.rand.org/publications/MR/MR924/

Hanushek, E. A. (1996). School resources and student performance. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 43-73). Washington, DC: Brookings Institute Press.

Hedges, L. V., & Greenwald, R. (1996). Have times changed? The relation between school resources and student performance. In G. Burtless (Ed.), *Does money matter? The effect of school resources on student achievement and adult success* (pp. 74-92). Washington, DC: Brookings Institute Press.

Hedges, L. V., Laine, R. D., & Greenwald, R. (1994). Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational Researcher, 23*(3), 5-14.

Hill, H., Cohen, D., & Moffitt, S. (1999). Instruction, poverty and performance. In G. Orfield & E. Debray (Eds.), *Hard work for good schools: Facts not fads in Title I reform* (pp. 55-76). Cambridge, MA: Harvard University.

Howley, C. (1995). The Mathew principle: A West Virginia replication? *Education Policy Analysis Archives, 3*(18). Retrieved June 3, 2003, from http://epaa.asu.edu/epaa/v3n18.html

Jencks, C., Smith, M, Ackland, H. Bane, M. J., Cohen, D., Gintis, H., Heyns, B., & Michelson, S. (1972). *Inequality: A reassessment of the effect of family and schooling in America.* New York: Basic Books.

Krueger, A. B. (1999). An economist view of class size reductions. In D.W.Grissmer, A. Flanagan, J. Kawata, & S. Williamson (Eds.), *Improving student achievement: What state NAEP test scores tell us.* Retrieved December 1, 2004, from http://www.rand.org/publications/MR/MR924/

Ladd, H., & Hansen, J. (Eds.). (1999). *Making money matter: Financing America's schools.* Washington, DC: National Academy Press.

Lee, R., Jr., & Johnson, R. (1998). *Public budgeting systems* (6th ed.). Gaithersburg, MD: Aspen Publication.

Musgrave, R. (1959). *The theory of public finance: A study in public economy.* New York: McGraw-Hill.

National Center for Education Statistics. (1998). *NAEP 1998 reading report card for the nation and the states.* Retrieved May 28, 2003, from http://nces.ed.gov/nationsreportcard//pubs/main1998/1999500.asp

National Center for Education Statistics. (2000). 1998 reading results for low-performing students. *NAEPFACTS, 4*(4), 1-4.

National Center for Education Statistics. (2001). *Lower scores for students on lunch program*. Retrieved May 28, 2003, from http://nces.ed.gov/nationsreportcard/reading/lunchprogram.asp

National Center for Education Statistics. (2002). *Overview*. Retrieved March 15, 2002, from http://nces.ed.gov/nationsreportcard/about

National Center for Education Statistics. (2003a). *The condition of education 2003*. Washington, DC: Author. (NCES 2003-067)

National Center for Education Statistics. (2003b). *Important aspects of No Child Left Behind relevant to NAEP*. Retrieved May 28, 2003, from http://nces.ed.gov/nationsreportcard/nclb.asp

National Education Association. (2002). *NAEP and ESEA testing: Confirming state test results*. Retrieved January 14, 2003, from http://www.nea.org/accountability/naep-accountability.html

Sauer, R. (2003, May 21). Jencks reassessed, one career later. *Education Week*, pp. 32, 34.

Schuler, D. B. (1990). Effects of family mobility on student achievement. *ERS Spectrum, 8*, 17-24.

Steinberg, L. D. (1996). *Beyond the classroom: Why school reform has failed and what parents need to do*. New York: Simon & Schuster.

Tanaka, J. S. (1993). Multifaceted conceptions of fit in structural equation models. In J. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 10-39). Newbury Park, CA: Sage.

University of San Francisco. (2001). *2001 inflation report*. Retrieved March 22, 2004, from http://www.usfca.edu/budget/EIAReports/cpi2001.html

U.S. Census Bureau. (1999). *Statistical abstract of the United States*. Retrieved January 5, 2004, from http://www.census.gov/prod/www/statistical-abstract-us.html

Wenglinsky, H. (1997). *When money matters: How educational expenditures improve student performance and how they don't.* Princeton, NJ: ETS.

Wenglinsky, H. (2002, February 13). How schools matter: The link between teacher classroom practices and student academic performance. *Analysis Archives, 10*(12). Retrieved August 2, 2003, from http://epaa.asu.edu/epaa/v10n12

Small Steps: Teacher Change in a Reform Mathematics Curriculum

Tom Fox
University of Houston-Clear Lake

Abstract

An experienced mathematics teacher in a rural high school in the United States was studied as she first implemented a reform calculus curriculum. Her instruction was compared to her previous practice using a curriculum not constructed with the reform movement in mind. Despite her early attempts to make her instruction more student-centered, the teachers' actions in the classroom were similar in many ways; demonstrating solutions to prototype problems was an important instructional strategy that she used in both curricula. However, the teacher showed an increased focus on conceptual knowledge and on the use of graphing calculator technology. The reform text was an influential part of this teacher's change process.

Problem and Its Background

Recently faculty at many universities have examined their calculus curriculum and the ways in which it is taught. To address concerns with calculus courses and their instruction, reform calculus curricula were developed, and their large-scale implementation began at various universities across the United States in the early 1990s. Currently these curricula are used in many secondary schools. They are only one component of the efforts to change mathematics education in secondary schools.

Classroom teachers are those who ultimately grapple with the changes advocated by reform movements. The changes advocated are not easily accomplished, and high school mathematics teachers can be among the most hesitant to modify their instructional practices (Wasley, Domoyer, & Maxwell, 1995). Knowledge of the process of mathematics teacher change can help inform those contemplating changes as well as those hoping to facilitate such changes.

The literature contains little research on high school mathematics teacher change, particularly regarding the implementation of a reform calculus curriculum at the high school level. Ferrini-Mundy and Graham believe that the calculus reform movement provides opportunities for "rich and interesting" (Ferrini-Mundy, & Graham, 1991, p. 633) research. They specifically called for research relating to teachers' use of reform calculus texts.

Goals and Conceptual Framework

The reform calculus movement had its roots in the 1980s and reflected a general dissatisfaction with the calculus course of its time. Among its goals are an increased focus on conceptual knowledge and use of more teacher strategies than just lecture. In general, the reform calculus curricula encourage the following:

(a) incorporation of technology;

(b) group work or collaborative learning;

(c) projects involving real-world applications;

(d) representation of concepts from numerical, algebraic, and geometric perspectives;

(e) verbal and written communication; and

(f) a focus on conceptual knowledge (Ross, 1996).

Of course, various curricula place different emphasis on each of these ideas. The text used by the teacher in this study, *Calculus* (Hughes-Hallett, et al., 1994) (to be referred to as RC, denoting 'Reform Calculus'), aspires to these goals. It is one of many reform calculus texts commercially available and is often considered a moderate attempt at reform.

My first goal was to compare one teacher's practice during her first-year implementing the RC relative to her practice using a text that is generally not considered to be reform in nature. Previously, this teacher used an edition of a calculus text authored before reform calculus texts were widely available: *Brief Calculus with Applications* (Larson, & Hostetler, 1987), (to be referred to as BR, denoting 'Before Reform'). These changes in approaches to the teaching of calculus led to the research question: What is the nature of any changes in one teacher's practice during her implementation of RC relative to her instruction using the BR? I define teacher practice as what the teacher does concerning instruction in the classroom or in preparation for instruction. Relative to her practice, a special focus was placed on the teacher's actions, her evaluation practices, technology use, representations used in instruction, and on the roles of conceptual and procedural knowledge. I define teacher beliefs as a "teacher's view or conception of the nature of mathematics, model or view of the nature of mathematics teaching, [and] model or view of the process of learning mathematics" (Ernest, 1989, p. 249).

The study described here was part of a doctoral dissertation completed at Illinois State University under the direction of Beverly S. Rich, Ph.D.

My second goal was to examine the nature of any changes in this teacher's practice throughout her first-year using the RC. As the teacher progresses through the RC, how stable are new strategies as she interacts with the text, with the particular content she is teaching, and with her students? Of particular interest is the extent to which any initiated changes endure. The difference in these two goals is that the second goal only looks at the changes in teacher practice over the course of the first year of implementation; it does not compare it to previous instruction using the BR text.

Research documents the extent to which both high school and university calculus teachers' have modified their practice. In general, many high school teachers have implemented more reform-orientated strategies into their mathematics classrooms (Bruckerhoff, 1994; Edwards, 1995; Ponte, Matos, Guimaraes, Leal, & Canavarro, 1994; Wilson, & Lloyd, 1995). In calculus classrooms, teacher change in practice (Edwards, 1996; Tucker & Leitzel, 1995) and beliefs (Edwards, 1996) have also been documented. Among the changes noted in these studies are an increased use of calculators and computers, cooperative learning, student projects, significant writing tasks, and modeling activities.

My third goal was to describe any influences the RC curricular materials may have on this teacher's change process. These effects are of interest because texts and other curricular material can serve as change agents: "Because many teachers rely on textbooks as a core for their teaching, a textbook is a reasonable candidate for communicating and providing guidance for change" (Ball, 1990, p. 257). The RC addressed in this study aspires to this goal; it is a harbinger of many of the changes espoused by the reform calculus movement.

While texts have been used in the past to try to influence instruction, they have had varying degrees of success in doing so. However, Ball states that texts "clearly can provide guidance to teachers . . . in selecting better mathematical tasks, and in creating different kinds of activities" (Ball, 1990, p. 257). Texts are unlikely a sufficient enough force to bring about all of the changes advocated by reform movements in mathematics education. However, texts reach a larger number of teachers than reform documents and policies (Ball, 1990).

Reform-orientated curricular materials have been found to influence teacher change at the secondary school level. Edwards (1995) studied three secondary school mathematics teachers implementing reform-orientated texts for the first time. All three teachers altered their practices in ways that included use of cooperative learning, students independently reading their texts, and a greater emphasis on mathematical connections. Edwards claimed that "the innovative ... materials seemed to have facilitated changes in instructional practices" (Edwards, 1995, p. 9). For certain teachers, the use of reform-orientated curricular materials has facilitated the change process. For others, their lack has been found to impede change (Wasley et al., 1995).

Method

In this qualitative study, the major techniques used to collect the data were nonparticipant observation, interviews, and written document collection. Data were used to generate a theory about one teacher's implementation of a reform calculus curriculum at the high school level. Grounded Theory (Strauss & Corbin, 1990) was used to generate this theory and verify it. It is theory building using an inductive process. This is particularly well suited to this study as there is little existing research about the process of teacher change in a reform calculus curriculum and even less about this process at the secondary level.

Description of Subject

In this case study, I focus on the teaching experiences of Beth (a pseudonym) who is a high school mathematics teacher. At the time when this study took place, she had taught mathematics for 13 years, during 7 of which she has taught a section of calculus. All 13 years were spent in her present high school located in a rural area of the Midwestern United States. She holds Bachelor's and Master's degrees in mathematics education and was one of three mathematics teachers in her department. Her class, the first year she implemented the RC, consisted of 10 students.

I selected Beth to be a participant in this study after it was learned that she had decided to choose a reform-oriented calculus textbook. Given her rather classical methods of teaching, I believed that the Beth's implementation of the reform curriculum would be a unique opportunity to examine a teacher involved in the process of change. Beth accepted my invitation without hesitation. Overall, she is a very open person; this openness included her willingness to examine her own instruction as she implemented the reform curriculum. Her openness yielded frankness regarding all facets of her own teaching during the data collection process.

Data Collection

Data collection techniques here included nonparticipant observation, teacher interviews, and written document acquisition. Use of multiple data sources allowed for "data triangulation" (Yin, 1994, p. 92). Triangulation of data reduces problems concerning construct validity.

I began data collection the summer before implementation of the RC when I conducted baseline interviews with the teacher. They focused on teacher beliefs and instructional practices. Beth's instruction from the previous curriculum was reconstructed using interviews that focused especially on lessons pertaining to limits, continuity, conceptual development of the derivative, the chain rule for differentiation, conceptual development of the integral, and integration by substitution. Informal observation data were also acquired during the second half of the last school year the teacher used the BR when the researcher spent 12 weeks in the teacher's calculus classroom. While no formal record of these

observations exists, it gave the researcher an opportunity to gain insight into the teacher's instruction in the previous curriculum.

The following school year, data collection consisted of observations of the teacher's instruction in the RC. The focus of the observations was Beth's instruction in the chapters devoted to the conceptual development of the derivative and integral. Also, lessons focusing on limits, continuity, the chain rule for differentiation, and integration by substitution were observed, with the exception of days entirely devoted to evaluation. For the remainder of the school year, a systematic sample of classroom observations was made. During this period, a minimum of one section in each chapter taught was observed. Lessons observed were audio taped, videotaped, and selectively transcribed. Handwritten field notes taken during each observation were used to construct a description of each observed lesson. This information was used to construct detailed notes of each class observed; on average, each description was five typed pages long. During the observations and the subsequent data analysis, the focus was on mathematics content, teacher actions and beliefs, assessment, technology use, and representations used in instruction. A total of 52 calculus lessons taught by the teacher in the RC were observed during the school year.

Teacher interviews were conducted after all observations, typically at the school immediately following the class observed. Any discussion of the Beth's planning process took place during these interviews that followed the observations. Depending on the teacher's schedule, the interviews were occasionally conducted later that day on the telephone. Longer interviews were also conducted before and after each chapter in the text and midway through the school year. All interviews were tape recorded and transcribed in their entirety. In all, 62 interviews were conducted. At the end of the school year, a final teacher interview was administered regarding the teaching and learning of mathematics and the Beth's own perceptions of her change process.

In addition, the following artifacts were collected from the teacher's instruction in both curricula:

(a) teacher lesson planning notes;

(b) all evaluation instruments and classroom handouts; and

(c) student notebooks to help document calculus lessons in both curricula.

Data Analysis

Qualitative methods were used to analyze the data collected for this study. They were analyzed using grounded theory methods (Strauss and Corbin, 1990). The particular manner in which the data were used is as followed. Field notes, interviews, and written documents were coded using conceptual codes. Coding the data helped the researcher find commonalties in it. Initially, these data were open-coded for rough categorization. During the coding process, the focus was on mathematics content, teacher actions and beliefs, assessment, technology use, and representations used in instruction. After initial coding, all documents were then re-examined as a part of the coding process. Axial coding techniques were then used to relate categories and their subcategories discovered during the open coding process. Any relationships between the different categories were examined to determine the presence of more abstract concepts that might link the less abstract categories (Strauss & Corbin, 1990). Concept maps were constructed to help in this process. Direct comparison of instruction in both curricula was also made. Multiple sources of data were used to validate the trends that emerged from the data. The multiple sources of data and their collection over 10 months reduced possible effects related to the teacher's own awareness as a participant in this study.

In order to compare evaluation instruments used by the teacher in the two curricula, all test and quiz questions were coded as procedural or conceptual. Evaluation questions were also coded as to whether or not they called for application of mathematics in a context. Comparison between the teacher's evaluation instruments was made using these criteria.

Results

At first glance, the teacher in this study maintained a relatively similar mode of instruction while using both curricula. Her practice revolved around demonstrating solutions to prototypical examples. However, several important changes in her practice were noted. These include an increased focus on conceptual knowledge and increased use of technology.

Teacher's Instruction in the BR curriculum

Beth's instruction in the BR curriculum followed a consistent pattern. Beth began class by grading students' homework, followed by students' questions on the previous night's assignment. Reading or introductory material in the text was rarely discussed in class and students were not required to read it: "The last book—myself, I read the sections [when planning], but nobody else did and all we did [were] problems". However, there were some exceptions to this. The BR's reading was occasionally, but not often, discussed in greater detail. When Beth did so, it was in the form of a lecture in which students participated by answering the teacher's questions the relating to various parts of the reading. These instances were early in the school year. For example, she discussed the reading's content in the text's first chapter in the lessons pertaining to limits and continuity.

Beth's presentation of new material followed a pattern that consisted primarily of solving prototype examples. She presented solutions to what she called the 'three-throughs' problems in the text (every third exercise in the problem set beginning with the third problem):

> I know a lot in this [previous] book ... I would give them enough examples so that mainly they had to just learn the examples that I did, and the other ones

> [in the assignment] matched fairly close[ly]. And so I didn't really challenge them to do new problems.

Here Beth largely guided students through the steps in the various types of problems. She would ask students to complete parts of the problems orally, using primarily lower-level questions. Normally, most of class time was used to present sample problems. Students would rarely have time to begin their assignment in class.

Teacher's Instruction in the RC

Insights into the Teacher's Plans For and Attempts at Change. Described here are the first stages of this teacher's change process, beginning with the teacher's choice of a reform calculus curriculum and continuing with its early implementation. Variations in her practice and the degree to which she was successful at her planned changes are also described. Insight into various factors that interacted with her change process is also provided.

In her last semester using the BR, Beth began her change process when confronted with the choice of a new calculus text. She enthusiastically began the process of text selection and sought out the advice of a university mathematics educator who recommended the RC (Hughes-Hallett, et al., 1994). After studying it, she decided that she wanted her school to adopt it. Beth chose this particular curriculum because she liked its application focus and its philosophy of learning: "Why I have picked this book. . . . in this [text's preface] about how the students learn from this book and the rule of three. I liked where they're going to make them do it. . . geometrically, numerically, and algebraically. . . Also they said their formal definitions and procedures evolve from the investigations of practical problems and that was another reason that I liked it. . . It had non-routine problems." (All quotes not cited are direct quotes from Beth.) The 'rule of three' refers to by the teacher pertains to studying various calculus concepts from three different points of view: algebraically, geometrically and numerically. This 'rule of three' is one of the cornerstones of the philosophy regarding the reform movement in calculus.

After the decision was made to adopt the RC, Beth made ambitious plans for its use, including many that were contrary to her past practices: "My number one goal would be to get them to read the book and number two would be for them to do more on their own ... for me not to do everything for them". She also hoped to have her class use graphing calculators. She also attended a week-long summer workshop for high school teachers on the reform calculus movement at a local university. This workshop's focus on projects had a particular effect on the teacher. Before the workshop, she seemed open to the idea of students completing projects, but after the workshop her enthusiasm for projects increased.

Overall, Beth looked forward to the upcoming implementation with enthusiasm, but she also had some concerns. Among her other concerns was the fact that the RC's problem sets were not constructed so that she could easily teach students how to solve the different problems using prototypical examples. This and the fact that the RC contained far fewer exercises than the BR influenced her to consider other instructional strategies.

When the school year began, Beth initiated some of her planned changes. She asked her students to read and take notes on the text reading and try some of the exercises before the lesson was discussed in class. Despite the Beth's enthusiasm for using the RC, the school year had a difficult start. Students' reaction to the course was both negative and vocal. One third of her students withdrew from the course in the first two weeks of the semester. Furthermore, the students that remained were often off-task and complained more than one might expect. Early in the school year, one student even commented in class: "I think we should go back to the old books and the old ways of teaching". Other students expressed similar sentiments.

Beth sensed her students' frustration but was unsure of its source: "They're frustrated . . . right now, and I don't know—I'm trying to teach it differently, but it's not that much different". She felt that one source of their frustration was her students' strong emphasis on earning high marks. This stressful beginning to the school year led to extreme teacher frustration.

