MID-WESTERN

EDUCATIONAL

RESEARCHER

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

White Hall

On The Cover—White Hall

White Hall is one of four buildings housing the new *College of Education, Health, and Human Services* (EHHS) at Kent State University. On July 1, 2005, six academic units merged to become this new College. In addition to producing highly qualified educators, faculty members prepare students for careers and leadership roles in health care, human services, and human development as the College seeks to enhance learning and well-being across the lifespan. The College now includes Speech Pathology and Audiology; Exercise, Leisure and Sport; and Family and Consumer Studies among its professional education programs.

The education of teachers for public schools in a democratic society has been central to the mission of Kent State University since its establishment as the "Northeastern Normal School" in 1910. As its first and largest professional academic unit, the former College and Graduate School of Education was at the center of transforming Kent State from a general purpose university designation in 1935, to its current status as a comprehensive national career center of higher education.

For nearly 100 years, the College has prepared teachers and researchers in professional education for work throughout Ohio and the world. The College has now graduated more than 40,000 teachers and over the next ten years, its graduates will touch the lives of one and a half million young people. The College is accredited by the National Council for the Accreditation of Teacher Education (NCATE) and offers programs through the doctoral level.

Kent State University ranks among the top 99 public universities in the country, according to the Carnegie Foundation for the Advancement of Teaching. This places Kent State among an elite group of the nation's nearly 3,900 colleges and universities. Kent earns national recognition for its breadth of high-quality undergraduate programs balanced with innovative research and graduate studies in selected areas. Kent State University is a Research I Institution. The University's eight campus system also serves the region and the state as a major cultural and recreational resource.

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

MID-WESTERN EDUCATIONAL RESEARCHER

Volume 19, Number 4 · Fall 2006   ISSN 1056-3997

Was Busing the Problem?

Joseph Watras
University of Dayton

Abstract

On 15 April 2002, the Dayton Board of Education, the Ohio State Department of Education, and the NAACP reached an agreement ending busing for racial balance in the city schools. Participants agreed that the era for litigated desegregation was over because busing had failed to raise academic achievement of African American children and court ordered racial balance caused white families to move to the suburbs. While this picture of racial desegregation fueled efforts to enhance the education of Dayton youth, it disguised the cause of the suburban sprawl that was endangering the region.

On April 15, 2002, negotiators for the Dayton Board of Education, the Ohio State Department of Education, and the NAACP reached an agreement to end busing for racial balance in the public schools. The next day, the local newspaper, the *Dayton Daily News*, printed a front page story claiming that the agreement marked the end of the most racially divisive chapter in the city's history. In accompanying stories, the newspaper added that no one spoke in favor of practices implementing racial desegregation. Instead, the newspaper printed pictures of African Americans and white people congratulating each other over the decision. Reporters quoted participants agreeing that the era for litigated desegregation was over because busing had failed to raise academic achievement of African American children and the court ordered racial balance caused white families to flee to the suburbs (Elliott, 2002a, p. A1, 6).

While this picture of racial desegregation portrayed the schools as the cause of the serious problems the urban area faces, surveys by land use experts claimed the state government caused many of those difficulties. In 2003, the Miami Valley Regional Planning Commission received a report from Ameregis that demonstrated that the Dayton region is highly segregated by income and race. On the one hand, according to the Ameregis report, the central cities of Dayton and Springfield faced growing rates of poverty and low average household incomes. These cities lost about 4 percent of their households between 1994 and 2000. Worse, this population loss left low income children behind in the schools. From 1993 to 2000, the percentage of children eligible for free lunch in Dayton elementary schools rose nine percentage points, and more than 57 percent of the children in these schools were non-Asian minority. On the other hand, according to the Ameregis report, the affluent suburbs such as Beavercreek, Sugarcreek, and Washington Township grew nine times faster than the regional average. Although these communities contained only 7 percent of the region's population, they had the highest percentage of school age children almost none of whom were non-Asian minority or eligible for free lunch (Orfield & Luce, 2003).

According to the Ameregis report, these patterns of change hurt every community because the townships and the cities are tied together. For example, in any region, in-comes in wealthy and in poor areas move up or down together. Further, regions with the smallest gaps between city and suburban incomes enjoy the highest income growth and property value appreciation. At the same time, the report claimed the solution was beyond the reach of any community. The uneven development was encouraged by the state finance system that pitted local governments against each other as they offered tax reductions to lure businesses and industries from surrounding communities. In the process, the communities lost tax money for such public services as schools (Orfield & Luce, 2003).

Since the authors of the Ameregis report blamed the process of suburban sprawl on the state's financial system, they recommended that communities cooperate and share their taxes with each other. In addition, they called for changes in tax laws of the state government in ways that would reward the development of abandoned sites in formerly industrialized areas. Above all, they noted that some form of wider regional planning such as exemplified in Smart Growth be enacted.

School officials and local participants ignored the need for wider reforms in state and regional governance when they blamed busing for the rapid and unfortunate changes associated with suburban sprawl. In this way, blaming school busing as the problem makes it difficult to find equitable solutions.

The Racial Desegregation of Dayton, Ohio Public Schools

Located forty-five miles north of Cincinnati, Ohio, Dayton changed dramatically from 1960 to 1980. In that time, the population in the city declined from over 260,000 to about 190,000. At the same time, the percent of African Americans increased from about 20 percent to around 40 percent. In that twenty year period, though, companies such as NCR (National Cash Register), Dayton Tire, and Frigidaire closed their manufacturing operations in the city while taking about half of the jobs to other parts of the country and leaving behind many abandoned factory buildings and associated toxic waste problems. National surveys repeatedly pointed out that Dayton was one of the most segregated cities in the United States. The pattern of segregation was simple. The city is shaped in a

square of about fifty square miles divided in half by a river. Black families lived on the west side of the river and white families lived on the east side. Around the city, the southern suburbs were almost exclusively composed of white families while northern suburban communities were African American (Watras, 1997, pp. 79-80).

Given this pattern of racial segregation, the schools would have been integrated if the buildings had been erected along the river. Unfortunately, the various superintendents and board members who oversaw the creation of the city schools placed the schools within neighborhoods. Since the neighborhoods were segregated, the buildings were segregated. Thus, the city schools were segregated although the Ohio General Assembly had adopted a resolution in 1887 that prohibited any school board from designating some schools for one race and other schools for another race or color (McGinnis, 1962).

In large part, the federal government encouraged residential segregation in the city. For example, until 1950, the Federal Housing Authority required deeds to contain covenants prohibiting African Americans from ownership of houses before the agency would provide financing. These covenants appeared in deeds for homes in Dayton except on the west side of the river (Hemmelgarn, 1972).

In 1951, the school board proclaimed a policy opposing racial segregation. Nonetheless, the board contributed to racial segregation. For example, during the academic year for 1951-52, Dayton had a total of forty seven schools with a total enrollment of about 35,000 students. In thirty-four of the schools, the student populations were 90 percent or more white. In four of the schools, more than 90 percent of the students were African American. Furthermore, officials placed most of the African American teachers hired by the school district in buildings with black students (Brinkman v. Gilligan, 1972).

In 1954, the U.S. Supreme Court declared that racially segregated schools contributed to the academic failures of African American students. As a result, people in Dayton became sensitive to the harm that segregation caused black children. For the first time, an African American minister took office as one of the seven members of the school board member elected at large from the city. In reaction to the slow pace of change there was a riot on the west side of the city on the morning of September 1, 1966. Although the violence quickly stopped when National Guardsmen arrived, several committees formed to study the causes of the disturbance and recommended that segregation had to end in order to give African American students equal opportunities to perform well in schools. In reaction to the committee reports, the school board adopted a policy in 1967 that affirmed the board's desire to achieve more balanced racial composition in the areas served by the schools. Since this seemed to be a call for fair housing rather than school desegregation, it is hard to tell what the board desired to do. Nonetheless, in 1968, the school board appointed a new superintendent. He claimed that he was not appointed to bring about racial desegregation. Instead, he noted that he had heard many comments implying that the community wanted to solve the problems of racial segregation when he was visiting (W. Carle, interview with author, February 11, 1990).

The most effective pressure for change came from the federal government. When the new, liberal superintendent took office in 1968, the U. S. Department of Health, Education, and Welfare served notice in a Title VI compliance review that the Dayton schools segregated students and teachers. Faced with the end of the district's eligibility to receive funds from the 1965 Elementary and Secondary Education Act, the new superintendent entered into a formal agreement to desegregate the faculties within two years. With the support of the Dayton Classroom Teacher's Association, the adoption of a middle school program that shifted programs to different buildings, and the inducement of special arrangements, the superintendent brought about the racial desegregation of the faculty (Dayton Public Schools, 1970).

While the superintendent succeeded in desegregating the teaching staff, he was unable to racially desegregate the students. Part of the problem was that his efforts to desegregate the teachers caused conservatives to create a political party, Serving Our Schools (SOS), complete with ward-like organization and membership dues, dedicated to halting desegregation and to preserving neighborhood schools. In 1969, three members of this conservative party won election to the school board. When they took office, they attacked the policies of the superintendent and campaigned against needed tax levies for the schools. To outside observers, the conservatives were so intractable that only the voters could remedy the antagonism (Split defies cure, 1971).

When the election took place in 1971, the voters elected a majority of SOS candidates. There had been three empty seats, and the SOS offered three candidates while four liberal candidates campaigned for election. The liberals received strong support from the two newspapers then publishing in the city. Although liberal candidates polled more than 60 percent of the votes on Election Day, they divided their votes in ways that allowed the conservatives to win (Goodman, 1971).

On December 8, 1971, before the conservative SOS members took office, the school board agreed to accept blame for segregation, to initiate a plan for school desegregation, and to ask the Ohio Department of Education to create a metropolitan plan of desegregation. The board had not made these proposals impulsively. On April 29, 1971, the Dayton board had formerly asked the Ohio State Department of Education for help in improving the educational opportunities for African American children. When the state department urged the Dayton board to make a plan, the school board appointed an advisory committee known as the Committee of 75. This group reported that some form of desegregation had to extend through out the metropolitan area because the white students would flee if the Dayton schools desegregated the students and the suburban schools remained one race schools (Advisory Committee, 1971).