As the semester progressed, students' negative reactions to the course did not decrease, but the teacher's level of frustration increased. She was also concerned that her students were having great difficulty understanding the review material found in the first part of the text. Student and teacher frustration finally influenced the teacher to modify her practice. Around three weeks into the school year, she stopped asking her students to read the text on their own and began demonstrating how to solve various problems before the students tried problems on their own:

> What happened at the beginning was that I tried to have them do it on their own and switched to trying to show them problems that matched [the assigned problems]. So I'll start them on [solving] the problems if I think they're going to struggle, but eventually I hope to get back where I don't have to do that, but with this group, I may not be able to.

However, Beth did not disregard the ideas in the RC's reading altogether. Instead, she began to present the reading material in the form of a lecture.

Although Beth's instruction remained similar in the two curricula in many ways, she was able to initiate important changes and hold on to hopes for continued change. Throughout the school year, she maintained a desire for students to complete projects. However, she never asked her students to complete the projects originally planned, citing time as a barrier.

General Portrait of the Teacher's Instruction in the RC. While some changes were never realized, Beth was able to accomplish other important changes, including an increased concept focus and graphing calculator use. Her increased

concept focus was influenced by the RC's conceptual focus. Over the course of the school year, she also used graphing calculators in important ways.

After the fluctuations in Beth's practice described earlier, a relatively stable pattern of instruction emerged during the teacher's first year using the RC. Class would normally begin with the teacher going around to students' desks to mark their homework. After students' homework was marked, Beth discussed the assignment. This discussion took on various forms. Sometimes she displayed a teacher-constructed answer key and made general remarks about how to solve various problems. When she had not constructed an answer key, she simply asked for student questions.

Following the discussion of homework, Beth presented the new material. This normally began with a lecture on the reading. Sometimes she would also discuss some of the examples found in the RC's reading. Occasionally, she would not cover a lesson's reading in class. This tended to be true when the topic was more procedural in nature.

An example of Beth's discussion of the text reading material can be found below in an extract from my notes. It comes from an introductory lesson on concepts related to the derivative. This lesson immediately followed one devoted to looking at the average rate of change of an object thrown into the air:

> The teacher introduced this section by saying: "We have not used the word derivative at all this year. Last year when I taught calculus and in previous years, my little definition of calculus was pretty much to learn to do derivatives and anti-derivatives and their applications. The derivative is a major topic in calculus and that is actually the start of what we're doing today. The derivative is just a rate of change." She then read from the text a discussion of the average change of the position function on the interval from a to $a + h$, making a connection with the first section that had discussed this idea in greater depth using the example of a grapefruit thrown in the air. She then referred to the point in the reading where the average rate of change of a function on the interval was defined for functions in general. She made connections between this and the definition of the average rate of change denoted by $\frac{\Delta y}{\Delta x}$.

Statements, such as these, that pointed out connections to previously studied ideas are indicators of the teacher's greater emphasis on concepts in the RC. This discussion continued:

> A student then asked the teacher if the average rate of change was "really just the change of y over the change in x", and the teacher replied "yes". This same student then asked "so what's the difference between this [the average rate of change over the interval from a to a+h] and the average velocity?" The teacher told him that they were the "Same. The average velocity and the difference quotient are the same … These two sections go together. The last section ... gives me one example [using the position function] and then this section goes for everything [all functions]. The average velocity that we found in the last section was just for the position function … It was a particular case. Now they're putting it for everything". The lesson continued where the teacher chose examples worked out in the text's reading to discuss. She intertwined this discussion with new, key definitions, which again connected back to the same ideas examined at in the previous day's lesson that examined the rate of change of an object thrown into the air. She also discussed, at length, the geometric visualization of the derivative of a function at a point. Finally, she summarized the two interpretations of the derivative at a point A.

After discussing the reading, Beth would typically present solutions to problems she chose from the text's exercise set. While presenting sample problems, she asked specific questions of different students as she carefully guided the class through the problems' solution processes. This discussion was often procedural in focus.

An important trend noticed during her presentation of sample problems was a tendency by Beth to sometimes give students rules beyond those found in the text to 'facilitate' solving the problems. This sometimes led to the proceduralization of a more conceptual problem. This tendency was especially prevalent when the teacher presented more procedural problems. Sometimes these rules were simply stated and not motivated by an explanation or discussion why the rule worked. However, this was not always true.

The final minutes of the class period were usually reserved for what Beth called "homework hints". The goal of the homework hints was to give the students a 'head start' on assigned problems that she had determined to be more difficult. Homework hints given varied from telling students what sample problems solved in class resembled assigned problems to starting a problem for students. Occasionally she would work out a problem out in its entirety.

At this point class time was nearly over. It was important to Beth that the majority of time was devoted to presenting new material. Little class time was allotted for students to begin their assignment as is common in many high school mathematics classrooms. At the end of class, Beth adjusted the homework assignment based on what sample problems had been presented in class.

Discussion

Comparing Beth's calculus instruction in both curricula, her practice remained unchanged in most ways. In both years, her goal was to show students how to do problems that she hoped would match those assigned. Despite this, she was

able to make modifications in her practice that included a greater focus on conceptual knowledge and the use of graphing calculators to teach calculus. She was able to accomplish these changes despite strong and sustained resistance from her students. This resistance and the teacher frustration that resulted from it influenced her to abandon several of the reform-orientated strategies implemented early in the school year.

The research on Beth's change shows that it is generally easier for teachers to make additions to practice than to modify existing practices (Garet & Mills, 1995). It is interesting to note that the changes in her instructional practice included both modifications and additions. The major addition to her instructional repertoire was her use of graphing calculator technology. The most important ways in which she modified her practice was an increased focus on concepts. For Beth, modifications of some prior practices proved especially difficult. This is demonstrated by the largely teacher-centered classroom that she maintained in spite a strong initial desire to put more of the responsibility for learning on her students.

This teacher was not only able to make additions but also modify her instruction in important ways. The primary way that she did this was through a greater focus on calculus concepts. The RC was very influential in helping her attain this focus. It is very likely that without it, she might not have changed to the same degree as she did. In a way, the RC served to help her overcome some of the barriers present in bringing about such a change.

Regarding my second goal, Beth's shift from practices early in the school year that encouraged more independent learning to instruction that focused more on demonstrating prototype examples was the major shift in her practice that took place during her first-year implementation. Giving students 'homework hints' was another instructional strategy that this teacher developed over the school year. This change also influenced the degree to which her classroom was teacher-centered.

The relative stability in this teacher's practice emerged despite her ambitious plans for change. Initially, she implemented many of her plans to use 'reform-orientated' strategies. However, her students resisted them. This resistance was key in this teacher's change process. Other research (Bruckerhoff, 1994; Duffy & Roehler, 1986) has found student resistance to be a change impediment. Without this resistance, she may have continued with the changes she had instituted at the beginning of the year and perhaps instituted even more as the year progressed. Had this occurred, her change process may have looked very different.

Changes in Instructional Practices

Two major trends emerged during Beth's first year using the RC that set it apart from her instruction using the BR: an increased focus on conceptual knowledge and use of technology. Her technology use was influenced by the RC, yet she also developed her own ways of integrating technology into her instruction. Her greater emphasis on conceptual knowledge was due to the more conceptual nature of the problems solved in class and completed on assignments as well as her choice to discuss concepts found in lesson's introductory reading of the RC.

Focus on conceptual knowledge. Beth's instruction in the RC focused to a much greater extent on important calculus concepts relative to that in the BR. This change was influenced, but not merely a result of, the RC text's conceptual focus. This greater focus on concepts was not without some struggle on the Beth's part: "You kind of struggle with the procedures or presenting concepts. I think that the older book—you really were doing more procedures where this one is actually more conceptually [orientated]—the new book is". This struggle was a difficult one for Beth, perhaps because of her own beliefs regarding the discipline of mathematics. She believed that she held more procedure-orientated beliefs: "In the new book they start from the conceptual view and so you're asked to do that. You know, I'm more of a numbers person". Her underlying beliefs are in many ways opposite to the views of the curriculum she had chosen. This underscores the importance in her change process of the instructional decisions she made that favored conceptual knowledge.

During her implementation of the RC, Beth made a *conscious choice* to focus on the introductory reading material. In the BR, she rarely discussed in-depth the conceptual ideas found in the reading even though they were often discussed in the BR's reading. Her decision to focus on the RC's reading was an important factor in her greater focus on concepts when using the new text. When she first used the RC, she asked students to read and study the text. While she quickly changed this practice, she did not abandon the ideas in the reading altogether and comfortably revert to her old routines. Instead, Beth adapted her strategies to include a presentation of the reading's ideas. Her choice to focus on the reading had important effects regarding her use of class time. It shifted the distribution of class time away from demonstrating solutions to prototype problems, which accounted for the majority of the class time when she used the BR, to a greater emphasis on discussing the concepts in the reading. The more problem solutions demonstrated in class, the more quickly she could cover material. For Beth, pacing and 'covering material' were of utmost importance. The fulfillment of pacing goals was often the greatest factor she used in determining whether or not a particular lesson 'went well'. Her decision to devote class time to the discussion of the reading was a difficult one for her, and she had to reaffirm it each time she planned a lesson. This underscores the importance of *her choice* to focus on the reading and its subsequent impact on her change process.

The nature of the problems discussed in class and those that Beth asked students to complete in their assignments also influenced her focus on concepts. The importance she placed on conceptually orientated exercises was particularly

influential. The emphasis in class that Beth placed on a particular problem was a *conscious choice* on her part and not a decision dictated by the RC. Again, this is a stronger indicator of teacher change, and not simply a case of a teacher carefully following a text. While using the RC, Beth gave added emphasis in several instances to problems of a highly conceptual nature. These often involved applications. This was done despite her strong concern for pacing. This is an indicator of stronger teacher changes than merely following closely a particular text.

Below I provide one instance in which Beth went beyond the RC's focus on concepts. In the chapter devoted to the conceptual development of the derivative, interpreting the meaning of the derivative in various contexts was studied. Beth liked the text's emphasis on asking students to explain the meaning behind a derivative in a contextual situation and saw it as typical of the RC's focus on concepts:

> Some of the problems ... it was just kind of neat some of the problems they would give—They would ask you to explain, in words, what [the derivative] meant. I've never been asked that before. It even made me stop and think 'what are they doing here?'.

Beth believed that the RC's focus on explaining the meaning of derivatives in applied settings was important and it influenced her to later write her own extensions of a similar nature for several problems in the RC. Such extensions were not something proposed by the RC and give further evidence of Beth's change relative to her focus on concepts.

Use of graphing calculator technology. Another important change in Beth's instruction concerns her use of graphing calculator technology. Edwards (1996) as well as Tucker and Leitzel (1995) describe calculus teachers' abilities to implement technology as they worked to reform their instruction. In her implementation of the BR, Beth used no technology other than the scientific calculator with the exception of a two-week add-on unit at the school year's end that focused on various graphing applications of the graphing calculator.

When using the RC, Beth furnished students with a graphing calculator, and they used it on a regular basis throughout the school year. Its use dropped off—but did not disappear—in her instruction chapters devoted to procedures for finding derivatives and anti-derivatives. Of course, the RC and its technological focus influenced this change; Beth commented several times how she would have had to eliminate some of the material in the RC if the appropriate technology had not been available.

While technology use by Beth was heavily influenced by the RC's technological focus, deeper changes regarding technology use by Beth were also noted. In several instances, she used technology in novel ways not called for by the RC. For example, she presented in class the solution to a problem that asked for the derivative of $f(x) =$ arctangent$(x) +$ arctangent$(1/x)$ for $x > 0$. After calculating a zero derivative for this function, she explained plans for showing students on the graphing calculator that $f(x)$ represents a horizontal line and, as such, has a derivative of 0. Beth wanted to "make sure they understand that's a constant function, and then what I was going to have them do was use their graphing calculator—that means the original function when I graph it—it better be a line" if it has a constant derivative. For her, the graphing calculator was becoming a tool that could help her reinforce and connect ideas together.

Further demonstrating the extent of her change with respect to technology, Beth started to use graphing calculators with her algebra classes in ways that she had not in the past. The algebra text she used did not advocate technology use, and it certainly could have been implemented without it. In prior years, she had used graphing calculators to a very limited extent to teach algebra. This was primarily during her presentation of linear and quadratic functions.

While the RC was very influential in the changes in her calculus course, it is certainly not responsible for the full extent of her change. The fact that technology use was becoming a part of her instructional decision-making in not just her calculus class, but also her other classes, demonstrates this.

For this teacher, the change process was influenced by several factors. Her students' beliefs and attitudes, the reform-orientated curricular materials, and prior and emerging teacher beliefs were all inter-related in their effects on the change process. The curriculum materials themselves, coupled with an open-attitude on the part of the teacher worked together as a catalyst for change. Student resistance worked to slow down the effect of this catalyst. In a way, it helped re-emerge the teacher's prior beliefs that were less aligned with the reform movement. The change process itself also influenced these same factors.

Conclusion

Because of the nature of this study, one can not conclude that all high school mathematics teachers would react the same way when using the RC for the first time. However, Beth's story helps those concerned with the teaching and learning of mathematics better understand the complex environment that is the high school calculus classroom and the positive impact that reform curricula can have on it. Since the RC materials appear to have had an important influence on Beth's practice, other high school calculus teachers in similar settings might also find them beneficial.

This teacher's change process was greatly influenced by the reform-oriented curricular materials as well as the students she taught. Such influences lead to potential research opportunities. What are the struggles of students new to learning in more reform-oriented classrooms, and how might their transition be eased? The students in this study had difficul-

ties reading the calculus text independently. How might teachers help students use more effectively reform mathematics texts?

The role that the RC played in this teacher's change process was critical. Reform-oriented curricular materials can be a change catalyst for high school mathematics teachers even in unexpected settings, such as this one. Studies of how teachers engaged in the reform process use successfully reform-oriented curricular materials and their accompanying ancillary materials would have important implications for teacher change. Which particular aspects of reform curricula promote change and which are less effective? Studies of ancillary materials may be an especially consideration for teachers using reform curricula who have little or no access to professional development opportunities. Such studies might have important implications for those hoping to promote teacher change as well as for those who design reform-oriented curricula.

The changes that this teacher underwent as she first used the RC are not 'revolutionary', and it would be difficult to argue otherwise. The changes discussed here are just 'small steps' towards a learning environment that focuses to a greater extent on understanding calculus and not merely on being able 'to do' calculus. Mathematics teachers hoping to change their instruction have had various degrees of success. Changing established classroom instructional routines is very difficult, yet this teacher made important progress in this area. The teacher's openness toward change and how this contributed positively can not be underestimated. It is particularly interesting that Beth accomplished these changes largely on her own, without any ongoing professional development or support of any kind. What Beth might have accomplished with the aid of an ongoing professional development program or collaboration with other teachers engaged in change will never be known but may have been significantly greater than seen here.

References

Ball, D. L. (1990). Reflections and deflections of policy: The case of Carol Turner. *Educational Evaluation and Policy Analysis, 12,* 247-259.

Bruckerhoff, C. (1994). *School routines and the failure of curriculum reform*. University of Connecticut: Department of Curriculum and Instruction. (ERIC Document Reproduction Service No. ED 384 977).

Duffy, G., & Roehler, L. (1986). Constraints on Teacher Change. *Journal of Teacher Education, 37,* 55-58.

Edwards, B. S. (1996, September). *Calculus reform: Change and the calculus instructor*. Paper presented at the annual meeting of the Research in Undergraduate Mathematics Education Conference, Mount Pleasant, MI.

Edwards, T. G. (1995, April). *Use of an innovative curriculum: Enabling and inhibiting change in practice*. Paper presented at the 73rd Annual Meeting of the National Council of Teachers of Mathematics, Boston, MA. (ERIC Document Reproduction Service No. ED 382 456)

Ernest, P. (1989). The impact of beliefs on the teacher of mathematics. In P. Ernest (Ed.) *Mathematics teaching: The state of the art* (pp. 249-253). New York: Falmer.

Ferrini-Mundy, J., & Graham, K. G. (1991). Overview of the calculus reform effort: Issues for learning, teaching, and curriculum development. *American Mathematical Monthly, 98*, 627-635.

Garet, M. S., & Mills, V. L. (1995). Changes in teaching practices: The effects of the curriculum and evaluation standards. *Mathematics Teacher, 88*, 380-388.

Hughes-Hallett, D., Gleason, A. M., Flath, D. E., Gordon, S. P., Lomen, D. O., Lovelock, D., McCallum, W. G., Osgood, B. G., Pasquale, A, Tecosky-Feldman, J., Thrash, J. B., & Tucker, T. W. (1994). *Calculus*. New York: John Wiley & Sons.

Larson, R. E., & Hostetler, R. P. (1987). *Brief calculus with applications* (2nd ed.). Lexington, MA: D. C. Heath & Company.

Ponte, J. P., Matos, J. F., Guimaraes, Leal, L. C., & Canavarro, A. P. (1994). Teachers' and students' views and attitudes towards a new mathematics curriculum: A case study. *Educational Studies in Mathematics, 26*, 347-365.

Ross, S. C. (1996). Visions of calculus. In A. W. Roberts (Ed.), *Calculus: The dynamics of change* (pp. 8-15). Washington, DC: The Mathematical Association of America.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques*. Newbury Park, England: Sage Publications.

Tucker, A. C., & Leitzel, J. R. (Eds.). (1995). *Assessing calculus reform efforts: A report to the community*. Washington, DC: The Mathematical Association of America.

Wasley, P. A., Domoyer, R., & Maxwell, L. (1995). Navigating change in high school science and mathematics: Lessons teachers taught us. *Theory into Practice, 34,* 51-59.

Wilson, M. R., & Lloyd, G. (1995). High school teachers' experiences in a student-centered mathematics curriculum. In D. Owens, M. Reed, & G. Millsaps (Eds.) *Proceedings of the seventeenth annual meeting of the North American chapter of the international group for the psychology of mathematics education*. (Vol. 1, pp. 162-167). Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.

Yin, R. K. (1994). *Case study research: Design and methods* (2nd ed.). Thousand Oaks, CA: Sage Publications.

Teacher Candidates' Conceptual Understandings of Mathematics Concepts

Angela T. Barlow
Jill Mizell Reddish
University of West Georgia

Abstract

As universities strive to produce the best mathematics teachers possible through both graduate and undergraduate programs, teacher educators must constantly work towards helping teacher candidates create their own conceptual understanding of mathematics. This pilot study examined the effect teaching in a constructivist manner had on teacher candidates' conceptual understandings of the arithmetic mean and on their ability to transfer this knowledge into their instruction. Results indicated that teaching in a constructivist manner can have a positive impact on teacher candidates' understanding of the arithmetic mean, but their abilities to transfer this new knowledge into their own instructional practices was inconsistent.

Having a highly qualified teacher in every classroom is a goal of not only the federal government, but also every parent, school administrator, and teacher educator. Being a highly qualified teacher entails having adequate content knowledge as well as being able to utilize that knowledge to bring students of diverse backgrounds to high levels of learning. Supporting high levels of learning in mathematics requires the teacher to have both a procedural and conceptual understanding of the mathematics being taught. Procedural understandings entail knowledge of the rules and sequences of actions taken in algorithms. Conceptual understandings involve making connections and understanding relationships between discrete pieces of mathematical information (Reys, Suydam, Lindquist, & Smith, 1998). Teachers who have conceptual and procedural understandings of mathematics produce students who "exhibit conceptual understanding, have the ability to learn and reason, and are able to achieve" (Knight, 2001, p. 21).