On January 3, 1972, the new board with the majority of SOS members met and reversed the decisions to desegregate the schools. As a result, in April 1972, with the help of the superintendent and the remaining liberal school board members, the NAACP gathered evidence and filed suit in U. S. District Court calling on the state to create a metropolitan plan that would racially desegregate the students from the suburbs into the city (Dimond, 1985).

Although the Committee of 75 had expressed an urgent need for a regional program of racial desegregation, the decision of the U.S. Supreme Court in 1974 about Detroit, Michigan ended any hope of bringing about metropolitan desegregation in Dayton. In the 1974 decision, *Milliken v. Bradley*, the justices decided that courts could not order the merging of school districts for racial balance unless there was evidence that officials created the different districts to racially segregate the students.

In 1973, the judge of the U. S. District Court decided the case in favor of the conservative board that resisted desegregation. The NAACP appealed and won. As a result, the case went back to the U.S. District Court where the judge strengthened his original objections and accepted a plan from the conservative board that would create some new programs but would not have addressed the problem of racial segregation. This time the NAACP carried the appeal to the U.S. Supreme Court whose justices allowed the reversal of the U. S. Court of Appeals to stand. Thus, as a result of the order in Brinkman v. Gilligan in 1975, the U. S. District Court judge had to call for the racial integration of Dayton schools.

Busing for Racial Balance in Dayton

On September 2, 1976, the school doors opened and buses rolled. "It was wonderful," an African American teacher told newspaper reporters about facing the first racially integrated class she had seen in her thirty-one years in the Dayton system. There was no violence or any threat of it. As attendance rates reached 80 percent of enrollment, Dayton became a national symbol. The Associated Press carried stories about the chief of the U.S. Justice Department's Civil Rights Division claiming that Dayton had one of the most successful large city school desegregation plans in the country (Pottinger, 1976).

Despite the official praise, the public reception of the success of busing was mixed. On the one hand, city leaders complimented each other on the success of the program. On the other hand, the SOS board members retained their negative attitudes. For example, the SOS members appointed the former principle of a local high school to be superintendent when they refused to renew the contract of the more liberal superintendent who came in 1968. In the midst of the celebrations over the successful implementation of racial desegregation, the conservative superintendent warned that the situation could explode. He claimed that more than 70 percent of the citizens opposed busing (Maxwell, 1977).

The conservative superintendent did not explain how he knew that 70 percent of Dayton's citizens opposed busing. If this assertion was true, one measure showed that students accepted busing at a much higher rate. At the end of the school year for 1976-1977, the district gave high students who had to attend different buildings the opportunity to return to their home schools. Although 85 percent of the white students who had been sent to a historically black high school in the city asked to return to their white schools, less than half of the white students sent to another black high school asked to leave. Further, less than 30 percent of the African American students asked to return to their formerly one race schools (48% choose, 1977).

During the first days of the busing program, newspapers reported several beneficial changes. For example, stories appeared about elementary students making new friends. Other stories described human relations efforts such as school buildings forming buddy systems to acclimate the new students to the changes. One surprising change came in school decorations. Principals in formerly one race buildings realized that the pictures on the walls depicted heroes of the same race as the students who had been in the building. As a result, principals traded pictures with each other so that the walls in the buildings displayed representative samples of distinguished people from both races (Truth was on the walls, 1976).

Most important, white flight was not immediately apparent. The enrollment in one private, fundamentalist church school doubled to 425 students. The pastor who founded this school had complained about racial integration on what he called biblical grounds. This increase was short lived. By 1986, the enrollment was down to 220, and the school had accepted some African American students. Other private schools, such as Dayton Christian and the Dayton area Catholic schools, did not have a large influx of new students. The superintendent of Dayton's Catholic schools had asked his principals not to accept new students fleeing desegregation (Adams, 1976).

Despite these optimistic signs, white flight could have been disguised by a general exodus of population from the area that resulted from the loss of almost half of the manufacturing, sales, and service jobs available in the metropolitan region between 1970 and 1980. Thus, although the Dayton board of realtors did not report a rapid increase in homes for sale as a result of busing, the enrollments in area schools declined from about 140,000 in 1968-1969 to about 100,000 in 1980-1981. Perhaps most drastic, the enrollments in the Dayton city schools declined from 1969, when about 62 percent of the 60,000 students were white, to 1980 when about 40 percent of the 32,000 students were white (U.S. Bureau of the Census, 1988).

The decline in enrollments and in percentage of white students in the Dayton city schools took place at the relatively steady rate of about one percent per year over several years. They did not come at one burst with the advent of racial desegregation of the schools (Rev. D. Holm, letter to Sr. Christina Bruno, April 20, 1980).

Ignoring Wider Changes in the Region

Since the decline in white enrollments took place slowly, the changes may have been the result of a larger process. This is what the authors of the Ameregis report claimed when they blamed the process of suburban sprawl on a state financial system that rewarded communities for offering incentives to draw commercial interests away from the cities.

Despite the evidence that the problems facing Dayton schools derive from the economic situation in the region, newspapers quoted parents who said that busing had hurt the schools because desegregation dismantled the black neighborhood with a black school and a black staff. According to these spokespeople, such black institutions enabled African American children to succeed. Testing the popularity of segregation in an opinion poll in 2001, newspapers polled black parents of whom about 53 percent wanted an end of busing while 34 percent wanted it to continue. In the same poll, questioners found that about 84 percent of the white parents opposed desegregation of schools.

Members of the Dayton school board, the Ohio Department of Education, and the NAACP repeated the view that court supervision of racial desegregation had caused problems in the schools. In September 2002, a local university hosted a symposium to illuminate the factors that led to the end of the last remaining racial desegregation plan in Ohio. According to newspaper reports, the judge who oversaw the process warned the audience not to look on busing as a failure. Although he acknowledged that it may have caused middle class black and white people to leave the schools, he felt that it had taught the community that different children cannot receive different opportunities. Despite the judge's warning, the president of the school board said that busing had caused academic failure because, among other things, the school officials paid more attention to bus routes than to academic plans (Elliott, 2002b).

As might be expected, the school board sent several brochures to the families in Dayton proclaiming that a new day was dawning when the U.S. District Court released the Dayton schools from supervision. The ads claimed the court sent a message that schools should return to their rightful place at the heart of the neighborhoods. The brochures promised that a one time payout of $32 million from the State of Ohio would help in a school rebuilding program. Further, the ads thanked the voters for approving Issue Four, a school construction levy in an election held on November 5, 2002. According to the ads, this vote qualified the community for state monies to begin rebuilding schools in neighborhoods to which children could walk. The ads promised these initiatives would bring about a revival of proud neighborhoods.

Care is needed in criticizing any public relations campaign such as the one to which these brochures belonged. After all, public relations is designed to build good relations not create discerning citizens. Nonetheless, by repeating the slogan that the problem was busing, the school board over-looked the fact that they were following academic measures previous administration had employed and drew attention away from the economic factors the Ameregis report described.

In 2002, when the school board president asserted that school officials spent more time with bus routes than with academic remediation, she offered no evidence for the claim. More important, two factors suggest that she was wrong. First, in 1975, when the school master, John Finger, planned the busing program, he claimed that the process was easy. He paired schools from one side of town with similar schools on the other side of town and moved entire classrooms together. While the judge gave the school board latitude in following Finger's ideas, the small size of Dayton and the ease with which children could go from one place to another simplified the mechanics of the plan. Second, at the same time that officials planned busing routes, they adopted a series of innovative curriculum reforms to advance student achievement. These included magnet schools, individually guided education, multicultural education, and extensive programs of remediation. Thus, busing was an incentive to curriculum innovation (Watras, 1997, pp. 190-192, 261-290).

The claim that busing caused white flight implies that the end of busing would reverse the exodus of middle class people from the city and the school. In Dayton, the end of busing did not bring about improved enrollments. According to figures on the state department's web page (*www.ode.state.oh.us*), the enrollment in Dayton public schools dropped from 22,500 students in 1999-2000 to 20,586 in 2000-2003. While there may be several reasons for the continued decline in public school enrollment, it is clear that busing is not the only or most important cause of the decline.

Occasionally, people blamed the cost of busing for weakening academic programs. This was a misconception because the state department of education paid the transportation expenses as part of the original decision. The rationale was that the state had allowed Dayton to cause segregation and was therefore culpable. Nonetheless, the ads overstated that benefits that the infusion of $32 million on a one time buyout from the state would bring to academic improvement. On January 15, 2003, the headline of the *Dayton Daily News* proclaimed that Dayton schools were the worst in Ohio. None of the other 607 school districts in Ohio had performed as badly on the state's report card. The state department of education had established 22 standards based on proficiency tests, graduation rate, and student attendance. The officials evaluated different schools and districts according to how many standards they met. Dayton school district had met three. Responding to the bad news, Dayton school officials claimed the scores came from tests students took before academic reforms were in place and things would improve next year (Elliott, 2003). Despite the optimism of official statements, in August 2004, when the state issued its report cards, Dayton was on the bottom for the third consecutive year. Reporters quoted the superintendent saying that the Dayton

schools had shown progress on three of the standards but the improvement was not sufficient to bring about passing grades. According to the superintendent, the problem was that the students were at very low points when his administration began its work (Fisher & Elliott, 2004).

The effort to build neighborhood schools will contradict the aims of the new academic initiatives if segregation by race and social class exacerbates academic failure. In May 2004, the Dayton newspapers published a study of enrollment in Dayton's 36 elementary and secondary schools. Compared to enrollments in 2000, six of those schools had ten percentage points more white enrollment and one school had increased black enrollments by ten percentage points. More important, six of the ten schools that had increased black enrollments also had a number of students who qualified for free lunch that exceeded the district's average. The schools that had increased white enrollments had fewer students who qualified for free lunch than the district average. Thus, the move toward neighborhood schools placed minority children from low income families together into the same schools. Since most studies show that academic achievement follows social class, the policy of neighborhood schools should lead to increased academic failure because it will increase the concentrations of students from low income families in certain schools.

It may be true that schools with more students from higher social classes enjoy more success. Unfortunately, the numbers of middle class children in Dayton decline as prosperous families seek suburban homes.

Conclusion

As noted in the beginning, the Ameregis report found the middle class leaving the cities in the Dayton area because of suburban sprawl fostered by state tax policies. The report did not find school desegregation to be responsible for so-called white flight. To the report writers, the solution had to come through some form of intercommunity cooperation or from changes in state tax policies. Nonetheless, in its campaign for revitalized neighborhood schools, the Dayton school board is claiming that the people in each neighborhood can help improve their own school.

According to the Ameregis report, turning things back on any one community or neighborhood will increase the economic difficulties and the academic problems in the region.

Characterizing racial desegregation as the cause of urban decline disguises this possibility because it focuses attention on the local school board. The rationale is that busing did not work; therefore, the school board has to try something else and everyone should support the board in this effort. Thus, while the criticism of busing reinforces the efforts of the new school board, the complaints allow the causes of the problems associated with suburban sprawl to continue to operate.

References

Adams, S. (1976, July 23). School attendance problems cropping up. *Journal Herald.*

Advisory Committee to the Board of Education. (1971). Report of the Committee of 75. Dayton, Ohio: Dayton Public Schools.

Brinkman v. Gilligan. (1972). Plaintiffs exhibit 2A: One race schools and classrooms in one race schools. Dayton, OH: NAACP Archives.

Dayton Public Schools. (1970). *Task force on staff desegregation.* Dayton: OH: DPS.

Dimond, P. (1985). *Beyond busing: Inside the challenge to urban segregation.* Ann Arbor, MI: University of Michigan Press.

Elliott, S. (2002a, April 16). Desegregation busing ends: Pact avoids court hearing. *Dayton Daily News,* pp. A1, A6.

___. (2002b, September 25). Busing placed city on progress path, Judge Rice says, *Dayton Daily News*, pp. B1, B6.

___. (2003, January 15). Dayton schools worst in Ohio, *Dayton Daily News*, pp. 1A, A4.

Fisher, M. & Elliott, S. (2004, August 25). Valley Districts move up on report cards. *Dayton Daily News*, pp. A1, A5.

48% of students choose to return to home school. (1977, April 20) *Journal Herald.*

Goodman, D. (1971, January, 7). How Bill Goodwin won. *Journal Herald.*

Hemmelgarn, J. H. (1972). Affidavit, *Brinkman v. Gilligan.* Dayton, OH: NAACP Archives.

Maxwell: Most oppose busing here. (1977, April 7). *Dayton Daily News.*

McGinnis, F. (1962). *The education of Negroes in Ohio.* Blanchester, OH: Curless Printing Co.

Orfield, M., & Luce, T. (2003). *Miami valley metropatterns: A regional agenda for community and stability in the Miami valley.* Minneapolis, MN: Ameregis.

Pottinger: Calm here helps others. (1976, September 15). *Dayton Daily News.*

Split defies cure. (1971, July 14). *Dayton Daily News.*

Truth was on the walls. (1976, October 9). Editorial. *Dayton Daily News.*

U. S. Bureau of the Census. (1988). *County and city data book.* Washington, DC: GPO.

Watras, J. (1997). *Politics, race, and schools: Racial integration, 1954-1994.* New York: Garland Publishing.

The Statistically Significant Exact Replication Method: A Programming and Conceptual Extension

David A. Walker
Northern Illinois University

Abstract

The purpose of this article was to extend the research pertaining to the idea of a statistically significant exact replication (SSER) method published recently in the Mid-Western Educational Researcher. *The current study continues the idea of SSER and provides researchers with a software program that will calculate easily the SSER value when there is a statistically significant finding, via an independent samples* t-*test, to assist in determining the chance that an exact replication will be statistically significant beyond 50%. This article also presents an expansion of the SSER concept to the one-way analysis of variance.*

Background

In a recent issue of the *Mid-Western Educational Researcher*, Newman, McNeil, and Fraas (2004), working from earlier theoretical and applied research initiated by Greenwald, Gonzalez, Harris, and Guthrie (1996) and continued by Posavac (2002), presented a method for estimating a study's replicability. Replicability, as defined by Greenwald et al., was "…a test conducted with additional subjects sampled in the same fashion as those in the initial study and tested under conditions identical to those of the initial study" (p. 181). This method of replicability was termed a statistically significant exact replication (SSER), and was premised on the idea that, "… the probability of a statistically significant exact replication (SSER) can be estimated from the probability of the statistical test" (Newman et al., p. 37). Further, it was noted by Greenwald et al. that this concept of replicability was applicable in the prevalent null hypothesis testing (NHT) framework and not in the estimation milieu. That is, in the NHT context, Greenwald et al. found that "… [the] *p* value does provide a continuous measure that has orderly and monotonic mapping onto confidence in the replicability of a null hypothesis rejection" (p. 180).

Additionally, three points should be emphasized when computing SSER values: (a) an "exact replication means that the initial experiment is repeated using the same independent and dependent variables with the same number of participants selected in the same way from the same population" (Posavac, 2002, p. 102); (b) the SSER probability is an estimate derived from a probability value of an observed test statistic that an exact replication will be statistically significant, or as Greenwald et al. (1996) stated its usefulness "… the *p* value computed in NHT is informative … as an indicator of the likelihood that an exact replication would similarly reject the null hypothesis" (p. 182); and (c) the SSER should be thought of as an upper bound value of replicability, which Greenwald et al. described as, "… the effect size for an exact replication of an isolated finding is likely to be smaller that the observed effect size" (p. 180).

For an application of this approach to replication, Newman et al. (2004) presented an example of the SSER method via an independent samples *t*-test, with an observed *t* value of 2.150, 38 degrees of freedom (df), and a two-tailed *t* critical value of 2.024 at the .05 alpha level (i.e., *p* value). Using these data, they calculated the SSER at .55 or the upper bound of replicability (see Newman et al. for the mathematics and order of operation behind this computation). Calculation of the SSER answered the question: how much beyond a 50/50 chance was there of replicating the example's statistically significant finding of p = .038 for an observed *t* value of 2.150 at 38 df? The SSER indicated that for these data, there was just a slight likelihood over a 50% chance of replication, with the upper bound of an exact statistically significant replication estimated at .55, which was not very reliable. Greenwald et al. (1996), citing Cohen's work (1977), suggested that a benchmark of SSER $\geq$ .80 be sought to assure a Fisherian (1951) NHT demonstrable result of the likelihood of a finding being repeated successfully.

Purpose

With this SSER primer in mind, the purpose of this research was not to revisit exhaustively the theory behind the previously-mentioned three articles that form the foundation of the SSER method (cf. Macdonald, 2003 for a critique of Posavac, 2002 and the response from Posavac, 2003, as well as comments from Froman & Shneyderman, 2004). Instead, this article has two equal purposes. The first intention was to provide researchers with a software program in SPSS (Statistical Package for the Social Sciences) that would calculate easily the SSER value when there was a statistically significant finding, via an independent samples *t*-test, to assist in determining the chance that an exact replication would be statistically significant beyond 50%. The second purpose of this article was to present an extension of the SSER concept to the one-way ANOVA (analysis of variance) through an SPPS program. Thus, if a quintessential feature of science is replication, then these two programs and an

extension of the SSER idea should help researchers strengthen their findings by providing them with more data in the area of demonstrable replicability of statistically significant results.

Results

SSER for t-Tests

Table 1 demonstrates the same *t*-test example provided in Newman et al. (2004), but calculated by the SPSS SSER program. From the output, researchers can compare the critical *t* value they desire to examine (e.g., at the .05, .01, or .001 level for a two-tailed test at particular degrees of freedom) and the resultant SSER value with the probability value of the observed *t*.

In Appendix A, the SPSS program syntax is provided. This syntax has been constructed for easy use in the sense that researchers need to supply only two pieces of data: the observed *t* value and the test's degrees of freedom. For example, in the line of the syntax between BEGIN DATA and END DATA, the user inserts two values: the first value is the observed *t*-value and the second value is the degrees of freedom. The entire syntax is highlighted and then run to produce results for the SSER, the critical *t* value, and the probability value of the observed *t*. The program's default is set at the critical *t* value at the .05 alpha level for a two-tailed test. If this default level needs to be changed, the user can follow the instructions embedded within the syntax to do so easily.

SSER for F Tests

In Table 2, the results for two examples using the SSER program for the one-way ANOVA are presented. This program allows the researcher to see their observed F value versus the critical F value they desire at the .05 or .01 alpha

Table 1
SSER t-*Test Example from Newman et al. (2004)*

		Test Statistics			
Degrees of Freedom	Your Observed *t* Value	For Your DF, the Critical Value of *t* Alpha = .05 Two-Tailed Test (Program Default Value)	The Probability of Your Observed *t* Value	For Your DF, the Critical Value of *t* Alpha = .01 Two-Tailed Test	For Your DF, the Critical Value of *t* Alpha = .001 Two-Tailed Test
38	2.150	2.024	.038	2.712	3.566

How Much Beyond a 50/50 Chance Do You Have of Replicating Your Statistically Significant Findings?

The Upper Limit of the SSER Probability Value
.550

Table 2
Example of the SSER Program for the One-Way ANOVA

		Test Statistics			
Degrees of Freedom for the Numerator	Degrees of Freedom for the Denominator	Your Observed F Value	For Your DF1 and DF2, the Critical Value of F Alpha = .05 (Program Default Value)	The Probability of the Observed F Value	For Your DF1 and DF2, the Critical Value of F Alpha = .01
3	99	3.798	2.696	.013	3.986
3	99	2.706	2.696	.049	3.986

How Much Beyond a 50/50 Chance Do You Have of Replicating Your Statistically Significant Findings?