As colleges and universities strive to produce highly qualified mathematics teachers through both graduate and undergraduate programs, teacher educators must work towards helping teacher candidates create their own conceptual understanding of mathematics. To develop this level of deep understanding, teacher candidates must experience a learning environment that is much different than the one typically found in a traditional college mathematics course. In order to take an active role in constructivist teaching, teacher candidates must first take an active role in constructivist learning. The experiences teachers have as learners will have a tremendous impact on the beliefs and attitudes brought into their own classrooms (Chappell & Thompson, 1994).

The preparation of elementary teachers is of particular importance when trying to establish a conceptually-based foundation of mathematics for teacher candidates. Teachers at the elementary level often are aware of the need to teach algorithms associated with mathematical concepts such as addition, subtraction, multiplication, and division in a conceptually-based manner. However, this is not the case when teaching more advanced concepts such as finding the arithmetic mean of a data set, hereafter referred to as the mean. Teachers typically provide the conventional algorithm yet have no understanding of how the algorithm is derived or why it is useful in solving posed problems. Despite its simple nature, "developing a conceptual underpinning that allows one to use the mean sensibly is surprisingly difficult" (Konold & Higgins, 2003, p. 204).

In this article, we present a pilot study that examines the effect learning in a constructivist manner has on teacher candidates' conceptual understandings of the mean and on their ability to transfer this knowledge into their instruction. Finding methods for empowering teachers to effectively teach mathematical concepts such as the mean is of great importance if the goal of teacher educators is to prepare teachers to teach in line with the National Council of Teachers of Mathematics' (NCTM) *Principal and Standards for School Mathematics* (2000). In addressing this need, the following research questions were posed:

1. Does learning about the mean in a constructivist manner extend teacher candidates' definitions of the mean from a primarily procedurally-based definition to a more conceptually-based definition?
2. Does learning about the mean in a constructivist manner have an impact on teacher candidates' ability to use the algorithm?
3. Are teacher candidates able to connect the algorithm to the equal distribution interpretation of the mean?
4. Does learning about the mean in a constructivist manner have an impact on teacher candidates' ability to design lessons that develop a conceptual understanding of the mean in elementary-aged students?

Sample and Methodology

Subjects

This pilot study was conducted during the spring semester of 2004 and involved 45 teacher candidates enrolled in one of two sections of a probability and statistics course for teachers. This population sample was taken from a southeastern state university which has a current enrollment of over 10,000 students.

Each teacher candidate involved in this pilot study was an early childhood major seeking initial certification. The sample consisted of 43 females and 2 males. Of the 43 females, there were 33 Caucasians and 10 African-Americans. Both males were Caucasians. The sample was comprised of 42 seniors, 2 non-traditional graduate students, and 1 junior. Forty-three of the teacher candidates had successfully completed their mathematics methods course which was designed to develop an understanding of mathematics content, methods, and materials appropriate for the cognitive development of the elementary child. These teacher candidates were in their last semester of coursework prior to their student teaching. The remaining two teacher candidates were taking the statistics course before beginning their junior/senior years of professional coursework. As a result, they had not completed the methods course. This sample was heterogeneous in that the teacher candidates had taken a variety of mathematics courses.

Procedures Used

Prior to the pilot study, teacher candidates completed a pre-survey designed to assess their definition of the mean and their ability to compute the mean and determine numbers that could be added to a data set without affecting the mean. The survey included the following questions.

1. In your own words, explain what the mean (or average) is.
2. Find the mean of the data set: 1, 1, 2, 2, 2, 2, 3, 3, 4, 5
3. Find three numbers that can be added to the given data set and not change the mean. Explain how you chose these three numbers.

The pilot study was conducted over 2 class periods and consisted of three constructivist-based lessons. The lessons were deemed constructivist in that they were aligned to constructivist theory and adhered to the following: the instructor served as a coach or facilitator; a real-world context was employed (Jonassen, 1991); instruction focused on knowledge construction; multiple representations of the content were examined; "collaborative construction of knowledge through social negotiation" was supported (Jonassen, 1994, p. 35); and scaffolding was used to engage students in higher-order thinking. The objective of the first lesson was to understand the mean as equal distribution. The second lesson's objective was to understand the mean as the balance point of the distribution. The final lesson allowed the teacher candidates to utilize technology in exploring the properties of the mean.

Following the three lessons, teacher candidates completed a post-survey that was identical to the pre-survey. In addition to the surveys, teacher candidates were asked to respond to a journal prompt that dealt with their understanding of the mean and to create a lesson plan designed to develop a conceptual understanding of the mean appropriate for fourth graders.

Data Analysis

The researchers used the four research questions previously presented to guide the data analysis process. To investigate research question one, teacher candidates' pre-survey responses to survey question one were compared to their post-survey responses to the same question. For this open-ended question, the researchers utilized codes to represent responses: C represented the ability to conceptually define the mean; P the ability to procedurally define the mean; and W the inability to define the mean correctly. Each researcher independently analyzed the responses using the codes. Together, we reviewed the coding and found the ratings to have inter-rater reliability. Data analysis consisted of a comparison of the percentages of teacher candidates in each of the categories from pre-survey to post-survey.

For further information on teacher candidates' understanding of the mean, we coded the responses to survey question two with a Y representing the ability to correctly calculate the mean or an N representing the inability to calculate the mean. We calculated the percentage of teacher candidates who were able to correctly compute the mean on the pre-survey.

Data for research question two was gathered by comparing teacher candidates' ability to add numbers to a given data set on the pre-survey and the post-survey. We coded the responses with a Y representing the ability to successfully complete the task and an N representing the ability to do so. Percentages of teacher candidates falling into each of the categories were compared.

To investigate research question three, we compared the teacher candidates' responses to the journal prompt which read, "Explain in your own words why it makes sense to sum the numbers in a data set and divide by the quantity of numbers when finding the average." We independently coded the responses to the journal prompt with Y representing the ability to associate the algorithm to the equal distribution interpretation of the mean or an N representing the inability to do so. We verified that the codings had inter-rater reliability. The percentage of teacher candidates in each category was determined.

For research question four, we analyzed the teacher candidates' lesson plans. We independently coded the lesson plans with C representing a conceptually-based lesson or P representing a procedurally-based lesson. Our coding had a high level of inter-rater reliability. In addition to determin-

ing the percentage of teacher candidates in each category, we characterized each teacher candidate according to his or her lesson plan type and journal entry in an effort to explore the relationship between the two.

Results and Limitations

This pilot study examined the effect teaching in a constructivist manner has on teacher candidates' conceptual understandings of the mean. This section presents the results of the data collected and the limitations of the study.

Results

The findings of this pilot study indicate that participating in constructivist-based lessons can have a positive impact on teacher candidates' understanding of the mean. After having completed the 3 constructivist-based lessons, the participants of this study were more capable of conceptually defining the mean. At the onset of the pilot study the pre-survey data showed that only 2.2% (one teacher candidate) of the teacher candidates defined the mean in conceptual terms. The sole conceptually-based definition given was:

> The mean or average is like making all of the numbers in a set even and the outcome is the average. For example, if I have 70, 80, and 90 I would even the numbers out by taking 10 from 90 and giving it to 70 then the average is 80, 80, 80—you add all the # together and divide by how many you have all together.

Twenty percent (nine teacher candidates) failed to provide an accurate definition of the mean, procedurally or conceptually. The following are examples of teacher candidates' inaccurate definitions:

> "The most likely or popular to occur"
>
> "The number that is common among other numbers"
>
> "The mean is number [*sic*] that occurs most when both extremes are combined and divided by two."

The remaining thirty-five teacher candidates (77%) provided an accurate definition for the mean that was exclusively procedural. The following are examples of the procedural definitions provided:

> "The mean is when you take a set of numbers, add them together, and divide by the total number of the numbers in a set. The mean is your answer."
>
> "The number you get when you divide the sum of a set of numbers by the amount of numbers in the set."

Interestingly, regardless of the accuracy of the definition or the type of definition provided, all teacher candidates were able to correctly compute the mean.

After the implementation of the three constructivist-based lessons, a higher percentage (53.3%) of teacher candidates provided a definition that was based on conceptual understandings of what the mean represents. This is a 51.1% increase in the number of teacher candidates who were able to extend their vision of the mean beyond the procedure. The two candidates' responses provided in Table 1 illustrate the shift made in some candidates' definition of the mean from an either inaccurate or procedural definition to one that demonstrates some level of conceptual understanding.

Candidate #7's original definition of the mean was very vague and is not true in general. Candidate #25's original definition of the mean was entirely procedurally-based. She defined how to arrive at the mean versus what the mean is. Following the implementation of the three lessons, both candidates provided a more conceptually-based interpretation of the mean. Candidate #7's definition is based on her understanding of the mean in terms of equal distribution, while Candidate #25's definition is based on her understanding of the mean as the balance point of a distribution. The teacher candidates' references to the mean as the sum of a data set equally distributed and as the balance point of a data set both demonstrate that exposure to conceptually-based lessons can increase teacher candidates' ability to communicate definitions of the mean that extend beyond a procedurally-based definition to a more conceptually-based definition. The example provided by candidate #25 indicates that her new definition was not only an adjustment in language, but also a transformation in her level of understanding.

In examining the candidates' ability to use the algorithm, results indicated that participation in conceptually-based lessons enhanced their ability to use the algorithm for finding the mean. At the onset of the pilot study, 17.7% (8 teacher candidates) were able to correctly identify three numbers

Table 1
Comparison of Pre-lesson Definitions of the Mean with Post-lesson Definitions

Candidates	Pre-Lesson Definitions	Post-Lesson Definitions
#7	It is the typical average found in a group of numerical data.	It is when you find the sum of a data set and you equally distribute the amount you computed to each group/member in the set.
#25	The mean is when you take a set of numbers, add them together, and divide by the total number of the numbers in the set. The mean is your answer.	The mean or average is when you have a set of numbers that balances the set. For example, if you have a 3 and a 5 the mean is 4. There is 1 space on either side of the 4 so the sides are balanced.

that could be added to a given data set without changing the mean. Knowledge of the algorithm should have enabled them to complete this task. However, the non-traditional nature of this task seemed to prevent the candidates from solving the problem.

After the implementation of the three constructivist-based lessons, 82.2% of the teacher candidates were able to add numbers to a data set and maintain the value of the mean. These findings demonstrate that participation in conceptually-based lessons not only increases teacher candidates' ability to understand the mean but also to apply that understanding in a variety of settings.

Journal entries were used to examine teacher candidates' abilities to connect the algorithm to the equal distribution interpretation of the mean. To demonstrate this ability, teacher candidates were asked to respond to the following journal prompt after receiving conceptually-based instruction:

> Explain in your own words why it makes sense to sum the numbers in a data set and then divide by the quantity of numbers when finding the average.

In these writings, 60% of the teacher candidates provided evidence that they could relate the algorithm to the equal distribution interpretation of the mean. Many teacher candidates provided a sample problem similar to those used in the conceptually-based lessons in an effort to support and clarify their explanation. This can be interpreted as evidence that the conceptually-based examples provided in class had an influence on the teacher candidates' ability to understand and solve problems involving the mean. For example, one student wrote this journal entry:

> The average is the one number that every piece of data needs to be moved to so that all of the data is equal. When all of the data is added together and then divided by the quantity of numbers used it gives you that number. For example, if Mom gave Charlie 5 cookies, Jeff 3 cookies, Sarah 3 cookies and Suzy 1 cookie, then there are 12 cookies all together. If 12 is divided by 4 (the number of people with cookies) the quotient is three. If each person gets 3 cookies then everyone would have an equal amount.

Many of the teacher candidates who were unsuccessful in making sense of the algorithm provided reasoning exemplified in the following journal entry:

> To me the definition of average is a number that represents a large group of numbers. It makes sense that if all the numbers in a group are added up and the sum is divided by the total of how many numbers are in the group it will find the middle or average number of that group.

Responses such as this seem to indicate that these teacher candidates have not yet made sense of the algorithm in terms of how it relates to equal distribution. These teacher candidates continue to see the average solely as the number that you obtain when applying the add-and-divide procedure.

In studying the impact of constructively-based lessons, of particular importance was the teacher candidates' ability to transfer this newly acquired knowledge into their own instruction. An examination of teacher candidates' lessons on the mean revealed that 40% of the teacher candidates designed a conceptually-based lesson. These teacher candidates introduced the mean using a problem that involved equal distribution. The remaining 60% of the teacher candidates submitted lessons that focused on the procedure without providing an explanation of the procedure's origin or why the procedure made sense.

Table 2
Relationship between Lesson Plans and Ability to Connect the Algorithm to Equal Distribution

Journal Entry	Number
Connected algorithm to equal distribution	
Conceptual lesson plan	14
Procedural lesson plan	13
Did not connect algorithm to equal distribution	
Conceptual lesson plan	4
Procedural lesson plan	14

Table 2 categorizes the teacher candidates according to their ability to connect the algorithm to equal distribution and the type of lesson plan submitted. As the table indicates, not all teacher candidates who were able to connect the algorithm to equal distribution in their journal writings utilized this understanding in their lesson plans. Eight of these 13 teacher candidates, however, submitted a lesson plan that had been either copied directly or altered slightly from an internet website. Conversely, among teacher candidates' who failed to tie the algorithm to equal distribution, not all submitted procedurally-based lesson plans.

Limitations

Before providing any conclusions from this research, limitations of this pilot study need to be noted. The first limitation of this pilot study is its unique setting and sample. Given that the university is located in the southeast region of the United States and has a current enrollment of primarily in-state students, the results of the pilot study cannot necessarily be generalized to other universities.

The second limitation lies in the teacher candidates' reliance on the state's curriculum resources in providing lesson plans. Had teacher candidates not had access to such a database, the results from the lesson plans may have been different. A total of 15 teacher candidates submitted lesson plans that were acquired from the state's internet-based lesson plan resource website under the fourth grade standard that read, "explores the concepts of mean and median." Teacher candidates may have possibly read the title and sub-

mitted the corresponding lesson without carefully reading the lesson for its conceptual development. Because teacher candidates failed to create original lesson plans, their true ability to develop conceptually-based lesson plans could not be observed.

The third limitation of this pilot study is a result of the journal prompt which did not force the teacher candidates to connect the algorithm to equal distribution. Instead, it instructed them to make sense of the procedure. Perhaps these instructions were not clear for the teacher candidates who had not already connected the algorithm to equal distribution. Teacher candidates would have benefited from a journal prompt that made it clear that they were to connect the two. In aiding these teacher candidates to make sense of the algorithm, they need to be involved in discussions with one another during the constructivist lessons. In these discussions, teacher candidates could share with one another how they have made sense of the mean and its relationship to the algorithm.

The fourth limitation of this pilot study is the lack of field data. At the time of the pilot study, these teacher candidates were not placed in actual classrooms in order to implement their lessons. Had this been the case, more insight would have been given as to their ability to transfer what they had learned from the conceptually-based lessons into their instruction.

Conclusions and Implications

The purpose of this pilot study was to examine the effect teaching in a constructivist manner has on teacher candidates' conceptual understandings of the mean and on their ability to transfer this knowledge into their instruction. Teachers who have a conceptual understanding of mathematics produce students who "exhibit conceptual understanding, have the ability to learn and to reason, and are able to achieve" (Knight, 2001, p. 21). Based on the findings of this pilot study, constructivist-based teaching can have a positive impact on teacher candidates' understandings of the mean. The teacher candidates in this pilot study were able to: extend their understanding of the concept of the mean as a result of participating in constructivist-based lessons; move from a very rigid, narrow interpretation of the mean and its associated algorithm to a more flexible and conceptual interpretation of the algorithm; and associate the algorithm with conceptual definitions such as the equal distribution interpretation.

While increasing teacher candidates' personal content knowledge is important, teacher educators must also be concerned with their ability to transfer this knowledge into instructional practices. As a part of this pilot study, teacher candidates were asked to create a lesson designed to develop conceptual understanding of the mean at the fourth grade level. Sixty percent of the teacher candidates submitted procedurally-based lessons despite having participated in three constructivist-based lessons. This unanticipated result serves as evidence that taking an active role in constructivist learning is not always sufficient for producing teachers who are able to transfer this newly acquired learning into their own instructional practices. This reality is further highlighted by the fact that only half of the teacher candidates who were able to tie the algorithm to a conceptual definition of the mean such as equal distribution submitted a conceptually-based lesson. This finding suggests that in order for teacher candidates to develop a conceptual understanding of the mean in others, they must be required to do more than take an active role in constructivist learning. Teacher candidates need to reflect upon how the constructivist-based lessons enabled them to gain a better understanding of the mean and how such lessons would be beneficial for the children they will teach. They should be engaged in discussions concerning what elements of the lessons could be used with elementary-aged students. To make a more powerful impact on instructional practices, lesson study could be a part of the class as well, enabling teacher candidates to examine lessons on the mean similar to those found on the internet, and contrast these with the lessons they have just experienced.

The goal of all mathematics teacher educators is to produce qualified mathematics teachers. All would agree that the first step to being an effective mathematics teacher is having strong content knowledge which includes both procedural and conceptual understandings. The pilot study reported here has shown that this is clearly possible through constructivist-based teaching. However, if teachers are not able to transfer their own knowledge into their instructional practices, the increase in conceptual understanding has not fully served its purpose. As reported here, participation in a constructivist-based classroom is not necessarily enough to influence instructional practices of all teacher candidates. Future work with teacher candidates should include providing opportunities as part of mathematics courses for them to assess the pedagogical merit of this approach to teaching that they have experienced as learners. Then teacher educators can address how teacher candidates can incorporate these practices into their own teaching.

The findings of this pilot study verify that further research is needed in developing a comprehensive understanding of the connections between the teacher candidates' instructional experiences and the development of their mathematical knowledge as it applies to their instruction. Interaction effects may exist between the instructors of mathematics and mathematics education courses and the conceptual development of teacher candidates. These interactions should be investigated in follow-up studies. In addition, future studies should employ the use of a control group thus enabling statistical comparisons to be made and thereby providing reliable data that supports the positive impact of constructivist-based teaching on teacher candidates' instructional practices. Studies of this nature should involve teacher candidates from multiple settings so that the results will be generalizeable.

References

Chappell, M. F., & Thompson, D. R. (1994). Modeling the NCTM *Standards*: Ideas for initial teacher preparation programs. In D. B. Aichele (Ed.), *Professional development for teachers of mathematics, 1994 yearbook of the National Council of Teachers of Mathematics*, (pp. 186-199). Reston, VA: National Council of Teachers of Mathematics.

Jonassen, D. H. (1991). Objectivism vs. constructivism: Do we need a new philosophical paradigm? *Educational Technology, Research and Development,39*(3), 5-14.

Jonassen, D. H. (1994). Thinking technology: Toward a constructivist design model. *Educational Technology, 34*(4), 34-37.

Knight, G. (2001). What is fundamental mathematics? In Mathematics Teacher Preparation Content Workshop Program Steering Committee (Eds.), *Knowing and learning mathematics for teaching,* (pp. 20-21). Washington, DC: National Academy Press.

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics,* (pp. 193-215). Reston, VA: National Council of Teachers of Mathematics.

National Council of Teachers of Mathematics (NCTM). (2000) *Principles and standards for school mathematics.* Reston, VA: NCTM.