The Upper Limit of the SSER Probability Value
.648
.001

level, the degrees of freedom for the numerator (DF1) and the denominator (DF2), and the SSER value compared to the probability value of the observed F. In the line of the syntax between BEGIN DATA and END DATA, the user inserts three values: the first value is the observed F, the second value is the DF1, and the third value is the DF2. As with the first program, the default value is set at the F critical value at the .05 alpha level. If this default level needs to be altered, the user can follow the instructions embedded within the syntax.

It is interesting to note that in the first ANOVA example, the probability value of the observed F was p = .013 and the SSER value was .648, or a fair chance, yet not at the previous-noted benchmark of .80, beyond 50% that an exact replication would be statistically significant. In the second example, the probability value of the observed F was statistically significant at p = .049, but the SSER value was .001 or virtually no chance, beyond 50%, that an exact replication would be statistically significant, which illustrates Newman et al.'s (2004) caution that a statistically significant result affiliated with an observed test value will not always generate an exact replication that will be statistically significant as well. In fact, Newman et al. advocated for presentation of the probability value of the observed test statistic along with the SSER data because "… researchers need to be careful not to assume that statistically significant findings automatically mean that the chance of obtaining statistically significant exact replications for the study will be high" (p. 37).

Finally, the accuracy of the two programs was checked by an independent source whose hand calculations verified the formulas utilized throughout the programs via various situations employing observed t and F values and their re-spective degrees of freedom. To obtain a copy of the program, please send an e-mail to the author at dawalker@niu.edu.

References

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). New York: Academic Press.

Fisher, R. A. (1951). *The design of experiments* (6th ed.). Edinburgh, Scotland: Oliver & Boyd.

Froman, T., & Shneyderman, A. (2004). Replicability reconsidered: An excessive range of possibilities. *Understanding Statistics, 3*(4), 365-373.

Greenwald, A. G., Gonzalez, R., Harris, R. J., & Guthrie, D. (1996). Effect sizes and *p* values: What should be the reported and what should be replicated? *Psychophysiology, 33*, 175-183.

Macdonald, R. R. (2003). On determining replication probabilities: Comments on Posavac (2002). *Understanding Statistics, 2*(1), 69-70.

Newman, I., McNeil, K., & Fraas, J. (2004). Two methods of estimating a study's replicability. *Mid-Western Educational Researcher, 17*(2), 36-40.

Posavac, E. J. (2002). Using *p* values to estimate the probability of a statistically significant replications. *Understanding Statistics, 1*(2), 101-112.

Posavac, E. J. (2003). Response to Macdonald. *Understanding Statistics, 2*(1), 71-72.

Appendix A and Appendix B appear on the following pages.

Learn more about

The Annual Meeting

The Westin Great Southern Hotel

Columbus, Ohio

October 11–14, 2006

visit http://www.mwera.org

Appendix A
SPSS Syntax for the SSER *t*-Test Program

Copyright David A. Walker, 2005
Contact dawalker@niu.edu
Northern Illinois University, 101J Gabel, DeKalb, IL 60115
   APA 5th Edition Citation

Walker, D. A. (2005). The probability of a statistically significant exact replication (SSER) for the t-test [Computer program]. DeKalb, IL: Author.

```
**NOTE** During your initial research, if the probability value of the observed
t test is > .05, there is NO need to run this program.

* Data enter *.
data list list / tobs (f9.3) DF (f8.0).

*Between BEGIN DATA and END DATA below, put your observed t value (tobs) and also
the degrees of freedom (DF). For the independent sample t-test, the DF = n-2.

BEGIN DATA
2.150    38
END DATA.

**NOTE** Below in TDIFF for the critical value of t, choose the alpha level
for a two-tailed test, either tcrit.05, tcrit.01, or tcrit.001
Currently, the program default is set at the .05 level.

COMPUTE tcrit.05 = ABS(IDF.T(.025,DF)).
COMPUTE tcrit.01 = ABS(IDF.T(.005,DF)).
COMPUTE tcrit.001 = ABS(IDF.T(.0005,DF)).
COMPUTE TDIFF = tobs-tcrit.05.
COMPUTE trep = CDF.T(TDIFF,DF).
COMPUTE sig1 = CDF.T(tobs,DF).
COMPUTE SIG = (1-sig1)*2.
EXECUTE.
* FINAL REPORTS *.
FORMAT tcrit.05 TO SIG  (f9.3).
VARIABLE LABELS tobs 'Your Observed t Value'/SIG 'The Probability of Your Observed t Value'/tcrit.05
'For Your DF, the Critical Value of t, alpha = .05, Two-Tailed Test (Program Default Value)'/tcrit.01
'For Your DF, the Critical Value of t, alpha = .01, Two-Tailed Test'/tcrit.001 'For Your DF, the
Critical Value of t, alpha = .001, Two-Tailed Test'/DF 'Degrees of Freedom'/trep 'The Upper Limit of
the SSER Probability Value'/.
REPORT FORMAT=LIST AUTOMATIC ALIGN (CENTER)
  /VARIABLES= DF tobs tcrit.05 SIG tcrit.01 tcrit.001
  /TITLE "Test Statistics".
REPORT FORMAT=LIST AUTOMATIC ALIGN (LEFT)
MARGINS (*,100)
  /VARIABLES= trep
  /TITLE "How Much Beyond a 50/50 Chance Do You Have of Replicating Your Statistically Significant
Findings?".

**NOTE** A SSER value >= .80 is desired.
```

This research idea was based off of earlier work conducted by:

Greenwald, A. G., Gonzalez, R., Harris, R. J., & Guthrie, D. (1996). Effect sizes and *p* values: What should be the reported and what should be replicated? *Psychophysiology, 33*, 175-183.

Newman, I., McNeil, K., & Fraas, J. (2004). Two methods of estimating a study's replicability. *Mid-Western Educational Researcher, 17*(2), 36-40.

Posavac, E. J. (2002). Using *p* values to estimate the probability of a statistically significant replications. *Understanding Statistics, 1*(2), 101-112.

Appendix B
SPSS Syntax for the SSER One-Way ANOVA Program

Copyright David A. Walker, 2005
Contact dawalker@niu.edu
Northern Illinois University, 101J Gabel, DeKalb, IL 60115
　　APA 5th Edition Citation

Walker, D. A. (2005). The probability of a statistically significant exact replication (SSER) for the one-way ANOVA [Computer program]. DeKalb, IL: Author.

```
**NOTE** During your initial research, if the probability value of the observed
F test is > .05, there is NO need to run this program.

* Data enter *.
data list list / Fobs (f9.3) DF1 DF2 (2f8.0).

*Between BEGIN DATA and END DATA below, put your observed F value (Fobs)
and also the degrees of freedom (DF1 for the numerator and DF2 for the
denominator) from your F test.

BEGIN DATA
3.798   3   99
2.706   3   99
END DATA.

**NOTE** Below in FDIFF for the critical value of F, choose the alpha level
for a two-tailed test, either Fcrit.05 or Fcrit.01. Currently,
the program default is set at the .05 level.

COMPUTE Fcrit.05 = ABS(IDF.F(.95,DF1,DF2)).
COMPUTE Fcrit.01 = ABS(IDF.F(.99,DF1,DF2)).
COMPUTE FDIFF = Fobs-Fcrit.05.
COMPUTE Frep = CDF.F(FDIFF,DF1,DF2).
COMPUTE Fsig = SIG.F(Fobs,DF1,DF2).
EXECUTE.
* FINAL REPORTS *.
FORMAT Fcrit.05 TO Fsig  (f9.3).
VARIABLE LABELS Fobs 'Your Observed F Value'/Fcrit.05 'For Your DF1 and DF2, the Critical
Value of F, alpha = .05 (Program Default Value)'/Fcrit.01 'For Your DF1 and DF2, the
Critical Value of F, alpha = .01 /DF1 'Degrees of Freedom for the Numerator'/DF2 'Degrees
of Freedom for the Denominator'/Fsig 'The Probability of the Observed F Value'/Frep 'The
Upper Limit of the SSER Probability Value'/.
REPORT FORMAT=LIST AUTOMATIC ALIGN (LEFT)
MARGINS (*,110)
  /VARIABLES= DF1 DF2 Fobs Fcrit.05 Fsig Fcrit.01
  /TITLE "Test Statistics".
REPORT FORMAT=LIST AUTOMATIC ALIGN (LEFT)
MARGINS (*,110)
  /VARIABLES= Frep
  /TITLE "How Much Beyond a 50/50 Chance Do You Have of Replicating Your Statistically
Significant Findings?".

**NOTE** A SSER value >= .80 is desired.
```

This research is an extension of similar work conducted by:

Newman, I., McNeil, K., & Fraas, J. (2004). Two methods of estimating a study's replicability. *Mid-Western Educational Researcher, 17*(2), 36-40.

Community, Collaboration, and Creativity: The Potential of Art Education to Create Change

Sandra Spickard Prettyman
Elisa B. Gargarella
The University of Akron

Abstract

The authors used qualitative methods to examine a high school summer arts program in the Midwest, and researched how the program promotes the creative and aesthetic development of students and teachers, as well as an appreciation for cultural and environmental diversity and conservation. We argue the program provides a space, both physical and emotional, where students could come together, recognize the power of community and collaboration in their lives, and engage in powerful creative work. This experience helped them recognize their own responsibility in creating cultural, environmental, and social awareness and change, and moved them toward becoming change agents in their own communities.

Art education and an aesthetic foundation can enhance academic achievement, increase self-esteem, and develop creative and critical thinking skills (Wachoviak & Clements, 2001). In addition, recent explorations in art education focus on the use of art and creative endeavors to develop ethical as well as aesthetic sensibilities (Billings, 1995; Greene, 1995; Hicks, 2002). The use of art education to foster prejudice reduction, cultural awareness, and social transformation has also been developed in recent years (Haynes, 1995; Hicks, 2002; Noel, 2003). Our challenge becomes finding ways to help future and practicing art teachers, along with art students, to develop such relationships between self and a broader social world, between aesthetic appreciation, creation, and transformation. "Understanding better where we are culturally will allow aspiring artists to create powerful and persuasive images of where we might be headed" (Haynes, 1995, p. 50). Such imagining helps us open ourselves to the possibilities that we might envision and create, and without it we are relegated to the status quo (Albers, 1999; Dewey, 1934; Greene, 1995). Thus, art education can foster more than just creative development, but holds open the possibility for transformation of both self and society.

This article presents findings about an art education program that aims to engage in the type of transformation that writers like Greene, Haynes, and others advocate. This unique summer program is run by a Midwestern university and focuses on art education for urban high school students. It seeks to foster collaboration among high school art students, art education teachers and students at the university, community conservationist organizations and artists-in-residence. Through this collaboration, students and artists create a permanent or semi-permanent art installation in a public outdoor space. The program is designed to promote the creative and aesthetic development of students and teachers, and also an appreciation for cultural and environmental diversity and conservation. This research project grew out of our desire to understand the benefits such a program might provide, if any, other than learning specific art skills and techniques. Specifically, we wanted to know:

1. Do students gain artistic appreciation as a result of their participation in the program?

2. How does Arts UP provide a way for students to use their art in critical and transformative ways?

After collecting and reviewing a wealth of data, we argue that this program provided a space, both physical and emotional, where students could come together, recognize the power of community and collaboration in their lives, and engage in powerful creative work. This experience helped them recognize their own responsibility in creating cultural, environmental, and social awareness and change, and helped them move "from the artist's personal experience of aesthetics to development of an ethical point of view" (Billings, 1995, p. 22).

Theoretical Grounding

This paper is guided most broadly by critical theory that focuses on challenging oppressive social formations, including those fostered in educational arenas. Kincheloe (2004) argues that: "A social theory should not determine how we see the world but should help us devise questions and strategies for exploring it" (p. 49). Critical theory framed this research—the questions we asked, how we asked them, and our reasons for asking them. Our reasons for engaging in this study grew out of our own needs to understand how and if students used their art in critical and transformative ways, and if Arts UP contributed to this. In part this is because we see ourselves as engaged in critical pedagogy, a kind of work that is based on several different principles including the need for social and educational justice and equality, the political nature of schooling and education, and the dedication of our efforts toward alleviation of human suffering in everyday life (Kincheloe, 2004). These also guide our own teaching and research.

Using Critical Theory as a Lens and a Framework

Most recently, critical theory in art became prevalent during the decades of great social movements: feminism, civil rights, and gay and lesbian liberation. Artists sought to bring awareness of the ways social, economic, political and historical conditions had an impact on daily lives. As Desai (2002) points out, "Social issues of environment, racism, homophobia, gender, sexuality, homelessness, AIDS, to name just a few, have increasingly provided the battleground for artistic interventions" (p. 309). Like the processes of engaging in art making or visual literacy, critical theory encourages participation and agency. At its heart is the questioning of often taken for granted assumptions about the self and the world that can help us to challenge oppressive social formations. This means exploring possibilities for self understanding and empowerment, for self and social transformation, and for recognition of the historical, social, and political nature of knowledge (including aesthetic knowledge and production). This research sought to understand how a program like Arts UP might help students recognize the social and political nature of their art, and help them engage in the type of self and social transformation articulated in critical theoretical perspectives.

Community Based Art Education

We approached this research with a critical theoretical lens, but also used research and theory about community based art education as a foundation. Taylor and Ballengee-Morris (2004) argue that through such community based education, art students can learn to see the world as a transformative place, and not as a static entity, and that the transformation and civic renewal that take place through this learning "could be viewed as a social movement" (p. 11). We were interested in whether and how Arts UP provided such a transformative space for students, one that could result in self and civic renewal and change.

Researchers and practitioners of community based art education often extol the benefits it offers for students and communities. Ulbricht (2005) argues that through community collaborations, art students can design and create art that can have a positive social impact, thus providing art students with a real life forum for their art and communities with resources they would otherwise not have. Wallot and Joyal (1999) posit, "Students should have the opportunity to demonstrate their sense of belonging by participating in the events that punctuate the social life of the community" (p. 29), and that through such participation students and communities can learn and grow together. In this study, we specifically sought to understand if, how, and why students used their art to create "a positive social impact" and if the program provided a space for the type of community-building and collaboration so often delineated in the literature.

Methodology

Our research examined the complex world of high school students who were brought together from a variety of schools for one purpose, to participate in a university-run summer art program. Our research was exploratory and context specific, criteria which call for a qualitative methodology (Glesne, 1999). The study design utilized qualitative data collection methods of participant observation, document analysis, and interviewing. Data analysis took place both in the field and once data collection was completed (LeCompte & Schensul, 1999; Maxwell, 2005).

Research Population and Researcher Roles

Our study focused on a group of 12 students who were part of a four week summer art program sponsored by a Midwestern university in an urban setting. Students came from five different public high schools in the area, and in the 2004 program there were eight girls and four boys. Some of the students were currently enrolled in art classes, while others were not, but all students had some previous experience with art. Elisa serves as director of the Arts-UP program, and thus access to the population was not a problem. In addition, several of the students had participated in previous Arts-UP programs and knew Elisa from these experiences. While several of the students knew each other, either from going to the same high school or from previously participating together in Arts-UP, others did not know anyone in the program. This meant that we, as researchers, needed to work early on to create rapport and trust both with and in the group. This was done through participating in daily activities with the group including both artistic and social opportunities. We often wrote in our journals as they worked in theirs, ate lunch with them, and attended tours with them. Both of us were accepted very quickly as members of the group, although with different roles. Elisa, as director of the program, was clearly a full participant. Sandra, while she participated actively in many aspects of the program, retained more of an observer status, and students often joked about the notebook that was "permanently attached" to her. Each of our positions provided us with opportunities for interaction and involvement with the students and the program on different levels, giving us a well-rounded portrait of the program and the participants.

Data Collection

Both researchers engaged in daily participant observation during the course of the four week program, taking extensive fieldnotes for each session, which were transcribed and word processed into descriptive, analytic, and autobiographical notes within 24 hours of each observation (Glesne, 2006). Descriptive notes were used to "capture specific interactions among research participants" (p. 58) in order to portray the context and content of the observation, or what was happening at any given point. Analytic notes served as a tool for recording initial ideas about what was going on in the field, identifying problems we experienced, and developing further questions for exploration. Autobiographic notes allowed us to "create a record of [our] behavior and emotions throughout the research process" (p. 60). This

helped us explore our subjectivities and situate ourselves as researchers both in and on the project. Together, we generated over 300 pages of transcribed fieldnotes during the course of the four week project.

We also collected a variety of documents and artifacts, providing us with another source of data. These included student journals, student artwork, artist statements, and documents about the program, such as newspaper articles and fliers about the program. Student journals included responses to specific journal prompts given by Elisa. These included asking students to respond to a quote about diversity, to respond to the idea of metamorphosis based on Kafka's idea in his short story of the same name, and to record their emotions and ideas about the local historic site they visited and worked with for the project. Their journals also included daily reactions to their activities, which could be represented in words or pictures, ideas about the skills they were learning, and interpretations of their experiences in the program. The student art created over the course of the program was quite varied. Much of it represented work done with the professional artist-in-residence for the program, such as the fused glass butterflies created for the community partner, along with personal fused glass pieces the students were able to design and make toward the end of the 2004 program. Students were also given specific art assignments over the course of the program, such as creating a piece of art about an issue that was important to them or creating a piece of art that represented a quote or idea discussed previously in the program. Artist statements were written statements students produced to discuss and present their work, and their thinking behind it. This art, in conjunction with the journals and other documents, provided us with tremendous insight into the lives of the students.

Semi-structured interviews took place with almost all of the students, as well as with the artist-in-residence and with participants from the local historic property that was the 2004 Arts-UP partner. Students were asked questions about how they would portray art and its relationship to the public world: to community, history, ethics, culture, and the environment. They were also asked to consider benefits of participating in the arts, how they prefer to engage in artistic activities, and whether their participation in Arts-UP affected their views on art and its uses and production. The resident artist and community partners also were asked about their views on the relationship between art and public life, but also about their commitment to educating future generations about art and its role in public life. These questions allowed us to explore how those who participated in the program understood the role of art as a transformative force. Through the use of daily observation, document/artifact collection, and interviews, we obtained a wealth of data, providing us with the ability to create rich, thick description about the lives of these students and this program.

Triangulation provides a means to increase the validity, reliability, and trustworthiness of qualitative data (Glesne,

2006). We utilized multiple forms of triangulation in order to strengthen our findings, including:

- the use of multiple methods in our approach to data collection and analysis

- the use of multiple forms of data such as participant observation fieldnotes, student journals and artwork, documents about the program, and semi-structured interviews with students, artists, and community partners

- the use of data that was collected across time, space, and persons

- the use of multiple researchers

Denzin (1970) argues that the use of triangulation helps the researcher in "partially overcoming the deficiencies that flow from one investigator and/or one method" (p. 300), thus serving as one means to increase the validity of the study and the trustworthiness of the data (Denzin, 1970; Glesne, 2006).