Reys, R. E., Suydam, M. N., Lindquist, M. M., & Smith, N. L. (1998) *Helping children learn mathematics* (5th ed.). Needham Heights, MA: Allyn & Bacon.

The Role of Classroom Experience in Preservice and Inservice Teachers' Assessment Literacy[1]

Craig A. Mertler
Bowling Green State University

Abstract

Assessing student performance is one of the most critical aspects of the job of a classroom teacher; however, many teachers do not feel adequately prepared to assess their students' performance. In order to measure and compare preservice and inservice teachers' "assessment literacy," two groups were surveyed using the Classroom Assessment Literacy Inventory *(*CALI*) which was designed to parallel the* Standards for Teacher Competence in the Educational Assessment of Students*. Inservice teachers performed highest on Standard 3*—Administering, Scoring, and Interpreting the Results of Assessments *and lowest on Standard 5*—Developing Valid Grading Procedures. *Preservice teachers performed highest on Standard 1*—Choosing Appropriate Assessment Methods *and lowest on Standard 5*—Developing Valid Grading Procedures*. Comparisons between the two groups revealed significant differences on five of the seven competency areas, as well as on the total scores. In all cases where significant differences were found, the inservice teachers scored higher than their preservice counterparts.*

Background

It has been estimated that teachers spend up to 50 percent of their time on assessment-related activities (Plake, 1993). Regardless of the amount of time spent on it, classroom assessment is a vitally important teaching function; it contributes to every other teacher function (Brookhart, 1998, 1999b). Sound assessment and grading practices help teachers to improve their instruction, improve students' motivation to learn, and increase students' levels of achievement (Brookhart, 1999a). According to Stiggins (1999a), "The quality of instruction in any ... classroom turns on the quality of the assessments used there" (p. 20). For all of these reasons, the information resulting from classroom assessments must be meaningful and accurate; i.e., the information must be valid and reliable (Brookhart, 1999a).

In recent years, public and governmental attention has shifted to school achievement as evidenced by performance on standardized achievement tests (Campbell, Murphy, & Holt, 2002; U.S. Department of Education, n.d.). Additionally, there has been an increase in expectations regarding teachers' assessment expertise. Teachers have been required to develop classroom assessments that align curriculum with state standards as a means of improving test scores (Campbell, Murphy, & Holt, 2002). New research on the relationship between classroom assessments and student performance on standardized tests reveals that improving the quality of classroom assessments can increase average scores on large-scale assessments as much as 3/4 of a SD (as much as 4 grade equivalents or 15-20 percentile points), representing a huge potential (Stiggins, 1999a). This is important research since it makes a connection between the quality of assessment in the classroom and assessment resulting from standardized testing programs.

Ironically, in this age of increasing emphasis on testing and assessment, many Colleges of Education and state education agencies do not require preservice teachers to complete specific coursework in classroom assessment (Campbell, Murphy, & Holt, 2002; O'Sullivan & Johnson, 1993). This continues to be an interesting phenomenon since many inservice teachers reported that they are not well prepared to assess student learning (Plake, 1993). Furthermore, these teachers claimed that the lack of adequate preparation is largely due to inadequate preservice training in the area of educational measurement (Plake, 1993). Brookhart (2001) also cited literature that calls for an increase in emphasis in teacher preparation programs on classroom assessment and a decrease in emphasis on large-scale testing. Studies have generally concluded that teachers' skills in both areas are limited.

Brookhart (2001) quite accurately summarized the research on teachers' assessment practices when she stated that teachers apparently do better at classroom applications than at interpreting standardized tests, perhaps due to the nature of their work.

What is "Assessment Literacy"?

Assessment literacy has been defined as "the possession of knowledge about the basic principles of sound assessment practice, including terminology, the development and use of assessment methodologies and techniques, familiarity with standards of quality in assessment...and familiarity with alternative to traditional measurements of learning" (Paterno, 2001). An alternative, simpler definition is offered by the North Central Regional Educational Laboratory who suggested assessment literacy is "the readiness

[1] This article was accepted for publication by the previous editorial team.

of an educator to design, implement, and discuss assessment strategies" (n.d.).

Others choose not to formally define assessment literacy, but rather to describe the characteristics of those who possess it. One such characterization proposes:

Assessment literate educators recognize sound assessment, evaluation, [and] communication practices; they

- understand which assessment methods to use to gather dependable information and student achievement.
- communicate assessment results effectively, whether using report card grades, test scores, portfolios, or conferences.
- can use assessment to maximize student motivation and learning by involving students as full partners in assessment, record keeping, and communication (Center for School Improvement and Policy Studies, Boise State University, n.d.).

Stiggins (1995) provided another similar description when he stated that "Assessment literates know the difference between sound and unsound assessment. They are not intimidated by the sometimes mysterious and always daunting technical world of assessment" (p. 240). He continued by stating that assessment-literate educators (regardless of whether they are teachers, administrators, or superintendents) enter the realm of assessment knowing what they are assessing, why they are doing it, how best to assess the skill or knowledge of interest, how to generate good examples of student performance, what can potentially go wrong with the assessment, and how to prevent that from happening. They are also aware of the potential negative consequences of poor, inaccurate assessment (Stiggins, 1995).

"The Standards for Teacher Competence in the Educational Assessment of Students"

The concept of assessment literacy is a key component of *The Standards for Teacher Competence in the Educational Assessment of Students* (AFT, NCME, & NEA, 1990). Additionally, *The Standards* are central to the study at hand, so it is imperative that they be described here. *The Standards for Teacher Competence in the Educational Assessment of Students* (AFT, NCME, & NEA, 1990) were a joint effort between the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association. This joint effort began in 1987 in order to "develop standards for teacher competence in student assessment out of concern that the potential educational benefits of student assessments be fully realized" (AFT, NCME, & NEA, 1990). They were originally developed in order to address the problem of inadequate assessment training for teachers (AFT, NCME, & NEA, 1990).

According to *The Standards* (AFT, NCME, & NEA, 1990), *assessment* is defined as "the process of obtaining information that is used to make educational decisions about students, to give feedback to the student about his of her progress, strengths, and weaknesses, to judge instructional effectiveness and curricular adequacy, and to inform policy." *The Standards*, of which there are seven, provide criteria for teacher competence with respect to the various components of this definition of assessment. *The Standards for Teacher Competence in the Educational Assessment of Students* consists of the following seven principles:

Standard 1—*Teachers should be skilled in choosing assessment methods appropriate for instructional decisions*.

Standard 2—*Teachers should be skilled in developing assessment methods appropriate for instructional decisions*.

Standard 3—*The teacher should be skilled in administering, scoring and interpreting the results of both externally produced and teacher-produced assessment methods.*

Standard 4—*Teachers should be skilled in using assessment results when making decisions about individual students, planning teaching, developing curriculum, and school improvement*.

Standard 5—*Teachers should be skilled in developing valid pupil grading procedures that use pupil assessments*.

Standard 6—*Teachers should be skilled in communicating assessment results to students, parents, other lay audiences, and other educators*.

Standard 7—*Teachers should be skilled in recognizing unethical, illegal, and otherwise inappropriate assessment methods and uses of assessment information*.

The *Standards* acknowledge and specify the importance of teacher education and professional development in the area of classroom assessment (Brookhart, 2001). All 7 standards apply to teachers' development and use of classroom assessments of instructional goals and objectives that form basis for classroom instruction. Standards 3, 4, 6, 7 also apply to large-scale assessment, including administering, interpreting, and communicating assessment results, using information for decision making, and recognizing unethical practices (Brookhart, 2001).

Research on Assessment Literacy and "The Standards"

Numerous research studies have been conducted over the past 10 years that have addressed one or more of the seven *Standards* (Brookhart, 2001). However, only one (Plake, 1993) addressed *all* teacher competencies—as specified by *The Standards*—for inservice teachers. Additionally, one other study (Campbell, Murphy, & Holt, 2002) attempted to apply *The Standards* to groups of undergraduate preservice teachers.

In 1991, a national study was undertaken in order to measure teachers' assessment literacy (Plake, 1993). *The Standards* were used as a test blueprint for the development of the survey instrument used in the study. The survey in-

strument (the *Teacher Assessment Literacy Questionnaire*) consisted of 35 items (5 per standard). Items were developed as application-type questions that were realistic and meaningful to teachers' actual practices. The instrument went through extensive content validation and pilot testing. A representative sample from around the country was selected to participate and a total of 98 districts in 45 states participated. There was a total usable sample of 555 surveys (Plake, 1993) and the KR–20 reliability for the entire test was equal to .54 (Plake, Impara, & Fager, 1993).

Teachers answered an average of slightly more than 23 out of 35 items correct. The teachers' highest performance occurred on Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments* (M = 3.96/5.00); the lowest performance occurred on Standard 6—*Communicating Assessment Results* (M = 2.70/5.00). On 10 of the 35 items, 90% or more of teachers answered the item correctly. These items addressed issues including selecting appropriate assessments, acceptable test taking behavior for standardized testing situations, explanation of the basis for a grade to a child's parent, and the recognition of unethical practices in standardized test administration. On 5 items, less than 30% answered correctly. Two of the five came from Standard 5—*Developing Valid Grading Procedures*. Only 13% answered correctly an item that focused on steps to increase reliability of a test score. The two remaining items with low performance addressed Standard 7—*Recognizing Unethical or Illegal Practices*.

A similar study, conducted by Campbell et al. (2002), attempted to apply the identical previously described assessment literacy instrument to undergraduate preservice teachers. The renamed *Assessment Literacy Inventory* (*ALI*) was administered to 220 undergraduate students following a course in tests and measurement. The course included topics such as creating and critiquing various methods of assessment, discussing ethical considerations related to assessment, interpreting and communicating both classroom and standardized assessment results, and discussing and evaluating psychometric qualities (i.e., validity and reliability) of assessments.

The data from the undergraduate preservice teachers exhibited a higher level of reliability (α = .74) than their inservice counterparts in the Plake et al. study (Campbell, Murphy, & Holt, 2002). The preservice teachers (M = 21) averaged two fewer questions answered correctly than did the inservice teachers (M = 23). Six items (numbers 5, 7, 22, 28, 31, and 35) demonstrated poor item discrimination values (< .20). The inservice teachers in the Plake et al. study scored higher than the preservice teachers on all but one standard (Standard 1—*Choosing Appropriate Assessment Methods*). The preservice teachers scored highest on Standard 1, whereas the inservice teachers scored highest on Standard 3. Both groups of teachers scored lowest on Standard 6—*Communicating Assessment Results*.

Purpose of the Study

My intent in this study was to investigate the concept of "assessment literacy" and attempt to measure it as delineated by *The Standards for Teacher Competence in the Educational Assessment of Students*. Specifically, the purposes of this study were: (1) to measure and describe the relative levels of assessment literacy for both preservice and inservice teachers, and (2) to statistically compare the relative levels of assessment literacy for these two groups. This is the first study that attempts to measure assessment literacy for both preservice and inservice teachers and statistically compare the results.

The specific research questions addressed in the study were:

- How does the assessment literacy of preservice teachers compare to the assessment literacy of inservice teachers?
- Are there any significant differences between the two groups?

Methods

Participants

During the fall of 2002, the researcher surveyed both preservice and inservice teachers with respect to their assessment literacy. The group of preservice teachers was comprised of 67 undergraduate students, all majoring in secondary education, at a midwestern university. At the time of data collection, they were enrolled in methods courses scheduled during the term preceding student teaching and had just completed a course in classroom assessment. The group of inservice teachers consisted of 197 teachers representing nearly every district and school in a three-county area surrounding the same institution. The schools were selected based on convenience due to their geographic location. All grade levels and content areas were represented in the final sample.

Instrumentation

Both groups of teachers were surveyed using an instrument titled the *Classroom Assessment Literacy Inventory*, or *CALI*, which was adapted from a similar instrument called the *Teacher Assessment Literacy Questionnaire* (Plake, 1993; Plake, Impara, & Fager, 1993). This inventory is based on the *Standards for Teacher Competence in the Educational Assessment of Students* (AFT, NCME, & NEA, 1990). The *CALI* consisted of the same 35 content-based items (five per standard) with a limited amount of rewording. The researcher assigned pseudonyms to represent the names of the teachers and changed word choice to improve clarity. Additionally, 7 demographic items were included. The items were grouped by Standard; Table 1 shows the alignment of items with their respective Standard.

The 35 items presented the respondents with assessment-related scenarios, followed by a question with a specific

correct answer. Each item had the same format featuring four options presented in a multiple choice format with one option being the correct response. The complete instrument can be viewed at the following URL: http://edhd.bgsu.edu/mertler/cali.html.

The original instrument has been shown to have reasonable reliability with both inservice teachers, $r_{KR-20} = .54$ (Plake, Impara, & Fager, 1993), and preservice teachers, $\alpha = .74$ (Campbell, Murphy, & Holt, 2002). Furthermore, the original instrument was subjected to a thorough content validation, including reviews by members of the National Council on Measurement in Education and a pilot study with and feedback from practicing teachers and administrators.

Procedures

Inservice teachers were sent the *CALI* in both paper and Web-based formats. Two weeks after the initial mailing of the paper version and posting of the Web-based version, teachers were sent a reminder about completing the instrument. The instrument was administered to the preservice teachers at the final class meeting in their classroom assessment course. They were informed that their individual decision about participation, as well as their individual score on the instrument, would in no way affect the grade received for the course.

Table 1
Alignment of The Standards with Respective CALI Items

Standard	Item Numbers
Standard 1 *Choosing Appropriate Assessment Methods*	#1, 2, 3, 4, 5
Standard 2 *Developing Appropriate Assessment Methods*	#6, 7, 8, 9, 10
Standard 3 *Administering, Scoring, and Interpreting the Results of Assessments*	#11, 12, 13, 14, 15
Standard 4 *Using Assessment Results to Make Decisions*	#16, 17, 18, 19, 20
Standard 5 *Developing Valid Grading Procedures*	#21, 22, 23, 24, 25
Standard 6 *Communicating Assessment Results*	#26, 27, 28, 29, 30
Standard 7 *Recognizing Unethical or Illegal Practices*	#31, 32, 33, 34, 35

Note: The *Classroom Assessment Literacy Inventory* (*CALI*) can be viewed at http://edhd.bgsu.edu/mertler/cali.html

Table 2
Demographic Characteristics of Inservice and Preservice Teachers Responding to the CALI

Demographic Characteristic	Level	Inservice Teachers ($n = 197$)	Preservice Teachers ($n = 67$)
Gender	Female	77%	43%
	Male	21%	57%
Teaching Level	Elementary	57%	—[a]
	Secondary	26%	—[a]
Education Level	Pre-Bachelors	0%	100%
	Bachelors	29%	0%
	Masters	67%	0%
Years of Experience	None	0%	100%
	1-5	16%	0%
	6-10	14%	0%
	11-15	17%	0%
	16-20	12%	0%
	21-25	22%	0%
	> 25	18%	0%

[a] Preservice teachers could not provide responses to the demographic item addressing teaching level.

Analyses

Descriptive analyses at the individual item level included frequencies and reliability analyses. Descriptive analyses were also conducted for the seven composite scores based on *The Standards*. Inferential analyses included *t*-test comparisons, evaluated at an α-level equal to .05, of the preservice to inservice teacher mean scores for each of seven composite scores, as well as the total score for the entire instrument. All analyses were conducted using SPSS (v. 11).

Results

One-hundred ninety-seven (N = 197) inservice teachers completed the instrument. Seventy-seven percent of the sample was female; 21% was male. With respect to teaching level, 57% of teachers in the sample reported that they taught at the elementary level and 26% indicated that they were secondary teachers. Over one-fourth (29%) had earned bachelors degrees and two-thirds (67%) had earned masters degrees. Finally, 16% reported having 1-5 years of teaching experience, 14% reported having 6-10 years of experience, 17% had 11-15 years, 12% had 16-20 years, 22% reported having 21-25 years, and 18% indicated that they had more than 25 years of teaching experience.

The sample of preservice teachers consisted of 67 students. The only demographic information available for this group consisted of the gender of each student, as the participants would not have been able to respond to the other previously listed demographic items such as educational level, years of experience, and so on. Forty-three percent of the preservice sample was female; 57% was male. The demographic characteristics for both groups are summarized in Table 2.

It is important to note that, although the sample sizes for both groups were not large, the demographic characteristics of each as reported here very closely resemble those of the entire population of teachers not only in the three county region, but also in the entire state of Ohio, as reported by the Ohio Department of Education. Therefore, it could be assumed that the two groups of teachers did in fact constitute representative groups.

Table 3
t-Test Results for Comparisons of Scores for Preservice[a] and Inservice[b] Teachers

Standard	Group	Mean[c]	*t*-statistic	*p*-value
Standard 1				
Choosing Appropriate Assessment Methods	Preservice	3.25	3.79*	<.001
	Inservice	3.74		
Standard 2				
Developing Appropriate Assessment Methods	Preservice	2.78	3.28*	.001
	Inservice	3.18		
Standard 3				
Administering, Scoring, and Interpreting the	Preservice	3.24	5.23*	<.001
Results of Assessments	Inservice	3.95		
Standard 4				
Using Assessment Results to Make Decisions	Preservice	2.67	4.36*	<.001
	Inservice	3.36		
Standard 5				
Developing Valid Grading Procedures	Preservice	2.06	–.03	.975
	Inservice	2.06		
Standard 6				
Communicating Assessment Results	Preservice	2.27	1.69	.093
	Inservice	2.57		
Standard 7				
Recognizing Unethical or Illegal Practices	Preservice	2.69	2.77*	.007
	Inservice	3.10		
Total Score	Preservice	18.96	4.85*	<.001
	Inservice	21.96		

[a] $n = 67$
[b] $n = 197$
[c] The mean score for each *Standard* ranges from a possible low score of 0 to a high score of 5 (indicating the average number of items per *Standard* answered correctly).
* $p < .01$.

Descriptive results for preservice teachers

Data resulting from the preservice teacher group (N = 67) demonstrated a reasonably good level of internal consistency reliability, α = .74. On average, preservice teachers answered slightly less than 19 out of 35 items correctly. Out of the seven competency areas, as delineated by *The Standards*, the highest overall performance for preservice teachers was found for Standard 1—*Choosing Appropriate Assessment Methods* (M = 3.25; maximum possible score = 5). The lowest performance was found for Standard 5—*Developing Valid Grading Procedure* (M = 2.06). The results for the preservice teachers on each of the seven standards are presented in Table 3.

On only 4 of the 35 items did 90% or more of the preservice teachers answer the item correctly. One item each came from Standard 1—*Choosing Appropriate Assessment Methods* and Standard 2—*Developing Appropriate Assessment Methods*; two items came from Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments*.

On five of the 35 items, 25% or fewer answer the item correctly. One item came from Standard 2—*Developing Appropriate Assessment Methods*; two items each came from Standard 5—*Developing Valid Grading Procedures* and Standard 7—*Recognizing Unethical or Illegal Practices*.

Descriptive results for inservice teachers

Data resulting from the inservice teacher group (N = 197) demonstrated a mediocre level of internal consistency reliability, α = .57. On average, inservice teachers answered slightly less than 22 out of 35 items correctly. Out of the seven competency areas, the highest overall performance for inservice teachers was found for Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments* (M = 3.95; maximum possible score = 5). The lowest performance was found for Standard 5—*Developing Valid Grading Procedures* (M = 2.06). The results for the inservice teachers on each of the seven standards are also presented in Table 3.