Data Analysis

Initial data analysis occurred in the field as the two of us (Sandra and Elisa) wrote fieldnotes and began to ask ourselves questions about what we were seeing, hearing, and experiencing. We also came together frequently over the course of the project to discuss our observations, which allowed us to explore initial interpretive ideas about the data and focus future observations. These meetings did not specifically address the project and business associated with it, but rather provided a venue for us as researchers to reflect on our daily observations and directions for future observations. After completion of the project, both researchers began re-reading fieldnotes, interviews, and documents for emergent themes and patterns. We created code lists from this initial screening of the data, which we refined over the course of multiple readings. Once a working codebook was in place, we coded all data separately and then compared our analyses. Interrater reliability was high, with over 80% overlap in coded sections. Many of the codes we developed emerged from the data, from the words and actions of the participants. However, we also read the data through multiple theoretical lenses, enabling us to explore it in different ways. In addition, other codes portrayed particular conceptual constructs, such as environmental awareness or community development, which informed the study. The use of inductive, theoretical, and conceptual approaches to the generation of codes allowed us to maintain flexibility in the analysis, while it also provided us some degree of efficiency.

The Arts-UP Program:
Building Artistic Skills and Understanding

Arts-UP was initiated three years ago after the art education program at an urban university received a sizeable contribution from a local folk artist to initiate a summer program for high school students talented in visual art. Students submit an application for the program, which includes an application form, a written statement about their rela-

tionship to art and its importance in the world, and teacher recommendations. Students are selected by a committee to work with a professional artist-in-residence for one month in planning, creating, and installing a site-specific art installation for a community partner. They are paid as apprentices over the course of the program, allowing many students who would have to work over the summer to participate. This practice also reinforces their value as artists, and the value of artists to the community. Over the course of the program, students also curate their own gallery exhibition and host a public reception to showcase their work.

One goal of the program is to promote collaboration among high school art students, art education teachers and students at the university, community conservationist organizations and artists-in-residence. While it is designed to enhance the creative and aesthetic skills of students, it also strives to foster an appreciation for cultural and environmental diversity and conservation. Another goal of the program is to help students recognize how they are connected to their local communities and ways they can engage with them through their art. It is hoped these goals provide benefit to both students and the community, and that its effects are able to be sustained beyond the course of the program.

Arts-UP 2004: Creating Community, Developing Collaboration, and Promoting Transformation

Over five dozen students applied to participate in the Arts-UP 2004 summer program. That year, the artist-in-residence was commissioned to work with students to create glass butterflies for the renovated conservatory at a local historic site. In addition, students curated and produced an exhibition of their artwork, both work they created during the previous year and work they created over the course of the project. Of the 60 applicants, only twelve students ranging from ages 14 to 18 were selected to join the program.

The students selected for participation in the 2004 Arts-UP summer program spent the first three days at the university's art school gallery, matting and framing artwork they had created during the previous year. In order to display this artwork, students would curate their own show, deciding where pieces would be displayed and how best to display them. They painted pedestals, measured and leveled prints and paintings, and typed labels, planning and executing virtually every aspect of a gallery exhibition. Experiencing the outcome of curating their first public show, we observed students admiring the gallery walls and the effect of their pieces in a new contextual environment. Many students commented on how much work the show entailed, and how much they learned from the process. This process gave them hands-on experience with the kind of work that many artists and museum specialists do regularly.

Students spent the remainder of their first week at the installation site with docents, curators and educators. Though several of the participants had visited this historic site on school field trips, several had never been to the estate before, even though they lived within ten miles of it. Vincent, a senior participant, expressed this irony when he wrote, "I've been to the Palace of Versailles, yet I have never crossed the threshold of [this place]." Participating in Arts-UP provided these students with opportunities to discover their own community and the treasures it holds, as well as with the recognition of the importance of preserving these treasures.

Throughout their tours of the estate and education about its historical and social significance, participants were challenged to think of ways they were responsible for preserving local culture and history. Kate, a junior in the program, summarized her thoughts on preservation:

> I do feel a personal responsibility to preserve history. I think that is one of the more important things that communities should do…I think that keeping people informed about things can be done just by being informed first, then just talking to people or making art—somehow making knowledge public…But telling people things, showing them stuff—doesn't matter much unless you get them to think for themselves. Getting people to think makes them care about things and then they will want to change or preserve things.

This attitude permeated the students' writing and artwork over the course of the program. Most felt it was important to create awareness, but through that awareness to get people involved in their communities and causes that mattered to them.

To create a positive learning environment, Arts-UP participants spent three weeks in small groups planning, designing and implementing five large scale (4 to 5 foot) fused glass butterflies for the conservatory that was being restored at the historic estate. The subject matter of the project was agreed upon by the site administrators and the Arts-UP planning committee to compliment the theme of butterflies at the gala opening of the restored structure. A resident artist was selected to implement a curriculum that would introduce participants to glass media and techniques, including fusing, etching, sandblasting, and cutting forms. While students learned about historical and contemporary glass, they were also encouraged to consider the impact of public art on helping to educate their communities about the preservation of an often over-looked historic and cultural beacon. Throughout the month-long collaboration, participants also created several other significant personal pieces in addition to their group butterflies. Students each designed personal fused glass pieces of art, but also utilized other forms and media to create art as a response to journal prompts or assignments given during the course of the project. Some students created sculptures, others framed poetry, others designed multimedia pieces. All of the student artwork was displayed at the closing reception, where nearly one hundred visitors came to experience the impressive exhibition of students' and artists' individual and group works.

Results

Finding the Self and Self Expression in the Company of Others

Research in art education suggests "that the impacts of the arts on young people's self and social development dovetail in such a way as to build community" (Smith & Stevenson, 2003, p. 1), which is often defined as a set of positive and supportive relationships. Interview, journal, and artifact data from this project indicate that students involved in the program developed positive relationships with others, including peers, teachers, and artists. Several students claimed they felt greater freedom to be themselves when they were around other artists, and that Arts-UP provided a safe place mentally and physically to express themselves. In her research on community art education and transformation, Lowe (2001) argues that: "The experience of coming together must involve an explicit emphasis on the processes of building community relationships and being creative" (p. 461). ArtsUP actively provided mechanisms through which the process of relationship building could be enhanced, providing students with venues for building community in the group. These included active, daily interaction and involvement, opportunities for collaborative, creative work, and attention to individual as well as group dynamics.

The development of a sense of community was already apparent when on the third day of the program students were asked to write down their thoughts about initial experiences in the program. Cori, a junior student in the program wrote:

> When artists, young artists are brought together an ora [aura], a vibe is created that makes us feel included. Not weird, because we are all 'weird' and it brings us together. Working together has shown me that team work is essential in life.

Annie, a freshman participant, also expressed a sense of belonging in her journal entry. She wrote:

> I think I'm the youngest one at Arts-UP, but it's okay because nobody here treats me like I am. We're all equal. I've never really gotten to hang out with people who are a few years older than me, so this is awesome. I'm learning a lot about meeting people and trying not to be so nervous - I have anxiety disorder. It's also weird because usually when I meet people, I am really uncomfortable, but everyone here seems to have the opposite effect. I feel at home.

Another participant commented that he was glad he did Arts-UP because "you get more culture. It's kind of like an upbringing. You're raised into a certain state of mind with Arts-UP. It's kind of like a family." The term "family" was often used to describe the close bond students felt throughout their participation. When asked to define the concept of family, Hilary wrote in her journal, "A group of people who nurture and support each other in a positive growth; both mentally and physically." Joan, another participant, explained that she "learned how to work in a group that [she was] not familiar with…to collaborate…to agree on one idea." Joan also said she had to keep an open mind about the ideas of others, and try to allow everyone in her group to have a voice. Smith and Stevenson's work (2003) highlights the reciprocal interaction between self and social interaction in arts programs, and "how together they give rise to a powerful sense of community" (p. 1). Students in the project connected with each other, creating a "family" of artists which provided them with a safe physical and emotional space in which to express themselves.

Arts-UP students, engaging in this reciprocal process, experienced transformation of themselves as individuals, and also of the social interactions they engaged in and developed. Participants also learned about the relationship between themselves and the broader social world. Holloway and Krensky (2001) argue that art education is a means to promote personal and social transformation, "a notion rooted in the work of early learning theorists, such as John Dewey, as well as in the history of art itself" (p. 356). Many Arts-UP students recognized the need for self and social transformation and the possibility that art held to engage in such transformation. Joan, a junior participant who makes her own clothes and plays drums for a rock band, spoke of her transformation from being more independent as an artist: "I used to be more independent, but now I'm like independent but also the one who doesn't like to exclude other people." Mason, a sophomore who expresses himself with black clothes, combat boots and gory art, explained that, "Art acts as a catalyst for relationships, for influencing people and for helping others see your side of the spectrum." Gina also recognized the ways in which the program could have a greater impact on the world. She added, "By affecting people in the program, you reach to their families."

Several participants showed depth in understanding the exponential impact they could have on family, friends and a larger community when trying to communicate progressive or alternative ways of thinking. Joan wrote about her beliefs on the subject:

> I've always had this subconscious belief that kids can't do anything (the grownups are the only ones that can make an impact). That idea is starting to change. I think the school system's method of teaching does contribute to a child's feeling of inferiority. It should build their confidence instead of teaching them to compare themselves to others in order to measure accomplishment.

Like Joan, all participants agreed that even their art classes in school reinforced a status quo that was rarely challenged by their teachers. They were pleased that the Arts-UP program pushed their ideas about the world, and their impact on it through artistic expression. Students in the program wanted to know that their ideas and work "mattered,"

and that they could have an impact on their social world (Smith & Stevenson, 2003).

Art, Social Justice, and the Self

Themes of cultural and environmental awareness were prevalent, as was an awareness of the connection between these issues and broader political agendas. Throughout the program, participants were given journal and art assignments as well as quotes to contemplate and interpret into visual images. Hilary created a collage in response to the first journal assignment and provided an overview of her response:

> In the onslaught of everything that happened when ArtsLIFT started, the 12 of us were given an assignment. It was nothing more than a quote, but of course sometimes those are more triggering than anything else. The quote was: "No people are uninteresting, their fate is like the chronicle of the planets. Nothing in them is not particular, and planet is dissimilar from planet." The image that immediately came to me was one of accusation; people pointing fingers at others. I guess you could say that the paper that I glued and scribbled on represented the opinions of others and your opinion and how they can clash worse than polka dots and plaid. However, the interesting part is the fact that the two opinions are not stemming from two different species or planets, but two dissimilar humans. Opinions then eventually lead to accusations or insults but in the end, it all comes back to the fact that regardless what you think or believe, we all have to live with each other anyway.