On 8 of the 35 items, 90% or more of the inservice teachers answered the item correctly. Two items each came from Standard 1—*Choosing Appropriate Assessment Methods*, Standard 2—*Developing Appropriate Assessment Methods*, Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments*, and Standard 7—*Recognizing Unethical or Illegal Practices*.

On six of the 35 items, 25% or fewer answered the item correctly. One item came from Standard 2—*Developing Appropriate Assessment Methods*; three items came from Standard 5—*Developing Valid Grading Procedures;* and two items came from Standard 7—*Recognizing Unethical or Illegal Practices*.

Comparative results for the two groups of teachers

Standard and total scores for the two groups of teachers were compared by conducting independent-samples t-tests (α = .05). Examination of the results revealed that significant differences existed between the two groups for scores on 5 of the 7 Standards, as well as for the total scores. In all cases where there were significant differences, the inservice teachers scored significantly higher, meaning they were more assessment literate than their preservice counterparts. The largest discrepancies were found for Standard 3, the total score, and Standard 4, respectively. For Standard 3, the inservice teachers scored significantly higher (M = 3.95, SD = .95) than the preservice teachers (M = 3.24, SD = 1.00), $t(262) = 5.23$, $p < .05$, two-tailed. For the total score, the inservice teachers scored significantly higher (M = 21.96, SD = 3.44) than the preservice teachers (M = 18.96, SD = 4.65), $t(262) = 4.85$, $p < .05$, two-tailed. For Standard 4, once again the inservice teachers scored significantly higher (M = 3.36, SD = 1.08) than the preservice teachers (M = 2.67, SD = 1.19), $t(262) = 4.36$, $p < .05$, two-tailed. Significant differences were also found for Standards 1, 2, and 7. There were no significant differences found between the groups for Standards 5 and 6. Interestingly, both groups performed the poorest—and at the same exact level—on Standard 5. The results of all t-tests are shown in Table 3.

Discussion

Many of the results of this study parallel those of an earlier study (Plake, 1993; Plake, Impara, & Fager, 1993) that used the original version of the instrument and focused on the assessment literacy of inservice teachers. With respect to overall performance on the 35 items, the average score was equal to 22 items answered correctly—quite similar to the average score of 23 obtained by Plake (1993). In the earlier study, the highest mean performance for a given competency area was on Standard 3—*Administering, Scoring, and Interpreting the Results of Assessments*; the lowest performance was on Standard 6—*Communicating Assessment Results*. In the present study, the highest mean performance was also on Standard 3; the lowest was on Standard 5—*Developing Valid Grading Procedures*. Reliability analyses also revealed similar values for internal consistency (α = .54 and .57 for the original study and the study at hand, respectively).

The results for the preservice teachers also reflected those from a recent study, which also used the original instrument but collected data from preservice teachers (Campbell, Murphy, & Holt, 2002). In that study, the highest mean performance was on Standard 1—*Choosing Appropriate Assessment Methods*; the lowest performance was on Standard 6—*Communicating Assessment Results*. In the present study, the highest mean performance was also on Standard 1; the lowest was on Standard 5—*Developing Valid Grading Procedures*. Reliability analyses revealed identical values for internal consistency (α = .74 for both the original study and the study at hand).

Comparisons between preservice and inservice teachers of the seven competency area scores revealed signifi-

cant differences on five of the seven areas, as well as on the total scores. In all cases where significant differences were found, the inservice teachers scored higher than their preservice counterparts. Both groups demonstrated their poorest performance on Standard 5—*Developing Valid Grading Procedures*, followed closely by Standard 6—*Communicating Assessment Results*. It is reasonable to expect that practical experience with student assessment in classroom settings would result in teachers possessing greater knowledge of and superior abilities to apply various assessment terms and concepts, as compared to their preservice counterparts. Participants' performances in this study on five of the seven standards, as well as on the total *CALI* score, support this assertion. However, it is a bit alarming that the inservice teachers did not demonstrate this expected higher level of understanding and application skills on two of the Standards, namely Standard 5—*Developing Valid Grading Procedures* and Standard 6—*Communicating Assessment Results*. These are two very critical Standards—so much so that Brookhart, in two of her papers, chose to focus on improving the instruction provided to preservice teachers on only these two competencies (1998, 1999b). She believes that instruction provided in these areas are typically "simplified psychometric content" as opposed to the *application* of those concepts to what teachers are actually called upon to do in their classrooms. Therefore, teachers are not taught how to apply the theories and principles behind valid grading procedures and communication of results to the classroom setting.

Another possible reason—somewhat related to the first—for this lack of difference between the two groups may be due in part to the fact that both Standards address knowledge and skills that even the most experienced teachers struggle with. For example, a portion of Standard 5 states:

> *Teachers will understand and be able to articulate why the grades they assign are rational, justified, and fair, acknowledging that such grades reflect their preferences and judgments. Teachers will be able to recognize and to avoid faulty grading procedures such as using grades as punishment. They will be able to evaluate and to modify their grading procedures in order to improve the validity of the interpretations made from them about students' attainments.*

Brookhart (1993) studied teachers' grading practices and discovered that teachers apply grading scales differently for students depending on their ability levels. She also found out that many teachers continue to award missing work a grade of zero, indicating punitive consequences, even if it meant that a student would fail a course. Furthermore, she concluded that teachers' grading is often a miscellany of attitude, effort, and achievement (1993), and these factors may not always be applied equally across the board to grades assigned to students. With respect to this lack of difference in performance between preservice and inservice teachers in this study, it could be the case that competencies related to grading systems and communicating assessment results are not acquired through practice and experience in the manner that some other competencies such as selecting appropriate assessment methods or developing appropriate assessment methods.

It is important to recognize that the low reliability coefficients—especially that for the group of inservice teachers—serves as a substantial limitation to the results of this study. An apparent lack of reliability in the data resulting from the administration of this particular instrument limits the extent to which the results of this study may be generalized to other groups of both preservice and inservice teachers. At a minimum, it is recommended that the *CALI* be substantially revised—if not completely rewritten—prior to being used in future research studies as a means of measuring teachers' assessment literacy.

Although these low reliabilities are somewhat problematic in terms of generalizing the results of this study, it is also imperative to recognize that the *CALI* was merely a *slightly* modified version of a previously utilized instrument. However, these slight modifications did not result in meaningful—and, in some cases, *any*—differences between the psychometric qualities of the original and revised versions of the instrument. With respect to measuring preservice teachers' assessment literacy, the original instrument and its revised version resulted in identical values for internal consistency reliability. The reliability resulting from the inservice teachers' data in this study was somewhat lower than that for the comparable group of teachers in the original study.

Research has shown that traditional teacher preparation courses in classroom assessment are not well matched with what teachers need to know for classroom practice (Schafer, 1993). It is likely that one course in assessment and measurement may truly be insufficient to cover everything that secondary teachers need to know. The traditional focus of these teacher prep assessment courses has historically been on large-scale standardized testing (Schafer, 1993), although this trend is changing. This changing trend is evidenced by Popham's (2000) call to stop the "erroneous and educationally harmful appraisal of instructional quality via standardized tests…" (p. 15). Further evidence can be gleaned through a brief examination of older and newer classroom assessment textbooks. Older textbooks—for example, Ebel and Frisbie, 1991, and Hopkins, 1998—tend to contain more chapters on standardized testing (3 of 18 chapters, and 3 of 15 chapters, respectively) and fewer on classroom assessment techniques, and in particular, methods of alternative assessment (1 of 18 chapters, and none of 18 chapters, respectively). Newer textbooks on classroom assessment demonstrate a reversal of this tendency. For example, McMillan's (2001a) textbook contains 1 of 13 chapters on standardized testing and 4 chapters on alternative assessment. Similarly, Mertler's (2003) text includes 1 of 13 chapters on standardized testing and 3 chapters on alternative assessment techniques. However, it is also important that the current administration's emphasis on standardized testing, as out-

lined in the *No Child Left Behind Act of 2001* (U.S. Department of Education, n.d.), not be overlooked. Teachers must be proficient in all of these areas of assessment.

The fact that courses in classroom assessment are not well matched with what teachers need to know for classroom practice is made even more troublesome when considering that many teacher preparation institutions and states do not even require a course in assessment (Campbell, Murphy, & Holt, 2002; Shafer, 1993). As of January 1998, only 15 states had teacher certification standards that required competence in assessment, and 10 states explicitly required a course in assessment; however, 25 states held no expectation of competence in assessment (Stiggins, 1999b). The majority of states and institutions simply embed assessment content into other teacher education coursework; students then learn about assessment and measurement from instructors who typically possess no expertise in educational assessment (Quilter, 1999).

However, instruction from individuals with expertise in educational assessment may not be enough. It may be more important, not that the instruction is presented by experts, but that these measurement specialists better understand the reality of K–12 classrooms. Specifically, it is important that they understand that assessment is an integral component of instruction and goals for student learning (McMillan, 2001; Pilcher, 2001). Teachers have indicated that they are more concerned with the day-to-day issues related to the application of assessment processes and less with fundamental measurement principles (Rogers, 1991). Hopefully, then, those who teach courses in assessment and measurement can teach preservice teachers to see this vital connection between assessment and instruction, making assessment more applicable to their views of teaching.

With respect to the concept of assessment literacy, Popham (2003) has called for an increased effort among the measurement community at large to promote assessment literacy on the part of parents, policymakers, practitioners, teachers, administrators, and counselors. A more assessment literate citizenry is less likely to tolerate misuse of assessment and, specifically, assessment results. Stiggins (1995) offers several guiding principles for educators to follow in order to promote assessment literacy. These guiding principles suggest that educators should:

- start with a clear purpose for assessment,
- focus on achievement targets,
- select appropriate assessment methods,
- adequately sample student achievement, and
- avoid bias and distortion.

Stiggins (1995) continues by stating that these standards of assessment quality are not negotiable, nor is the expectation that they be met every time educators assess student achievement. However, research shows that these standards are seldom met—due to fear of assessment and evaluation, insufficient time to assess properly, or public perceptions of assessment practices.

Recommendations

The day-to-day work of classroom teachers is multifaceted, to say the least. However, none of these daily responsibilities is more important—or more central—to the work of teachers than that of assessing student performance (Mertler, 2003). Previous studies have reported that teachers feel—and actually *are*—unprepared to adequately assess their students (e.g., Mertler, 1999; Plake, 1993). They often believe that they have not received sufficient training in their undergraduate preparation programs in order to feel comfortable with their skills in making assessment decisions. This, coupled with the fact that inservice teachers outscored preservice teachers on nearly every subscale in this study, may raise substantial questions about the usefulness—or, perhaps more importantly, the *appropriateness*—of assessment training in preservice teacher education programs.

Another question worthy of consideration—and further research—is whether or not a majority of assessment training is an "on-the-job" type of training. In other words, are assessment skills best learned through *classroom experience* as a teacher, perhaps once teachers can place the notion of "assessment" in a specific context, as opposed to learning them as an undergraduate? Does undergraduate training provide the necessary foundation for this on-the-job training? At a minimum, the present study highlights specific competency areas—namely, developing valid grading procedures and communicating assessment results—where both preservice and inservice teachers need remediation and additional support.

Based on the findings of this study, as well as on the questions posed above, several recommendations for practice and research are offered here. It is the belief of this author that assessment training at *both* the preservice and inservice levels is crucial. Additionally, this belief is not meant to take away from the valuable knowledge and skills gained through practical classroom experience. Therefore, an initial recommendation is that, although the appropriateness of preservice training in classroom assessment was questioned above, it is certainly not being advocated that the profession abandon this training. On the contrary, preservice training of teachers in the concepts and techniques of classroom assessment is critical. This should be enhanced through thoughtful examination and research into the knowledge and skills that these teachers will need to possess once they assume the responsibilities for their own classrooms and students.

Second, even though assessment training for preservice teachers is important, ongoing training on various topics related to classroom assessment should be an essential component of any district's program of professional development for its teachers. Administrators at both the district and individual building levels need to stress to their teachers the

importance of sound assessment practice and the professional benefits of being assessment literate. Furthermore, they must provide sufficient opportunities for those teachers to improve their understanding and application of assessment techniques.

Third, future research should investigate various reasons behind the apparent discrepancy between the assessment literacy of preservice teachers and that of inservice teachers. The inservice teachers in this study appeared to be significantly more literate than their preservice counterparts with respect to (1) administering, scoring, and interpreting the results of assessments, and (2) using assessment results to make decisions. Additionally, the inservice teachers scored highest in the skill area of administering, scoring, and interpreting the results of assessments; whereas, the preservice teachers scored highest on their abilities to choose appropriate assessment methods. Examination of these differences and the relative impact of preservice training versus "on-the-job" learning certainly seems warranted.

Finally, the measurement community must take on the responsibility of improving assessment literacy among *all* educational stakeholders. These stakeholders include—but are not limited to—administrators, teachers, parents, policymakers, journalists, and the general public. The ability to assess student performance—and to do so in appropriate, valid, and reliable ways—is arguably one of the most important aspects of the job of teaching.

References

American Federation of Teachers, National Council on Measurement in Education, & National Education Association. (1990). *The Standards for Competence in the Educational Assessment of Students*. Retrieved July 22, 2003, from http://www.unl.edu/buros/article3.html

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, *30*, 123-142.

Brookhart, S. M. (1998). *Teaching about grading and communicating assessment results*. Paper presented at the annual meeting of the National Council on Measurement in Education, San Diego, CA. (ERIC Document Reproduction Service No. 419838).

Brookhart, S. M. (1999a). *The art and science of classroom assessment: The missing part of pedagogy*. Washington, DC: ERIC Clearinghouse on Higher Education and Office of Educational Research and Improvement.

Brookhart, S. M. (1999b). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice*, *18*(1), 5-13.

Brookhart, S. M. (2001). *The Standards and classroom assessment research.* Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Dallas, TX. (ERIC Document Reproduction Service No. ED451189).

Campbell, C., Murphy, J. A., & Holt, J. K. (2002, October). *Psychometric analysis of an assessment literacy instrument: Applicability to preservice teachers*. Paper presented at the annual meeting of the Mid-Western Educational Research Association, Columbus, OH.

Center for School Improvement and Policy Studies, Boise State University. (n.d.). What is assessment literacy? In *Assessment literacy*. Retrieved July 24, 2003, from http://csi.boisestate.edu/al/

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed.). Boston: Allyn & Bacon.

McMillan, J. H. (2001a). *Classroom assessment: Principles and practice for effective instruction* (2nd ed.). Boston: Allyn & Bacon.

McMillan, J. H. (2001b). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice*, *20*(1), 20-32.

Mertler, C. A. (1999). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education*, *120*(2), 285-296.

Mertler, C. A. (2003). *Classroom assessment: A practical guide for educators*. Los Angeles: Pyrczak.

North Central Regional Educational Laboratory. (n.d.). *Indicator: Assessment*. Retrieved July 24, 2003, from http://www.ncrel.org/engauge/framewk/pro/literacy/prolitin.htm

O'Sullivan, R. G., & Johnson, R. L. (1993). *Using performance assessments to measure teachers' competence in classroom assessment*. Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA. (ERIC Document Reproduction Service No. 358156).

Paterno, J. (2001). Measuring success: A glossary of assessment terms. In *Building cathedrals: Compassion for the 21st century*. Retrieved July 24, 2003, from http://www.angelfire.com/wa2/buildingcathedrals/measuringsuccess.html

Pilcher, J. K. (2001). *The standards and integrating instructional and assessment practices*. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Dallas, TX. (ERIC Document Reproduction Service No. ED451190).

Plake, B. S. (1993). Teacher assessment literacy: Teachers' competencies in the educational assessment of students. *Mid-Western Educational Researcher*, *6*(1), 21-27.

Plake, B. S., Impara, J. C., & Fager, J. J. (1993). Assessment competencies of teachers: A national survey. *Educational Measurement: Issues and Practice*, *12*(4), 10-12, 39.

Popham, W. J. (2000). The mismeasurement of educational quality. *The School Administrator*, *11*(57), 12-15.

Popham, W. J. (2003). Seeking redemption for our psychometric sins. *Educational Measurement: Issues and Practice*, *22*(1), 45-48.

Quilter, S. M. (1999). Assessment literacy for teachers: Making a case for the study of test validity. *Teacher Educator*, *34*(4), 235-243.

Rogers, W. T. (1991). Educational assessment in Canada: Evolution or extinction? *The Alberta Journal of Educational Research*, *37*(2), 179-192.

Schafer, W. D. (1993). Assessment literacy for teachers. *Theory Into Practice*, *32*(2), 118-126.

Stiggins, R. J. (1995). Assessment literacy for the 21st century. *Phi Delta Kappan*, *77*(3), 238-245.

Stiggins, R. J. (1999a). Are you assessment literate? *The High School Journal*, *6*(5), 20-23.

Stiggins, R. J. (1999b). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, *18*(1), 23-27.

U.S. Department of Education. (n.d.). *The facts about measuring progress—No Child Left Behind*. Retrieved June 12, 2003, from http://www.nclb.gov/start/facts/testing.html

Reviewers for 2005

Savilla Banister
Bowling Green State University

Tom Cody
Western Illinois University

Sharon Damore
DePaul University

Cathleen Doheny
University of West Georgia

Ros Duplechain
University of West Georgia

Mark Earley
Bowling Green State University

Eva Floyd
University of Kentucky

Cynthia P. Fraga-Cañadas
The Ohio State University

Myrna Gantner
University of West Georgia

Cher Hendricks
University of West Georgia

Marie Holbein
University of West Georgia

Sema A. Kalaian
Eastern Michigan University

Barbara Kawulich
University of West Georgia

Michael Kerstetter
The Ohio State University

Jenny Kilgore
Miami University

Charles E. Kline
Purdue University

Dorothy Kupsky
The Ohio State University

Terri Lane
University of Dayton

Bambi Bethel Leitshuh
University of Missouri-St. Louis

Robert H. Lombard
Western Illinois University

Eleni Makris
Northeastern Illinois University

Gail Marshall
University of West Georgia

Julia M. Matuga
Bowling Green State University

Sharon McNeely
Northeastern Illinois University

Harry Morgan
University of West Georgia

Glenda Moss
Indian University-Purdue University Fort Wayne

Timothy James Murnen
Bowling Green State University

Noriko Nagata
The Ohio State University

Sungworn Ngudgratoke
Michigan State University

Betul Ozkan
University of West Georgia

Lynn M. Papenfus
Kent State University

Jonnie Jill Phipps
University of Akron

Jean Pierce
Northern Illinois University

Hema Ramanathan
Butler University

Ronald S. Reigner
University of West Georgia

Laura Renzi
The Ohio State University

Elaine Roberts
University of West Georgia

Kenneth Royal
University of Kentucky

Mary Beth Slone
University of West Georgia

Ruslan Slutsky
University of Toledo

Janet Strickland
University of West Georgia

Kathleen Sullivan-Brown
University of Missouri

Sharon Valente
Ashland University

John vonEschenbach
University of West Georgia

Kathryn Wiggins
DePaul University

E. Jane Williams
The Ohio State University

James P. Wilson
University of Dayton

Lesson Plan Design for Facilitating Differentiated Instruction[1]

Pamela Hudson Baker
George Mason University

Louise Conn Fleming
Ashland University

Abstract

This study examines the role of lesson plan design in helping teacher candidates to meet diverse learning needs. The authors studied student teachers who had been taught to use a differentiated lesson plan to see if they were differentiating in their planning, assessment, and instruction. They concluded that, indeed, lesson plan design does facilitate differentiation but that instruction, coaching, and examples are also needed.