Repeatedly, the themes of cultural diversity and social justice were contemplated and included in students' thoughts and artwork. One student created a collage centered on the theme of HIV/AIDS. In her work titled Gay Disease, she highlighted the misconception that HIV/AIDS viruses were initially thought to afflict only homosexual populations. She pointed out that although every human could contract the virus, homosexuals remained vulnerable to prejudice and violence. Another student painted an image of Iraqi women and children being killed in the war. She challenged the inhumanity of our nation in several of her writings and works of art. Students were exploring critical issues in their world and hoped to use their art as a vehicle for awareness and change. They believed "the arts help us explore our own and others' thoughts and feelings, critique ourselves and our worlds, express our voices, and influence our social contexts by using nonviolent means" (Holloway & Krensky, 2001, p. 357).

Early in the program the themes of emergence and transformation of the self became prevalent in student work. Perhaps the subject for the project itself, butterflies, led to the frequency of metamorphosis as a recurring theme time and time again. In recognition of this pattern of thinking, we asked participants to interpret a second piece of litera-

ture. This assignment entailed a graphic quote from Franz Kafka's Metamorphosis (Kafka, 1915, p. 3). The quote (found below) opened possibilities for a broader media base and as one student put it, "a more individual interpretation."

> When Gregor Samsa woke up one morning from unsettling dreams, he found himself changed in his bed into a monstrous vermin. He was lying on his back as hard as armor plate, and when he lifted his head a little, he saw his vaulted brown belly, sectioned by arch-shaped ribs, whose dome cover, about to fall off completely, could barely cling. His many legs, pitifully thin compared with the size or the rest of him were waving helplessly before his eyes.

In response to the quote, several students created artwork that reflected their dreams, or perhaps nightmares about metamorphosis and change. They used this assignment as an opportunity "to name themselves, envision alternative realities, and engage in remaking their worlds (Holloway & Krensky, 2001, p. 358).

For example, Cori developed a sculptural piece comprised of bloody plaster hands arranged on a white sheet. Its stark and haunting presence on a well-lit pedestal drew many onlookers at the reception. Cori explained its significance in an artist statement that made several observers cry as their eyes moved back and forth between the artwork and the artist's words.

> So I took something that I would not want to ever be and re-wrote the excerpt my own way. I would not like to be my father, abusive and drunk. So here is my story…When (Cori) woke up one morning from unsettling dreams, she found herself changed in her bed into a monstrous vermin. She was lying on her side which was all bruised, when she opened her eyes a little she saw her girlfriend bleeding, lying unconscious-broken nose, whose breath became so shallow, she could barely hang on. Her bloody hands staining the white sheets as they laid there like her girlfriend, before her eyes.

Here, Cori was making a statement about cultural acceptance of homosexuality and her fear of not being accepted by her own family. She also used this piece as a visual and written message to her grandmother to tell her that she was a lesbian. Her artwork served as a powerful vehicle for communication, personal and political, in her life. Many students, like Cori, chose to expose personal issues and intimate truths, and said they found strength within the healing powers of artistic expression.

The theme of self transformation was again evident in a student's abstract painting dealing with the issue of body image. Joan's artist statement about her piece entitled *Yesterday but not Today* highlighted her need to examine herself in relation to broader social and cultural images. Her artist statement appears in Figure 1.

Joan's statement reflects her need to make the world aware of an issue that is important to her, and reflects her desire to create both self and social change. Like Joan, almost all of the participants began to use their art as a forum for creating awareness about social issues that were important to them. Albers (1999) argues that by examining student artwork, "educators are more able to identify how students see their world and how their visual constructions of meaning reveal their own beliefs about social locations such as gender, race, class, and sexual orientation" (p. 8). It is through this awareness that art educators can then encourage greater socio-political awareness and action on the part of their students. We argue that programs like Arts-UP are imperative in creating not only future artists, but also future activists who will be critical thinkers and active, democratic citizens, capable of effecting change.

Conclusions about Community, Collaboration, and Creativity

Arts experiences, through structured collaboration, learning, and creation, allow young people to explore and absorb multiple worlds into a cohesive, active, participatory environment where they become agents of their own change. (Smith & Stevenson, 2003, p. 12)

The Arts-UP program sought to marry the ideals of arts inquiry and production in heightening a sense of social consciousness in teenagers. The program allowed for students to connect learning with the real world. It provided opportunities for students to make choices and decisions about what, why and how they learn. Its structure, built around experiences that allowed students to feel a deeper connectedness to their social and cultural communities, helped to change beliefs about how they viewed their own learning potential. Perhaps most importantly, initial results suggest that students will remain engaged and committed to projects and to each other long after the end of the program.

In addition, some students began to use their art as a forum for creating awareness about social issues that were important to them. Many students created pieces of art over the course of the program that reflected what they saw as oppressive and unjust situations in the world, and many said their involvement in the program made them think about the impact of their art in a new way. In their artist narratives and in interviews they often articulated political positions and spoke about the importance of using their voices to create change, for themselves, their local community, and society in general. For some, this work was about the alleviation of suffering, in the self and in others, work that clearly fits into a critical theoretical perspective. Initially, some students believed their voices were not heard or taken seriously, questioning whether it was worthwhile to present their point of view as an artist. However, later in the program, many of these same students created pieces of art with explicit social and political themes, and in artist narratives or interviews spoke about how they had come to see the importance of using their art to represent their point of view and to raise

---

Yesterday but not Today is a piece that I created one day entirely from impulse. It is a reaction to and release of feelings that had built up inside of me in regards to the appearance of my own body.

There are no words that I can find to even begin to express the amount of pressure that is put on females to be thin. It all starts when a young girl [gets] the idea that beauty lies solely on the exterior (because that is all that a photograph or doll portrays). That idea of more beauty being the result of fewer pounds soon grows into a subconscious belief as its image is repeated millions of times throughout her life in advertisements, movies, magazines, cartoons, melodramas, and any other visual genre that is trying to attack viewers and make a profit.

Soon the day arrives when a girl becomes conscious of the fact that she is different from these women she sees in the media. To the girl, this difference may be her own height, feet, or the measurements of her figure. The most widely shared "flaw" in young girls (and in all women self-conscious about their body) is undoubtedly their size.

Next, having identified her deformity, the girl's body image will become even more negative and she will feel even more "fat" than she actually appears. She will develop a feeling of inferiority when comparing herself to the slender "pretty" girls that she sees around her. She will feel hopelessly alone and love will seem unattainable, believing that the opposite sex is attracted only to those "beautiful" girls. But one day the girl will become tired of feeling inferior and of having to hide her imperfections and will come to realize that her belief was not her choice, and that she does not agree with the decisions made by the profit-seeking business men.

HER ROUND STOMACH(E) IS BEAUTIFUL BECAUSE IT NEVER CHANGED, NO MATTER HOW HARD SHE TRIED TO FLATTEN ITS ELEGANT CURVES. HER UNCONTROLLABLE CURLS ARE A MIRACLE, STYLING THEMSELVES IN A WAY THAT BEST COMPLIMENTS HER FEATURES. HER FACE WOULD BE ALIEN WITH ANYTHING OTHER THAN THAT BEAST SHE CALLS "NOSE" ATTACHED TO IT. EVERY "FLAW" IS PERFECTION.

YESTERDAY I felt FAT and UNDESIRABLE, but TOMORROW I will NOT feel that way, nor will I EVER feel that way AGAIN.

Figure 1. Joan's Artist Statement

awareness. While many of the students spoke about the impact of the program on their artistic abilities, many also spoke about how it changed the way they related to the world and specifically to others in it.

Freedman (2004) writes that: "Considering how much art education takes place in the world, the field is remarkably under-theorized" (p. 99), and also researched. This research provides us with further empirical evidence of the effectiveness of art education in promoting self and social change, and of the mechanism by which such education can happen. Bushnell (2003) argues that: "Aesthetic education is about becoming aware of assumptions, challenging those assumptions, and subversion" (p. 437). Understanding how to help students engage in such challenging, of self and society, is a necessary precursor to helping them see themselves as transformative agents in a diverse, democratic society. Assisting students in understanding their ideological make-up becomes the job of teachers, artists, and conservationists to help restore the unity between education and community life. Such education has the ability to engage students "in a multicultural critical pedagogy which challenges them not only to examine their own beliefs about society, but also to create the artistic vision that will inspire others to take up the project of social transformation" (Noel, 2003, p. 18).

Programs like Arts-UP serve as models of success in educating its young community members to recognize the potentials of their own collective voice. Arts-UP, it seems, is one of the alternative organizations described by Trend (1992) as ideally situated to facilitate a renaissance of a people's pedagogy (p. 138). As all educators seek to employ effective methodologies for teaching and learning, collaborations with community conservation organizations can help students reach new levels of intrapersonal and interpersonal awareness through their art. As students construct their experiences in such programs, they can begin to understand their part in the process of changing not only their local community, but changing the world.

References

Albers, P. (1999). Art education and the possibility of social change. *Art Education*, *52*, 6-11.

Ament, E. (1998). Using feminist perspectives in art education. *Art Education*, *51*, 56-61.

Billings, M. (1995, January). Issues versus themes: Two approaches to a multicultural art curriculum. *Art Education*, 21-56.

Cathan, S. & Kocur, Z. (Eds.). (1996). *Contemporary art and multicultural education*. New York: Routledge.

Denzin, N. (1970). *The research act: A theoretical introduction to sociological methods*. Chicago: Aldine Publishing.