In the Middle Grades classroom, diverse learners are seemingly becoming the norm rather than the exception. Finding ways to meet the varied needs of all learners is a challenge for even the most experienced educator. This study examines the role of lesson plan design in helping teacher candidates to meet diverse learning needs. Specifically we examined differentiation, which is using different targets, instruction, and assessment to address both content standards and learner differences. It should be noted that differentiating is not limited to students whose needs have been officially identified. It also includes any students who have mastered given content or who need further support.

Background Information

In the state of Ohio, where this research was conducted, recent changes in teacher credentialing created an initial teacher license for Middle Childhood teachers, grades four to nine. In addition to education courses, these candidates are required to complete a program of study in two content areas. When our teacher education faculty at Ashland University were preparing to address this change, they created a middle childhood program team that developed the new licensure program to match the National Middle School Association's recommendations for middle grades teachers. For example, *Turning Points 2000* (Jackson & Davis, 2000) recommends teaming of teachers, differentiation in the classroom, Backward Design, integration of assessment and instruction, and flexible scheduling. In order to be effective in preparing teachers to function within these structures, the Ashland team decided to model them.

Program Context

The middle childhood program revolves around two blocks who integrate methods, assessment, and field experience. Block I presents middle grades philosophy, and Block II presents content methods, assessment, differentiation, and evaluation. In these blocks we teach and model middle level principles, hoping to help our students to develop attitudes and skills for building an environment of effective learning for all students. One of our goals is to provide some realistic strategies for meeting diverse learning needs in a middle grades classroom. Successful programs for educating diverse learners can often be found to include interactive teaming among stakeholders with expertise in different areas supported by a core set of values (Thomas, Correa, & Morsink, 2001; Walther-Thomas, Korinek, McLaughlin, & Williams, 2000; Wormeli, 2001).

We teamed with a colleague to co-teach what we call the "Junior Block," consisting of two courses: Middle Grades Methods and Assessment II and Middle Grades Field Experience II. The courses meet for 7 hours, two days a week, and feature class sessions and a field experience. In class we teach planning, assessment, instruction, differentiation, and evaluation. Assisted by adjuncts, we also coach the students while they complete a field experience at a middle school. The students are expected to carry the methods and assessment techniques into their field settings. One of our goals for the course is to have our students differentiate planning, instruction, and assessment in order to better address the needs of their learners. During this process, we developed a lesson plan to emphasize Wiggins and McTighe's (1998) "Backward Design" techniques in conjunction with a focus on meeting the instructional needs of differing learners. We teach the teacher candidates how to differentiate, require it in their lesson plans, and help them implement it.

Lesson Planning

Both modeling by instructors and mentorship by practicing teachers are of great importance in apprenticing preservice teachers into effective planning, teaching, and assessment. Lesson planning is a difficult process. It is loved by professors and hated by both students and teachers who declare they never plan lessons! We tell our students that a lesson that is not carefully planned is a lesson that goes nowhere. Instructional strategies and assessment must be designed to facilitate the learning of clear, worthwhile

[1] This article was accepted for publication by the previous editorial team.

objectives. Teaching the lesson planning process involves making the parts of lesson planning clear, as well as indentifying the interrelatedness of the sections. Teacher candidates must learn to see both the global aspect and the individual parts (Baylor, Kitsantas, & Chung, 2001). Taylor (2000) also notes that modeling and reflection are crucial in teaching lesson planning.

Researchers (Giebelhaus, 1999; Giebelhaus & Bowman, 2002; Salzman 1999) have explored the effects of mentoring on preservice teachers. Giebelhaus and Bowman (2002) found that the mentoring that student teachers receive is the primary link between theory and practice, yet, "The kind of mentoring that preservice teachers receive, is, at best, marginally effective" (p. 247). These researchers recommend using the Praxis III/Pathwise framework for selection and training of mentors, as well as for the mentor and student to work together. The role of the expert (cooperating teacher) should be to help the novice (prospective teacher) by providing "effective and appropriate feedback and modeling," (p. 249) providing the opportunity for reflection and discussion. They found that "Prospective teachers who have been trained, using a common framework for discussion, demonstrate more complete and effective planning, more effective classroom instruction, and greater reflectivity on practice than those whose cooperating teachers received only an orientation" (p. 250). Colleges and universities, they maintain, must give primary consideration to effective mentorship preparation of those teachers who agree to mentor their preservice teachers (Giebelhaus & Bowman, 2002).

Differentiation

Traditional classrooms view learner differences as deficits to be fixed, often neglecting strengths and talents that learners bring with them (Levine, 2003). Differentiation, by contrast, recognizes student differences and plans, instructs, and evaluates with difference in mind. Tomlinson (1995) defines the differentiated classroom as a place where "the teacher plans and carries out varied approaches to content (what students learn), process (how they learn), and product (how they demonstrate learning) in anticipation of a response to student differences in readiness, interest, and learning needs" (p. 10). Whereas the deficit model asks for labels, remediation, and forces students into the teacher's agenda, differentiation identifies interests, strengths, and innate motivation in order to adapt the agenda to learner needs (Tomlinson, 2003). A differentiated classroom does not lower standards, but rather modifies its demands to suit individual learners: No student should be permitted to work, study, or produce less than his or her peers. But we should not insist that everyone put forth identical output (Levine, 2003).

Thus, teachers who differentiate have ongoing assessment, including pre-assessment and learner self-assessment. This technique allows teachers to scaffold new information into existing information and avoids boredom on the part of children who already understand the concept (Brimijoin, Marquissee, & Tomlinson, 2003; Franklin, 2004), thus facilitating learning on the part of children at many levels and diverse abilities. Such a differentiated environment promises to be a key for success as we embark upon education in the twenty-first century, and our goal was to prepare our students to use this key in their teaching.

Methods

The dilemma we faced, which led to this study, was ascertaining the level to which the students actually exhibited the desired skills after they left our block. We needed evidence to indicate how successful we had been in building the ability and willingness to differentiate and to help us to do a better job in the future. To gather this evidence, we studied preservice teachers who we taught in the previous year. Because we were evaluating a specific process in teacher education for the purpose of understanding our program better, we used a qualitative design with the purposive selection of participants (Wiersma, 2000). They were selected because they had completed our junior methods block and to assure confidentiality of all participants and their cooperating schools, we used pseudonyms in place of their actual names. Students were told only that we were doing a study to help us to improve the Junior Block. They were asked to allow each of us to observe them once; to provide six lesson plans, three from their junior experience and three from student teaching; and later to meet with us for an interview. Because we did not want to bias the results, we did not tell the students that we were looking for evidence of differentiation.

Description of Informants and Contexts

Abrams Middle School was selected because of its reputation for exhibiting the features of a quality middle childhood program: teaching teams, support for students, engaging learning activities, and use of alternative assessment. Abrams, a school enrolling students in grades 6 through 8, is located in a suburb with above average family incomes. Both students and faculty were friendly, happy to accommodate our need to visit, and helpful to give directions or answer questions. Elaine and Kyle did their student teaching at this school.

Elaine was a traditional student teacher who was a very goal-directed student and a delight to have in the Junior Block. By traditional student teacher, we mean that she entered university after graduating from High school and was in her fourth year of collegiate study. She was energetic, caring, and always willing to listen to another person's viewpoint. Partly as a result of these attributes, she was offered a teaching position at Abrams before completing her Student Teaching experience.

Kyle was also a traditional student teacher. He was conscientious, and in the Junior Block tended to be serious about his studies both in class and in the field. His field experience placement was mediocre, partly because he emulated the level of preparation his cooperating teacher exhibited. Since the cooperating teacher was a veteran teacher, he

placed little effort into planning his lessons and Kyle did the same. Therefore, Kyle had to be prodded to do the necessary work to get background information for his lessons. By the time we observed Kyle's student teaching, however, he clearly felt comfortable with his students and prepared his lessons thoroughly.

Benedict Middle School is located in a below average income community and many of the students are from farming families. Recently a division has come about in the community as higher income families move into the rural area. This influx has caused a rift between the people whose families have lived there for generations and the newcomers. The faculty and secretaries at the school did not particularly welcome us, but they did not mind our being there. Lisa and Diana, both traditional student teachers, did their student teaching at this school.

In her Junior Block, Lisa was always very serious about learning, conscientious in all of her lesson preparation, and goal-oriented. Her cooperating teacher and her students enjoyed working with her. She was kind and caring, but she also maintained an appropriate distance from her students. She worked extremely hard, putting in long hours, preparing her lessons, bringing in extra information to her students, and being imaginative in her presentation. As a student teacher she did the same thing. The administrators, her cooperating teacher, and her students valued her.

As a junior Diana never went beyond the requirements. We allowed her to redo assignments that were not right and coached her in her Junior Block field experience. She was uncomfortable and unprepared, but with extensive coaching from her cooperating teacher and her supervisor, she rose to the level of adequate. As a student teacher, she appeared to dislike teaching, and, once again, she prepared only to meet the requirements. That is, she knew the minimum of how to instruct the students, but she did not have any background understanding either of the material or it's fit into the curriculum.

Charles Middle School is located in a small city surrounded by farms. The community, and thus the school, have a wide diversity of income and education levels. The grade seven and eight teachers at the school are fairly friendly and were willing for us to visit. They did not go out of their way to welcome us, but they did not seem to mind our presence.

Nancy was the only student teacher at this school. In the Junior Block she seemed to like teaching. She was not very confident but she tried to teach meaningfully and to use methods that would engage her students. She did not, however, put a lot of time into lesson preparation and into informing herself with background information. As a student teacher she appeared to do the same. She seemed capable of running a class but her methods were fairly traditional, and she had not gone out of her way to be prepared with more information than she was actually teaching.

Procedures for Data Collection and Analysis

The goal of data collection was to find evidence of the level of differentiation carried from the junior block to the student teaching setting. Therefore, a variety of methods were utilized by the researchers. Each of the five informants was asked to provide copies of three lesson plans from the junior block experience and three lesson plans from the current student teaching setting. A review of these documents allowed a comparison of planning for differentiated instruction in content, process, and product areas. The junior level lesson plan directly required differentiation whereas the student teaching lesson plan did not. In addition to the written materials, each informant was observed twice, once by each researcher. The observations were scheduled during the last two weeks of the student teaching experience in order to see the participants at the maximum level of experience prior to licensure. Finally, each student was sent a list of four interview questions via e-mail at the conclusion of the student teaching placement. The questions were:

(a) What is differentiation as you understand it?

(b) Did you differentiate as a student teacher?

(c) Did you feel prepared to differentiate?

(d) What do you think of the junior block lesson plan form?

The collection and review of data was undertaken by two researchers, one with a general education perspective and with a special education perspective. Each researcher examined the data independently and then compared findings to jointly identify themes.

Evidence of Trustworthiness

In any qualitative design, strategies can be used to enhance the value of the information gained (Johnson, 1999). Even though this study was conducted for the highly focused purpose of improving our own teacher preparation program, we were cautious to consider trustworthiness in hopes that this information would be useful to other educators. Krefting (1999) suggested four key techniques that can be used to establish a sense of trustworthiness in qualitative research. They include: (a) credibility, (b) transferability, (c) dependability, and (d) confirmability. Within each of these strategy areas, a variety of criteria may be utilized to support what has traditionally been viewed as the validity and reliability of the study. Credibility requires a sufficient level of time be spent with the informants for any patterns to emerge. The informants used in this study were known to the researchers in a variety of contexts (i.e., classes, junior field, and student teaching) over a period of one year. The informants were accustomed to observation from university personnel and had previously participated in discussion of their experiences with both researchers.

Credibility, dependability, and confirmability can each be enhanced by the use of triangulation. The documents gathered from the informants provided triangulated data sources

for both the junior level lesson plan and the student teaching lesson plan. Methods triangulation was also employed by adding observation and interview components to the document review. Member checking was not utilized until the interview phase of the data collection because informants were unaware, up to that time, that differentiation was the focus of the study. Peer review was utilized by sharing these findings with a colleague who has utilized this lesson format to obtain his reactions to the interpretations and recommendations we made. The researchers also used reflection as a vehicle to support greater credibility and confirmability. Each of us is aware of the professional and personal bias that we bring to any situation. We have discussed and documented the role of our backgrounds on the perceptions that we bring to the study. It is because of our differences that greater balance in the results can be inferred. Finally, regarding transferability, we have attempted to provide enough information about our program, the contexts, and the informants for readers to ascertain the applicability of our findings to their specific situation.

Results

The information collected came from a combination of document reviews, observations, and follow-up interviews with each participant. A summary of each segment of the data can be found in the following sections.

Document Review

Opportunities to plan for differentiation appear on the lesson plan template for the Junior Block. To support students in their lesson planning, our Junior Block team emphasizes the importance of planning differentiated learning experiences for children at various learning levels: core, enriched, and fundamental. Even though the lesson plans reviewed from the Junior Block actually included minimal evidence of differentiation, certain aspects of this philosophical framework did emerge. For instance, instruction often included different tasks for students at fundamental levels of readiness, but not for students at the enrichment level. All of the field experience sites included an intervention specialist (special educator) on the teaching teams, and it was common to have this teacher in the classroom during instruction. When this teacher with the special education endorsement was not present, other classroom teachers worked with the students who were experiencing difficulty to make sure they were on track with the lesson. The students who had already achieved lesson goals were not treated differently in any of the Junior Block classrooms, and it was rare for our teacher candidates to plan for or to treat them differently in their lessons. In the lesson plans it was common for all of the sections to be completed as if the teacher candidates were differentiating, but it was clear that all students would be doing the same things.

The lesson plan templates used for student teaching do not include any section specifically for differentiation. This is the format that the student teachers used, and none of the teacher candidates showed that they had planned for differentiation. Although there is a section for accommodations, most student teachers regarded this section as necessary to describe teaching activities for children with physical disabilities. Specific patterns found for each informant can be found below:

Elaine's lesson plans in the Junior Block were filled with activities. For one class she used a game of Jeopardy!; for another she taught fractions using candy; for another she taught the concept of negative numbers by using pictures of a mountain and an ocean; and for another she began using a K-W-L activity. All of these activities showed attempts to engage all students in learning but minimal differentiation was noted. Elaine did write differing elements into the process section of her plan for a math lesson that showed her intent to differentiate. When reviewing long division, she would have some students describe how to move the decimal, some students explain the steps of long division, and still other students create another word problem.

For student teaching she still used activities to engage her students, but her written lesson plans still showed no differentiation. She taught math and language arts. Her math lesson plans included a game of Jeopardy!, a pretest, a study guide, manipulative blocks, and student study teams. Her language arts lessons included writing stories to a prompt, writing descriptive paragraphs, using a semantic map, and having students conduct research on computers. For instance, she didn't write in her plan that she would provide different levels of questions for fundamental, core, and enriched learners. However Jeopardy! clearly allows the infusion of such levels. She explained that she relied on the intervention specialist to differentiate for the students on IEPs.

Kyle's Junior Block lesson plans included virtually no differentiation. His cooperating teacher did not differentiate and expected Kyle to follow his model of instruction, which was based on overheads and lectures. However, Kyle was aware that he should try to develop other teaching practices. He was concerned that he was not able to attempt alternatives in the same way as some of his classmates.

By contrast, Kyle's student teaching lesson plans were filled with engaging activities. He taught social studies and language arts. In social studies he donned an apron and "cooked" the Constitution, with students adding the ingredients. In language arts he used a Venn diagram to discuss characters, poetry stations, student power point presentations, and student-written "newspapers." He did not have his lesson plans clearly differentiated, but he did show evidence of infusing student interests and choices into the lessons. For instance, he planned for students to create ways to show key points from a story in a newspaper article. Kyle explained that he did additional work with students who had learning difficulties. He clearly infused some principles of differentiation and multiple intelligences in his actions. His rubrics for assessment were also well developed and clear.

Lisa's Junior Block lesson plans had no clear evidence of differentiation, partly because she deviated from the prescribed format. She did show a high level of preparation and a need to have a very sequential lesson plan. Upon a closer reading of her junior lesson plans, one finds that all of her lessons contain many levels of assessment and varied usages of multiple intelligences. In her junior field experience she taught language arts and used role-playing, multilevel questioning, predictions, writing, pictures, films, quotes, and students' experiences.

In her student teaching, Lisa continued to show minimal evidence of planning for differentiation and high levels of variety to engage multiple intelligences. She explained that all the students with IEPs were in one class and that she worked with the intervention specialist to tailor lessons for them. This practice provides some support for the students with disabilities, but from the lesson plans it appears as if the students who needed more of a challenge were getting no special treatment. While she read *The Diary of Anne Frank* to her students, she had them write journal entries, showed the movie, gave them a study guide and had students do outside projects for extra credit. These projects included discussing prejudice and segregation, designing a poster, drawing a hiding place, writing a character analyses from the viewpoints of other characters, and writing a letter to Mr. Frank. Therefore, with no apparent planning for enrichment, she challenged students at various levels. Her rubrics for assessment were also clear and she did infuse opportunities for students to choose activities and extra credit topics that would be of interest to them.

Diana's Junior Block lesson plans included some efforts to plan for differentiation, but she fell into several pitfalls. For instance, she often noted an accommodation such as extended time as a form of differentiated assessment. She also would use extra work upon completion of the core materials as a way of planning for enrichment, so she wasn't actually varying the learning experience itself, but rather just varying the parameters of the tasks. Her cooperating teacher mostly relied on the intervention specialist to work with the students who had special needs but never thought to challenge the ones who had already mastered learning outcomes. She taught language arts, and her university supervisor worked with her to think of ways to challenge students at different levels. Diana would typically read a book to her students and give the students a comprehension worksheet, but with some coaching by her supervisor she eventually had students doing research related to topics in the book. In her lessons she tried to engage her students by giving them imaginative writing activities, and one day she used a map activity related to the book. However she clearly misinterprets some of the key principles of differentiation.

Diana taught math during her student teaching experience and had the students dancing and doing tessellations. While this was a nice infusion of multiple intelligences, she had no differentiation or accommodations in her lesson plans, although she did accommodate for a student who was hard of hearing during the dance. The lesson plans in evidence for her student teaching experience showed far less depth of planning than did her previous efforts during her field experience as a junior. The students had no opportunities for choice or for variety within the experience. While they may have had some fun, the lesson plan laid the foundation for all of the students to experience the lesson in a singular way. Diana did not clearly establish an understanding of differentiation as a junior, and she did not show even incidental evidence of differentiation as a student teacher.

Nancy's Junior Block lesson plans for math included having students use manipulatives and work independently on sample problems. In language arts she was reading a story to her students and asking them questions regarding comprehension and characters. Although the lesson plan template sections were completed out for differentiation, clearly she intended all students to be doing the same things. She did plan for a little enrichment in part of her lesson by offering students an "opportunity to create their own problems." It was unclear whether this activity would be over and above the required elements or could be a vehicle for showing knowledge of multiplication.

During her student teaching experience Nancy taught math again. She used the Ashland University lesson plan template and did not show evidence of differentiation in writing. She had a cooperating teacher who did not use much variety and modeled this limited variety daily. Not surprisingly, Nancy's lessons were fairly traditional—homework review, activity, and homework assignment. She did, however, incorporate manipulatives and cooperative groups on occasion. However, her student teaching lesson plans lacked detail to such an extent that it was really hard to tell what would actually occur in the classroom. She didn't even plan for simple accommodations noting "accommodations will be addressed at that time."

Observations

Each of us observed each student once during their student teaching experience. The observations were completed near the end of the student teaching experience so the students had already received consistent support from their cooperating teachers and university supervisors. Not surprisingly, we found that the students who had detailed lesson plans also had the best presentation in class. While none of the student teachers showed clear evidence of differentiation or accommodations on their lesson plans, some of them actually did both differentiate and accommodate.

Elaine was clearly in charge of her sixth grade Math class. She used techniques that allowed for movement such as drum rolls for impending answers, and thumbs up, stand up, thumbs down for oral review of greater than and less than Mathematical concepts. She used a variety of engaging activities and a lot of questioning. Elaine asked easier questions of her students who were not yet grasping higher-level content and provided students with verbal praise for their correct answers. Although no obvious attempt was made to

challenge the students who were mastering the skill quickly, they appeared to be challenged by the concepts.

Kyle's teaching was inspired, and his students appeared to be enjoying his class. He used imaginative activities and group work. He provided samples to help students with ideas, and he coached them while they worked. Although he relied on the intervention specialist for his students with learning difficulties, he provided choices within his assignments, allowing all of the students to match their interests with the material. When Baker observed, Kyle held class in the school's media center so that each student could work at a computer. The project afforded students with choices of topic and style of the presentation slides they were creating. Kyle set parameters by providing students with titles and a sequence for the slides, but allowed them to choose the format and specific information to include. Most students were consistently engaged throughout the class and freely asked for assistance from Kyle.

Lisa's teaching was exceptional. We each observed Lisa teaching an 8th grade English class. Although her lesson plans did not include differentiation, she clearly asked a variety of levels of questions in order to challenge students who were able to deal with complex issues while also including all of the other students. It was clear that she took her students very seriously by listening carefully and responding to their comments. We were both impressed by the depth of the students' answers. She provided a variety of activities during class and kept the pace at a productive level. She used internet and other sources extensively in order to help students to connect with the material, and she was very aware of engaging as many of the multiple intelligences as possible. She provided study guides, videos, pictures, outside projects, predictions, and questions to help students empathize with the characters. The outside projects allowed students to choose topics of interest to research and present to the class using a format they thought appropriate. We observed one project presentation that was an oral book report and one that was a poster presentation of a research paper on a specific battle from the era represented in the novel.

Diana thought she was helping students to engage with the material in her sixth grade Math class. She used a dance and tessellations, but when Fleming observed her she spent the entire class threatening her students. All of her instructions were given orally and she seemed irritated by her students. She praised students who had done a good job on their tessellations but seemed unwilling to answer other students' questions. She told Fleming that she could hardly wait for her student teaching to be over so she could leave and go back home. When Baker observed her the next day, Diana tried to cancel the visit because she wasn't feeling well. She continued the tessellations activity, but seemed negative and totally disinterested in the students. While these tessellations may give the illusion of differentiation because each student creates their own drawing using patterns of shapes, there was no clear purpose to the activity. We could not ascertain the objective and Diana never told the students why they were doing these activities. It seemed like students were engaged early, but their off-task behaviors increased consistently throughout the class period.

Nancy had all of her students doing the same work, although she provided the accommodation of extra time on the quiz for students who needed it. Fleming observed students finishing a quiz and working on a worksheet. They worked individually at their desks while Nancy circulated to answer questions. She allowed her students to correct any questions that they missed for half credit. It appeared that neither the students with learning difficulties nor the students who had mastered the material were being served. Her students appeared to like her a lot although her instruction and activities were uninspired. When Baker observed another section of seventh grade math the outcome was similar. The lesson relied heavily on lecturing during which there were three inaccuracies and several missed opportunities to vary explanations and techniques for differing learner styles. After the instructional phase, students worked independently while Nancy sat at her desk grading papers. The students came to her to ask questions about the work.

Interviews

The first question the participants were asked to address was "What is differentiation, as you understand it?" Their responses reflected that they had each learned some portion of what it means to differentiate instruction. For instance, one participant noted the need to "take the time to understand the differences of your students." However this student did not address how to use this information to adjust instruction. Some participants noted only the enrichment end of the spectrum of learners while others noted only accommodating for the special education perspective. When the participants' comments did move in the direction of specific instructional techniques, no one remembered to utilize all three areas —content, process, and product—for differentiation. No one noted the importance of student choices and input in the learning process. All of the students did mention the importance of differentiating if teachers are to maximize student success; they just didn't consistently reflect the extent of how to make that happen.

The second question asked "Did you differentiate as a student teacher? Why or why not?" All five of the participants claimed to be using differentiation techniques. Philosophically, each participant noted the importance of trying to differentiate. However, the examples they used to illustrate their attempts to differentiate were often specific accommodations for special education students such as extended time or shortened assignments. Two participants noted the use of different levels of questions for lower to higher functioning learners, with one also noting differences in grading expectations for different students. One participant noted observing a cooperating teacher building choice into projects the students were doing as an example of differentiation. One student commented on consciously trying to infuse multiple intelligences experiences into lessons due to memories of her own experiences: "I remember when I was in grade school and how boring the lessons (and the teacher)

would be if they didn't shake it up a bit." All of the students reported that differentiating was difficult to do on a daily basis.

The third question asked "Did you feel prepared to differentiate?" Each of the participants reflected the importance of practice in feeling truly prepared to differentiate. Four of the five reported feeling ready, but one felt that she didn't have enough of a chance to practice in her previous field experiences. One noted that differentiation has become more comfortable with more practice noting that "time spent in class helped me learn how to differentiate; but until one actually uses it in an actual environment, one cannot truly know how to do it fully." One participant reflected that some subject areas were easier to differentiate than others. One participant felt more prepared to adapt lessons for learners who were functioning at lower levels than those at higher levels, again illustrating the accommodation mindset rather than the differentiation mindset.

The final question asked of the participants was "What do you think of the junior block lesson plan form?" The participants were split in their response to this question. Two participants noted really liking the Junior Block lesson plan format because it was "well-organized" and "useful in shaping our thinking." One participant waffled noting that the plan was very time intensive and "too complex" to be practical, yet she also noted that what she learned from that experience "always sat in the back of my mind while teaching." Two participants reported disliking the junior lesson plan format because of its length and a sense of repetition in some sections. One of these noted however, that having a section for differentiation on the lesson plan used in student teaching would have been beneficial.

Discussion

Given the data, we found that several issues emerged. These issues can be grouped into three categories: situations inherent in the student teaching experience or setting, instructional practices of the student teacher, and aspects of the teacher preparation experience that could be modified to enhance program outcomes.

Situational Issues

One aspect of the student teaching experience that continues to be problematic is the expectations of the cooperating teachers relative to the expectations of the university instructional team. The influence of modeling can, at times, be paradoxical to research-based, quality practices. For instance, some student teachers are told that lesson plans are something you won't have to write anymore once you get your own class. They then learn to minimize the importance of planning lessons that meet a clear purpose, but rather have nothing more than a series of groundless activities. Additionally, it is always difficult to find classroom settings where every student's needs are being met to the maximum extent possible. What seems to be occurring is that we infuse a foundational philosophical perspective in the students that is then minimized in the field when they see so little implementation of differentiation, especially for students who master concepts quickly and need higher degrees of challenge. Finally, our encounter with a student who showed no heart for the experience was discouraging. While we don't expect all of the student teachers to be excellent, we do expect them to have a positive attitude about teaching and learning.

Practice-based Issues

We observed minimal evidence that awareness of student needs and actual assessment data were the driving force of instruction. The gifted end of the spectrum was frequently ignored and the special education students were still doing the same work done by all the other students with some minimal accommodations. Series of activities with no clear sense of purpose surfaced in some situations with a heavy reliance on whole class instruction. The use of group work was rarely seen and consisted of small groups all doing the same tasks. When evidence of differentiation could be found, it was primarily in the product area; some evidence could be found in the process area, but no content differentiation was observed. Most of the actual differentiation seemed to be incidental as it was not reflected in any of the current lesson plans. At least three of the five students did demonstrate evidence of quality efforts to meet the needs of differing learners.

Preparation Issues

During their junior methods block, the students were each exposed to Tomlinson's (1995) definition that reflects the need to vary content, product, and process as they related to each student's level of interest, learning needs, and readiness level. Walker (2001) found that while many teachers in the middle grades setting think they employ differentiation, they are more typically offering only simple task accommodations. Walker also noted that many teachers overlook the importance of involving the student in learning decisions. Our students' responses mirrored these perspectives, but also showed that they each had latched onto some specific aspect of the differentiation concept without internalizing the whole picture. Clearly we need to infuse additional opportunities for students to experience differentiation in our methods setting. Given the findings of Baylor, Kitsantas, and Chung (2001) and Taylor (2000), we need to provide more modeling to enhance student experience with differentiated lesson planning. The lesson plan sample these students used as juniors (see Appendix A) needs to be revised to address the issues raised in this study such as length, repetition, and practicality. Additionally, the department lesson plan used for student teaching needs to include a section for differentiation. At least having the categories will encourage our students to think about differentiating. Finally, a continued emphasis on reflective teaching is necessary as an individual's need for lesson preparation varies based upon self-awareness of strengths and weaknesses. It was of great concern to us that a lack of actual content knowledge in some areas had an impact on instruction. Trying to achieve a differentiated learning environment when the student teacher is struggling with the accuracy of content is not a realistic venture.

Recommendations

As a result of the information we gathered from this study, we think we are headed in the right direction by requiring differentiation in the Junior Block. We were pleased to see that three of the five students were varying their instruction and assessment to encompass multiple intelligences. These three teacher candidates were also making an attempt to teach material on multiple levels, challenging the students at higher levels while supporting students at lower levels. However, we will also need to revise several aspects of our junior methods block. First and foremost, we have modified the lesson plan design in an attempt to streamline the format for easier use. Continued revision of the lesson plan form will be made until students report greater success and a stronger willingness to use it. Then we will propose that our Department of Education at Ashland University adopt the same format so that consistency can be attained. It may even be helpful to develop a streamlined lesson plan design to be introduced once students have shown mastery of the more detailed format. Providing a computerized version of both may also serve to enhance the usefulness of the form. Instructionally, we will provide students with examples of completed lesson plans designed to better illustrate how they can use the plan to support differentiation. Additionally, we believe that we need to spend more time teaching and coaching differentiation, including learning needs, multiple intelligences, learning styles, and student choice.

Multiple researchers (Giebelhaus, 1999; Giebelhaus & Bowman, 2002; Salzman, 1999) have suggested that mentoring is a critical component for the success of beginning teachers. Therefore, it would behoove us to examine their mentorship programs so we can train our cooperating teachers according to their models. Geibelhaus and Bowman recommend using the Pathwise and Praxis III formats for the selection and training of mentors. Creating partnerships with cooperating schools has also been shown to be an effective method to mentor teachers (Dever, Hager, & Klein, 2003; Salzman, 1999). Our university tends to take volunteers as cooperating teachers rather than using a set of criteria from which to select them. In the future, we hope to work more closely with our school partners to involve cooperating teachers who are willing to practice methods of differentiated planning, instruction, and assessment.

References

Baylor, A., Kitsantas, A., & Chung, H. (2001, March-April). The instructional planning self-reflective tool: A method for promoting effective lesson planning. *Educational Technology, 41*(2), 56-59.

Brimijoin, K., Marquissee, E., & Tomlinson, C. A. (2003). Using data to differentiate instruction. *Educational Leadership, 60*(5), 70-73.

Dever, M. T., Hager, K. D., & Klein, K. (2003). Building the university/public school partnership: A workshop for mentor teachers. *The Teacher Educator*, *38*(4), 245-255.

Franklin, J. (2004, November). Reaching potential: How GT educators are making a difference. *Education Update, 46*(7), 3-8.

Giebelhaus, C. (1999). Leading the way: State initiatives and mentoring. *Mid-Western Educational Researcher, 12*(4), 10-13.

Giebelhaus, C., & Bowman, C. L. (2002). Teaching mentors: Is it worth the effort? *The Journal of Educational Research, 95*(4), 246-254.

Jackson, A.W., & Davis, G. A. (2000). *Turning points 2000: Educating adolescents in the 21st century*. New York: Teachers College Press.

Johnson, R. B. (1999). Examining the validity structure of qualitative research. In A. K. Milinki (Ed.), *Cases in qualitative research: Research reports for discussion and evaluation* (pp. 160-165). Los Angeles: Pyrczak Publishing.

Krefting, L. (1999). Rigor in qualitative research: The assessment of trustworthiness. In A. K. Milinki (Ed.), *Cases in qualitative research: Research reports for discussion and evaluation* (pp. 173-181). Los Angeles: Pyrczak Publishing.

Levine, M. (2003). Celebrating diverse minds. *Educational Leadership, 61*(2), 12-18.

Salzman, J. A. (1999). With a little help from my friends: A course designed for mentoring induction-year teachers. *Mid-Western Educational Researcher, 12*(4), 10-13.

Taylor, P. M. (2000, May). When are we ever going to use this? Lessons from a mathematics methods course. *School Science and Mathematics, 100*(5), 252-255.

Thomas, C. C., Correa, V. I., & Morsink, C. V. (2001). *Interactive teaming: Enhancing programs for students with special needs* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Tomlinson, C. A. (1995). *How to differentiate instruction in mixed-ability classrooms*. Alexandria, VA: ASCD.

Tomlinson, C. A. (2003). Teach them all. *Educational Leadership, 61*(2), 7-9.

Walker, K. (2001). *The utilization of multiple intelligences to differentiate instruction in a middle school setting*. Unpublished doctoral dissertation from Bowling Green State University.

Walther-Thomas, C., Korinek, L., McLaughlin, V. I., & Williams, B. T. (2000). *Collaboration for inclusive education: Developing successful programs*. Boston: Allyn & Bacon.

Wiersma, W. (2000). *Research methods in education*. Needham Heights, MA: Pearson.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: ASCD.

Wormeli, R. (2001). *Meet me in the middle: Becoming an accomplished middle-level teacher*. Portland, ME: Stenhouse Publishers.

Appendix A

Lesson Plan Form

Name_____Class_____Date_____Topic_____

I. Essential Question(s)

1.

2.

II. Unit Question(s)

1.

2.

III. Differentiated Targets (Knowledge, Skills, Understanding)

Fundamental: The student will. . .

1.

2.

3.

Core: The student will. . .

1.

2.

3.

Enriched: The student will. . .

1.

2.

3.

IV. Differentiated Assessment (How students will demonstrate learning)

Explain assessments, and attach assessments/rubrics you will use.

Fundamental:

Core:

Enriched:

V. Differentiated Instruction (What and How You Will Teach: How you will assure that the students will achieve the targets. Activities you will have students do)

A. Introduction (What you will **say** to get students' attention: **WHERE**)

(*Appendix A continued on next page*.)

(*Appendix continued from previous page*.)

B. What teacher will do

Fundamental:

Core:

Enriched:

C. What students will do

Fundamental:

Core:

Enriched:

VI. Infusion of Multiple Intelligences (Linguistic, Logical-Mathematical, Spatial, Bodily-Kinesthetic, Musical, Interpersonal, Intrapersonal, Naturalist)

Name and explain which multiple intelligences you will use in the lesson.

VII. Learning Climate

Explain how you will establish a learning climate for best achievement.

VIII. Grouping

Explain how you will group students for best achievement.

IX. Other Accommodations

Explain how you will accommodate for other student needs.

X. Connections

Explain how this lesson connects with the essential and unit questions and previous and future lessons.

Attach a reflection of your teaching and additional sheets as needed.

(Fleming, Baker, & Rushton, 2001)

Pre-service Teachers' Perceptions of Character Education

Floyd D. Beachum
University of Wisconsin–Milwaukee

Abstract

The purpose of this study was to assess pre-service teachers' support for character education and analyze their perceptions of character education as an effective deterrent to negative school behaviors. In addition, the author of this study sought to ascertain pre-service teachers' opinions regarding the importance of character education in undergraduate teacher education courses. The instrument utilized in this study was PPCES (Pre-service Teacher Perceptions of Character Education Survey). The study's sample consisted of pre-service teachers enrolled in an undergraduate course at a mid-western university in February 2002. Character education received high levels of support and pre-service teachers felt character education was an effective deterrent to anti-social behavior. Pearson correlation coefficients revealed a statistically significant relationship between pre-service teachers' support for character education and their perception of character education as a deterrent to school discipline problems and school violence. In addition, pre-service teachers supported the notion of including character education issues in undergraduate teacher education courses.

Values are an important part of any society. The argument is that too many of us desire materialism over generosity, self-interest over altruism, and comfort over challenge (Beach, 1992; Boylan, 2000; Lickona, 1991). A possible way to address these problems could be the implementation of character education in elementary and secondary schools. McDonnell (1999) noted, "Character education is one of the most important, if not the most important, answer to our national crisis of character and is absolutely essential to any truly effective reform movement" (p. 251). Character education is viewed by many as a responsibility of schools and teachers (see DeRoche & Williams, 1998; Lickona, 1991; Unell & Wyckoff, 1995).

Due to the lack of research on character education in reference to pre-service teachers, the present study is of particular importance. The perspectives of pre-service teachers are important to the knowledge base of educational research and teaching pedagogy. Weist (1998) noted, "teacher attitudes and beliefs influence teaching behaviors, which affect student learning and behavior" (p. 358). Mahlios and Maxson (1995) state that pre-service teachers have attitudes and beliefs that impact their feelings toward students, themselves, and teaching practices.

The support, importance, and perceived deterrent effect of character education are critical aspects of its continued existence. In recent years character education has received federal, state, and university-level support (DeRoche & Williams, 1998; Jacob & Reetz, 1999; Ryan & Bohlin, 1999). Despite this level of support, many teacher education programs do not reflect the national level of interest in character education. DeRoche and Williams (1998) observed, "Both university-based pre-service teacher education and in-service staff development have all but ignored character education in recent decades" (p. xii). Milson (1999) adds "teacher education programs are not currently training teachers adequately to function as character educators" (p. 44).

Objectives and Purposes

The purpose of this study was to assess pre-service teachers' support for character education and to assess their perception of character education as a deterrent to negative school behaviors. In this study I also sought to ascertain pre-service teachers' opinions as to the importance of character education in undergraduate teacher education courses.

The major questions in this study were:

1. To what extent did pre-service teachers support character education in K-12 curricula?
2. To what extent did pre-service teachers perceive character education as an effective deterrent to school discipline problems and school violence?
3. What is the relationship between group and individual characteristics of pre-service teachers and their support for character education?
4. What is the relationship between group and individual characteristics of pre-service teachers and their perception of character education as an effective deterrent to school discipline problems and school violence?
5. What is the relationship between the support pre-service teachers have for character education and their perception of character education as an effective deterrent to school discipline problems and school violence?
6. Among pre-service teachers, what is the perceived importance of including character education issues in a teacher education course?

[1] This article was accepted for publication by the previous editorial team.

Review of Literature

Character education has gained increasing prominence among education stakeholders in recent years. Since pre-service teachers will be working within this context, it makes sense to understand some of the factors that influence the way character education is discussed and understood. In this section, I report professional and parental factors that have created prominence for character education. Historically, a large number of national professional organizations have rallied around the character education cause. Vessels (1998) found that the Association for Supervision and Curriculum Development (ASCD), the National School Boards Journal (NSBA), the National Association of Secondary School Principals (NASSP), the National Association for the Education of Young Children (NAEYC), the National Education Association (NEA), and the National Society for the Study of Education (NSSE) have all proclaimed the value of character education. Perhaps a part of the reason for this interest is the hope that character education might act to deter negative behaviors. For example, Peterson and Skiba (2001) noted that character education has the ability to address issues of ethical judgment, home and community values, and school violence. Research exploring character education's deterrent effects also has supported these assertions.

The Center for Health and Policy Studies of Columbia, Maryland conducted a three-year assessment of the Community of Caring Program (CCP), which is a comprehensive, integrated initiative that involves discussion, practice, and modeling. CCP encourages a community approach that includes students, all school personnel, parents, and community members. It was concluded that the CCP was "instrumental in promoting academic performance and school behavior among students in a manner that reduces the risk of dropping out of school, a common precursor to other negative behaviors among teenagers…" (Stephen & Stoodley, 1999, p.51).

In addition to examining how community programs work, researchers have also used questionnaires to document administrators, teachers, and parents perceptions of character education. In a study on South Dakota administrators' perceptions on character education, two hundred public school administrators were randomly selected and given an eight-item questionnaire (Wood, 1999). One of the questions focused on perceived teacher and parent support of character education. Wood found seventy-one percent of the administrators perceived parents to be supportive of teaching character education. In another study, parents and teachers at an elementary school in Georgia were surveyed in reference to character education. The purpose of the study was to determine if parents and teachers could agree on beliefs and values to be taught in elementary school (Montgomery & Plevyak, 2000). Out of 132 parents (the survey was sent home through first-graders) and 42 teachers, 112 (85%) of the parents and 37 (88%) of the teachers returned the surveys (Montgomery & Plevyak, 2000, p. 25). Based on these surveys, there was overwhelming support for character education. "Only two parents and one teacher said that values should not be taught in elementary classrooms" (Montgomery & Plevyak, 2000, p. 25).

Method

Participants

Because survey research has been useful for helping educators understand the milieu in which character education is based, I thought I would extend this research by surveying 263 pre-service teachers enrolled in a course entitled Organization and Administration of Education in American Society. The course was offered at a mid-western university during the Spring 2002 semester. Of the 285 pre-service teachers surveyed 263 responded to the survey; for a response rate of 93 percent. These students had been admitted into a teacher education program but had not yet fulfilled requirements for a statewide teaching license.

Of the 263 respondents, 79 (30.0 %) identified themselves as Early Childhood/Elementary Education majors. The other identified teaching majors were 3 (1.1 %) Intervention Specialist/Special Education, 51 (19.4 %) Middle Childhood Education, 27 (10.3 %) Multi-age, 84 (31.9 %) Secondary Education, and 19 (7.2 %) other. The Intervention Specialist/Special Education majors were collapsed into the group entitled other. Of the 260 pre-service teachers responding to the community item, 23 (8.8 %) stated that they were raised in an urban area, 146 (56.2 %) in a suburban area, and 91 (35 %) in a rural area. Three pre-service teachers did not respond to this item.

Instrumentation

The instrument utilized in this study was a compilation of two already existing instruments. The first instrument was developed by East (1996) and was used to investigate the perceptions of South Carolina principals regarding character education. This instrument was validated by the Modified Delphi Technique, an expert workshop supporting the validation of instruments, using a process that solicited content experts for the purpose of gleaning optimal expertise (Webler, Levine, Rakel, & Renn, 1991). The expert panel consisted of six former public high school principals and six people from character education-related fields. Their recommendations and suggestions were integrated into the final version of East's (1996) survey instrument.

The second instrument was used in an investigation of social studies teacher educators' perceptions of character education (Milson, 1999). The author of this instrument noted that frequency distributions from a pilot study were similar to results obtained in a similar study on social studies teachers.

After I obtained permission to use the instruments from both of the original authors I modified their work to create a third instrument, the Pre-service Teacher Perceptions of Character Education Survey (PPCES). Modifications were

necessary because I wanted to assess the perceptions of pre-service teachers rather than Principals or Social Studies teachers. To do this, I omitted or modified items to tailor the PPCES for a pre-service teacher audience. For instance, in the instrument devised by East, there were 12 relevant items and I deemed 5 useful for the PPCES. In the instrument devised by Milson, there were 38 items arranged in three parts: themes of contemporary character education; scope of character education; and character education in a curriculum/ methods course. Of the 38 items, I altered 9 items so that they were relevant to pre-service teachers instead of social studies teachers.

My decision to alter an item was typically made so that the item would be relevant to the population I surveyed. For example, in the instrument devised by East and targeted at school principals, there was an item that stated, "Even if a program of character education happened to be mandated by the South Carolina Department of Education, it would be a low priority item in my high school" (East, 1996). Since the pre-service teachers I surveyed were not in South Carolina and not subject to South Carolina Department of Education mandates, items like these were omitted. In addition to omitting some items, I reworded others. For instance, an item that read, "I believe the direct teaching of positive character traits is a legitimate function of the public high schools" was slightly modified by me to read, "I believe the direct teaching of positive character traits is a legitimate function of K-12 schools". Another example of an item I modified originally read "There is not enough time in a social studies curriculum/methods course to spend time discussing character education." I modified this item to read, "There is not enough time in a curriculum/methods course to spend discussing character education". Content validity was established through the submission of the instrument to a group of experts in the content area. This panel consisted of three professors and a school district superintendent. According to Fraenkel and Wallen (2000) content validation can be established by having someone "look at the content and format of the instrument and judge whether or not it is appropriate" (P.171).

In this study, the dependent variables of support, importance, and perceived deterrent effect of character education were measured through the use of the PPCES. The instrument included 19 items with a 4-point Likert-type scale for items 1 to 14. Items 15 to 19 addressed demographic information, including group and individual characteristics. In order to measure the three dependent variables participants had to make choices from the following options: 1 = strongly disagree; 2 = disagree; 3 = agree; 4 = strongly agree. The first part of the study measured pre-service teachers' support for character education. This included items 1-5. Items 6 to 10 were designed to determine pre-service teachers' perception of the importance of character education in a curriculum / methods course. Items 11 to 14 measured pre-service teachers' perceptions in reference to character education as a deterrent to school violence and discipline problems. Items 15 to 19 addressed personal information such as gender, teaching major, and the community in which one was raised.

The instrument was administered to a pilot study group and was estimated to take less than 10 minutes to complete. The results of the pilot study assisted in establishing confidence in the survey. Face validity was established through administration of the instrument to the pilot group. The instrument is a compilation of two previously existing instruments in which expert panels were utilized in establishing validity. In addition, frequency distributions from the pilot study were similar to results obtained in other studies (see East, 1996; Milson, 1999).

The reliability of data gathered from the entire survey as well as each of the subscale was calculated using Cronbach's index of internal consistency. The entire scale's reliability coefficient was $\alpha = .8354$. The reliability coefficients for the support subscale was ($\alpha = .5965$), the importance subscale ($\alpha = .7068$), and the perceived deterrent effect subscale ($\alpha = .6786$).

Procedures

The administration procedures first included providing respondents with a verbal definition of character education, distributing the instructions for the instrument, and reading these aloud to the participants. Character education was defined as the deliberate effort to help people understand, care about, and act upon core ethical values such as respect, responsibility, trustworthiness, fairness, diligence, self-control, and caring (Center for the Fourth and Fifth Rs, 2002). Administration materials were then distributed. The administration materials included the instructions and survey. Completed surveys were then scored for the purpose of producing statistical reports and analysis of the results.

Findings

Research Question 1: To what extent did pre-service teachers support character education in K-12 curricula?

A calculation of means and standard deviations was reported for this research question. The average score out of 20 was 17.07 with a standard deviation of 1.66 in reference to the support variable. There were five items on the instrument that specifically addressed pre-service teachers' support for character education (items 1-5). Item 5 received the highest percentage of strong agreement with regard to support. Of the 263 pre-service teachers, 166 (61.9 %) strongly agreed to item 5 which stated, "Both teachers and administrators should be responsible for establishing a positive moral climate in the school".

Research Question 2: To what extent did pre-service teachers perceive character education as an effective deterrent to school discipline problems and school violence?

A calculation of means and standard deviations was reported for this research question. The average survey score out of 16 was 12.40 with a standard deviation of 1.66. There were four items that specifically examined pre-service teachers' perception of character education as a deterrent to school discipline problems and school violence (items 11-14). Item 14 received the highest percentage of agreement with regard to perceived deterrent effect. Out of the 263 responding pre-service teachers, 237 (88.4 %) agreed or strongly agreed on item 14 which stated, "I believe that the direct teaching of positive character traits is an effective means of addressing the problems K-12 schools are experiencing with violence, vandalism, and other discipline matters".

Research Question 3: What is the relationship between group and individual characteristics of pre-service teachers and their support for character education?

To assess the amount of relationship between group and individual characteristics and support for character education, I employed the Point-biserial correlation. The highest correlation was between support for character education and race/ethnicity (r = .17). There was no significance found in any of the relationships (see Table 1).

Table 1
Correlation Coefficients for Personal Characteristics and Support, Perceived Deterrent Effect, and Importance

		Perceived deterrent	
	Support	Effect	Importance
Gender	–.09	–.12	.02
Race/ethnicity	.17	.06	.10
Religion	.11	.06	.06
Teaching Major	–.04	–.06	–.06
Community	–.06	–.07	–.01

Research Question 4: What is the relationship between group and individual characteristics of pre-service teachers and their perception of character education as an effective deterrent to school discipline problems and school violence?

In order to measure the amount of relationship between pre-service teachers' group and individual characteristics and perception of character education as a deterrent to school violence and school discipline problems, I used a Point-biserial correlation. Data analysis revealed that there was only a slight relationship in the variables. The highest correlation was between perceived deterrent effect and gender (r = –.12). There was no significance observed (see Table 1).

Research Question 5: What is the relationship between the support pre-service teachers have for character education and their perception of character education as an effective deterrent to school discipline problems and school violence?

Data relative to pre-service teachers' support for character education and their perception of character education as an effective deterrent to school violence and school discipline problems were compared using Pearson Correlation Coefficients. A Pearson Correlation Coefficient of .54 indicated that there was a statistically significant moderate relationship (see Table 2). The moderate relationship is also reflected in the r^2 value, .2916, indicating that 29 percent of the variance in support for character education is explained by its relationship with perceived deterrent effect.

Table 2
Correlation Coefficients for Support, Perceived Deterrent Effect, and Importance

		Perceived deterrent	
	Support	Effect	Importance
Support	–	.54*	.61*
Perceived Deterrent Effect		–	.57*
Importance			–

* Correlation is significant at the .05

Research Question 6: Among pre-service teachers, what is the importance of including character education in a teacher education course?

A calculation of means and standard deviations was reported for this research question. The average survey score out of 20 was 15.80 with a standard deviation of 1.87. There were five items that addressed the importance of character education issues in a teacher education course (items 6-10). With regard to importance of character education in a teacher education course, item 9 is of significance because of the high percentage of disagreement. Of the 262 pre-service teachers who responded, 192 (71.6 %) disagreed with item 9 which stated, "I would not include character education as a topic if I had to write a syllabus for a curriculum/methods course".

Discussion

As pre-service teachers move into teaching positions they will have to be even more prepared to teach students of varying races, ethnicities, and cultures. Character education can play a role in the preparation of pre-service teachers and guide them as they balance academic preparedness and morality (see DeRoche & Williams, 1998; Lickona, 1991; Ryan & Bohlin, 1999). The present study revealed that the pre-service teachers surveyed overwhelmingly support character education in K-12 curricula. This result is similar to the results from other studies conducted with social studies teacher educators (Milson, 1999), student teachers (Mathison, 1998), and principals (East, 1996; Wood, 1999). This means that educators representing different stakeholders view character education as a possible solution to many social issues. Since these studies were studies of perception, we can conclude that more research will be required to document how character education might be defined, implemented, and evaluated (Lockwood, 1997).

A major focus of this study sought to determine how important character education issues were to pre-service teachers in a teacher education course. The results suggest that pre-service teachers are committed to the idea of including character education issues in a teacher education course. The results of the present study mirror the results of Milson (1999) who surveyed social studies teacher educators. Historically, character and moral education were transition components of teacher preparation programs in the 1960s – 1970s (Jordan, Metha, & Webb, 2000). That is no longer the case. This study supports Mathison's (1998) claim that more needs to be done in terms of including character education in teacher preparation programs. Other studies have also indicated that character education is not a priority in teacher education even though it generates widespread support (see Jacobs & Reetz, 1999; Jones, Bohlin, & Ryan, 1998). Pre-service teachers' devotion to the importance of character education seems to be circumvented by the actions of those who determine teacher education curricula.

Pre-service teachers' support for character education was compared with their perception of character education as a deterrent to school discipline problems and school violence. With a Pearson Coefficient of .54, there was a moderate relationship between pre-service teachers' support for character education and their perceptions of character education as a deterrent to school discipline problems and school violence.

The results here are consistent with descriptive data described in previous studies (see East, 1996; Wood, 1999). Historically, many educators have supported themes of character/moral education in the hopes that it would deter negative, antisocial, or downright "sinful," behaviors (McClellan, 1999).

Limitations

Several limitations should be noted with regard to this study. Due to the controversial and debatable nature of the subject, respondents may have differing opinions and interpretations of what character education is. In addition, the respondents may have differing degrees of knowledge and experience with character education. Some may be well acquainted with character education literature and concepts, whereas others may be getting introduced to the subject. Furthermore, there may be pressure for participants to answer favorably to character education due to the positive, altruistic nature of the subject.

Ninety-three percent of the targeted pre-service teachers participated, but the sample of this study is one of convenience; thus, limiting the ability to generalize the results. Furthermore, the homogeneous makeup of the population surveyed is also a limitation. Due to the nature of the population surveyed the results are not easily generalized beyond this population. Finally, there are concerns about the calculation and presentation of group means as utilized in this study. Such a limitation is inextricable from the research posited here and must be duly noted.

Implications for Further Research

The following suggestions for further research are based on the results and conclusions of this study. First, additional research studies of perceptions with regard to character education are necessary. The perceptions of paraprofessionals, school counselors, and other populations are of particular importance. Second, additional research should be conducted on pre-service teachers' perceptions of character education. These studies should take place in a range of settings and comparisons should then be conducted. Third, a qualitative study is needed to further explore pre-service teachers' support, definition, and understanding of character education. In addition, there should be more examination of variables such as gender, race, ethnicity, religion, teaching major, and community in which one was raised, in order to find out if there are significant differences within these groups. Finally, additional research utilizing critical theory would better explore the definition and scope of character education. Though people seem to initially agree with the concept of character education, research needs to be undertaken into how and why character education is defined and implemented. A critical theory approach (see Beach, 1992; Lockwood, 1997) would assist researchers in evaluating these differences.

References

Beach, W. (1992). *Ethical education in American public schools*. Washington DC: National Education Association.

Boylan, M. (2000). *Basic ethics*. Upper Saddle River, NJ: Prentice Hall.

Center for the Fourth and Fifth Rs (2002). What is character education? Retrieved from www.cortland.edu/c4n5rs

DeRoche, E. F., & Williams, M. M. (1998). *Educating hearts and minds: A comprehensive character education framework*. Thousand Oaks, California: Corwin Press.

East, R. K. (1996). South Carolina public school principals' perceptions of character education programs. *Dissertation Abstracts International, 57*(11), 4608. (UMI No. 9711678)

Gutek, G. L. (1991). *An historical introduction to American education* (2nd ed.). Prospect Heights, IL: Waveland.

Hatch, A. J. (2002). Accountability shovedown: Resisting the standards movement in early childhood education. *Phi Delta Kappan, 83*(6), 457-462.

Jacobs, G. M., & Reetz, L. J. (1999). Faculty focus on moral and character education. *Education, 120*(2), 208-212.

Jones, E. N., Bohlin, K. E., & Ryan, K. (1998). Character education & teacher education: How are prospective teachers being prepared to foster good character in students? *Action in Teacher Education, 20*(4), 11-28.

Jordan, K. F., Metha, A., & Webb, D. L. (2000). *Foundations of American education* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Lashway, L. (2000). Who's in charge? The accountability challenge. *Principal Leadership (Middle School Education), 1*(3), 8-13.

Lickona, T. (1991). *Educating for character: How our schools can teach respect and responsibility.* New York: Bantam Books.

Lockwood, A. T. (1997). *Character education: Controversy and consensus.* Thousand Oaks, CA: Corwin Press.

Mahlios, M., & Maxson, M. (1995). Capturing pre-service teachers' beliefs about schooling, life, and childhood. *Journal of Teacher Education, 46*(3), 192-199.

Mathison, C. (1998). How teachers feel about character education: A descriptive study. *Action in Teacher Education, 20*(4), 29-38.

McClellan, E. (1999). *Moral education in America: Schools and the shaping of character from colonial times to the present.* New York: Teachers College Press.

McDonnell, S. N. (1999). Living up to a code of ethics: There is always room for improvement. *Vital Speeches of the Day, 65*(8), 249-251.

Milson, A. J. (1999). The perceptions of social studies teacher educators regarding character education (Doctoral dissertation, University of Georgia, 1999). *Dissertation Abstracts International, 60*(05), 1517. (UMI No. 9928968)

Montgomery, C., & Plevyak, L. (2000). What we value: A survey of one community's beliefs. *Social Studies and the Young Learner, 12*(4), 24-26.

Peterson, R. L., & Skiba, R. (2001). Creating school climates that prevent school violence. *The Clearing House, 74*(3), 155-163.

Ryan, K., & Bohlin, K. E. (1999). *Building character in schools: Practical ways to bring moral instruction to life.* San Francisco: Jossey-Bass.

Schaeffer, E. F. (1998a). Character education makes a difference. *Principal, 78*(2), 30-32.

Schaeffer, E. F. (1998b). Character crisis in and the classroom. *Thrust for Educational Leadership 28*(1), 14-18.

Stephen, C. & Stoodley, J. (1999). Community of caring: A character education program designed to integrate values into a school community. *NASSP Bulletin, 83*(609), 46-51.

Unell, B. C., & Wyckoff, J. L. (1995). *20 Teachable virtues: Practical ways to pass on lessons of virtue and character to your children.* New York: Berkley Publishing.

Vessels, G. G. (1998). *Character and community development: A school planning and teacher training handbook.* Westport, CT: Praeger.

Watson, M., Solomon, D., Battistich, V., Schaps, E., & Solomon, J. (1989). The child development project: Combining traditional and developmental approaches to values education. In L. P. Nucci (Ed.), *Moral development and character education: A dialogue* (pp. 51-92). Berkeley, CA: McCutchan Publishing Corporation.

Weist, L. R. (1998). Using immersion to shake up pre-service teachers' views about cultural differences. *Journal of Teacher Education, 49*(5), 358-365.

Wood, R. W. (1999). Administrators' perceptions of character education. *Education, 120*(2), 213-219.

Index of Authors: 2005

The Index of Authors and the Index of Articles are alphabetical listings of MWER articles from 2005 indicating the issues where they can be found.

(continued on next page)

Index of Authors: 2005 (cont.)

Index of Articles: 2005

The Index of Authors and the Index of Articles are alphabetical listings of MWER articles from 2005 indicating the issues where they can be found.

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