Desai, D. (Summer, 2002). The ethnographic move in contemporary art: What does it mean for art education? *Studies in Art Education,43*, 4, 307-23.

Freedman, K. (2004). Becoming a researcher in art education. Forming research questions. *Studies in Art Education*, *45*, 99-100.

Glesne, C. (2006). *Becoming qualitative researchers: An introduction*. Boston: Allyn and Bacon.

Greene, M. (1995). *Releasing the imagination: Essays on education, the arts, and social change*. San Francisco: Jossey-Bass.

Haynes, D. (1995, September). Teaching postmodernism. *Art Education*, 23-50.

Heath, S. B. & Roach, A. (1999). Imaginative activity: Learning in the arts during the nonschool hours. In E. B. Fiske (Ed.), *Champions of change: The impact of the arts on learning* (pp. 20-34). Washington, DC: The Arts Education Partnership and President's Committee on the Arts and Humanities.

Hicks, L. (2002). Feminist analyses of empowerment and community in art education. *Studies in art education*, *32*(1), 36-46.

Holloway, D. & Krensky, B. (2001). Introduction: The arts, urban education, and social change. *Education and Urban Society*, *33*, 354-365.

Kafka, F. (1915). *The metamorphosis*. Bantam Books, NY.

Kincheloe, J. (2004). *Critical pedagogy*. New York: Peter Lang.

Lowe, S. (2001). The art of community transformation. *Education and Urban Society*, *33*, 457-471.

Noel, J. (2003). Creating artwork in response to issues of social justice: A critical multicultural pedagogy. *Multicultural Education*, *10*(4), 15-18.

Smith, L., & Stevenson, L. (2003, September). *You want to be a part of everything: The arts, community and learning*. Retrieved February 25, 2005, from http://aep-arts.org/PDF%20Files/YouWantToBePart.pdf

Taylor, P. G. & Ballengee-Morris, C. (2004). Service-Learning: A language of "We." *Art Education, 57*(5), 6-12.

Trend, D. (1992). *Cultural pedagogy: Art, education, politics*. New York: Bergin & Garvey.

Ulbricht, J. (2005). What is community-based art education? *Art Education*, 6-11.

Wachoviak, F. D., & Clements, R. D. (2001). *Emphasis art: A qualitative art program for elementary and middle schools*. New York: Longman.

Wallot, J. A., & Joyal, B. (1999). School-community collaboration and artistic process. In R. L. Irwin & A. M. Kindler (Eds.), *Beyond the school: Community and institutional partnerships in art education* (pp. 29-35). Reston, VA: National Art Education Association.

A Practical Proposal for Special Education Teacher Induction

Leah Wasburn-Moses
Miami University, Oxford, Ohio

Abstract

Teacher attrition is a serious issue in the field of special education (Boe, Cook, Bobbitt, & Terbanian, 1998). Teachers leave the field in record numbers due to inadequate support in a variety of areas (Shoho, Katims, & Meza, 1998). However, the first few years of a special education teacher's career have been shown to be crucial for teacher retention (Singer, 1993). The author outlines a three-part plan for induction of special education teachers that directly addresses problems faced by teachers as well as issues in the field of special education. These three components include: IEP Assistance Teams, dual mentoring, and professional development workshops.

Introduction

Special education teacher burnout and attrition is an enormous problem in education. The annual attrition rate for special education teachers has been estimated to be between eight and ten percent (Whitaker, 2000). These teachers are leaving the field in much greater numbers than their peers in general education (Nichols & Sosnowsky, 2002). Nationally, there is a persistent annual shortage of about 33,000 fully certified special education teachers, with 98% of school districts reporting an immediate need for special education teachers (Bergert & Burnette, 2001).

These alarming statistics have prompted researchers to investigate the factors involved in teacher attrition, as well as the factors contributing to some teachers' commitment to remain in special education. Isolation, limited supports, and large caseloads are all highly related to attrition (Gersten, Keating, Yovanoff, & Harniss, 2001; Shoho et al., 1998). However, building administrators can play a crucial role in affecting teachers' sense of well-being and their decisions whether or not to leave their positions (Littrell, Billingsley, & Cross, 1994).

This proposal for a high quality teacher induction program for special education teachers involves three interrelated parts: IEP Assistance Teams, dual mentoring, and a series of professional development workshops. This program is based on research involving induction and professional development in both general and special education. It specifically targets areas of need that have been linked to attrition. The program also educates teachers in that it seeks to raise awareness of and help generate solutions to other persistent problems in special education.

IEP Assistance Teams

IEP Assistance Teams would meet with each novice teacher to help review students' IEPs . These teams would be composed of both new and experienced teachers and would be charged with assisting teachers in developing student action plans. These action plans are used to target students who are at risk for school failure and/or dropping out.

The purpose of this teaming is to initiate new teachers into a community of practice, attack some of the causes of special education teacher attrition, and develop strategies to ameliorate poor outcomes for students.

First, the study teams would provide new teachers entrée into a *community of practice*. Ball and Cohen (1999) cite the need for teachers to participate in communities of practice and to expand their knowledge about ways to understand and approach students. In addition, Stough and Palmer (2003) assert that the "expert" special education teacher is unique in his/her ability to assess and respond to students' individual needs. IEP Assistance Teams help address such critical issues.

Second, these meetings also have the capability to address some other leading causes of attrition such as excessive paperwork, isolation, insufficient support, and challenging students (Gersten et al., 2001; Nichols & Sosnowsky, 2002; Shoho et al., 1998; Wisniewski & Gargiulo, 1997). Researchers with the Center on Personnel Studies in Special Education (COPSSE) list eight problems with traditional special education teacher induction (Griffin, Winn, Otis-Wilborn, & Kilgore, 2002). Of these, four are directly addressed by IEP Assistance Teams: challenging students, lack of opportunity for collaboration, lack of professional development opportunities, and excessive paperwork. The teams would aid teachers in preparing IEPs and brainstorming solutions for addressing students' needs. For example, teams could create a Behavior Management Plan to be used across settings by special and general education teachers and school administrators (for example, see http://www.ericec.org/digests/prodfly.html for ERIC reports on instruction and management).

Third, outcomes among special education students are notoriously poor. Problems include high dropout, illiteracy and unemployment rates, and increased incarceration rates (Archwarnety & Katsiyannis, 2000; Malian & Love, 1998; Wagner & Blackorby, 1996). Malian and Love (1998) recommend identifying potential dropouts, developing interventions, and establishing support networks. The team would consider individual students' situations and assist the novice teacher in creating a cooperative plan to aid these students.

Dual Mentoring

Dual mentoring is the second facet of the induction plan. This arrangement would involve the novice both with a special education and a general education teacher. Preferably, the general education teacher is one who teaches in the same content area (Whitaker, 2000). This "challenge of dual socialization" was first discussed by Pugach (1992, p. 144). She predicted the growing need for special education teachers to be socialized into the school's culture. Ideally, dual mentoring would encourage the mentors and mentee to work together, modeling collaboration and bringing the separate spheres closer.

Mentoring is highly touted both in general and in special education, but it is also acknowledged that simply assigning a mentor(s) to a novice teacher is insufficient (Griffin et al., 2002; Whitaker, 2000). This is why the administrator must take steps to ensure a quality experience, including training, release time to visit each other's classroom, and/or shared dialogue journals.

Both Ball and Cohen (1999) and Feiman-Nemser (2001) characterize some of the main goals of induction as helping the novice create an identity, gain a repertoire, and learn how to learn from practice. This mentoring plan offers new teachers the opportunity to reflect upon and learn from their own practice and the practice of others. It would also provide collegial socialization. In this case, socialization would involve sharing general information about responding to students' individual needs as well as precise, technical knowledge about laws, paperwork, unwritten policies and administrative duties (Pugach, 1992; Stough & Palmer, 2003).

General education socialization is equally important to the new special education teacher. General education teachers offer the opportunity to do what Ball and Cohen (1999) term "developing and expanding their ideas about learning" (p. 9). For years, special education has overlooked general education curriculum and instructional techniques in its all-encompassing focus on the individual learner (Pugach, 1992; Winzer, 2000). Preservice programs in special education often do not emphasize content-area teaching and learning, although those teachers often find themselves teaching content-area classes (Mastropieri, 2001; McKenzie, 1995). As predicted shifts toward inclusion and collaboration take effect, and as demands of "access to the general education curriculum" for all students become universally recognized, such dual initiation is imperative (President's Commission on Excellence in Special Education, 2002; Sabornie & deBenttencourt, 1997; Wehmeyer, Sands, Knowlton, & Kozleski, 2002).

Professional Development Workshops

The final part of the induction plan involves *targeted professional development workshops*. These workshops would involve new teachers and their mentors in discussions and activities focused on problems of the field, personal issues, and learning through cases of others' practice. It would tap into community resources for leadership, such as universities, community agencies, and school leaders.

Ball and Cohen (1999) mention the importance of basing professional learning activities on changes that reformers wish to make in education. Several critical issues in special education are directly relevant to the daily practice of special education teachers. These issues include behavior management, student outcomes, parent involvement, and inclusion. Reformers are working in all of these areas to improve outcomes for individuals with disabilities (Malian & Love, 1998; Wagner & Blackorby, 1996; Zigmond, 2003). Important voices are be heard and connections to be made outside the school's normal boundaries (e.g. agencies serving adults with disabilities, university personnel, parent advocacy groups).

A focus on personal issues in professional development could help ameliorate teacher attrition (Gersten et al., 2001). These issues could include stress and time management and communication with general education teachers, administrators, and parents, all of which are known contributors to burnout and attrition (Gersten et al., 2001; Nichols & Sosnowsky, 2002; Sires & Tonnsen, 1993). These factors are also related to the kind of support special education teachers desire from school administrators (Littrell et al., 1994). In addition, paperwork completion, while seemingly trivial, would be a critical topic, as interference caused by paperwork has been named as one of the primary causes of attrition (Council for Exceptional Children, 2001). Problems with current professional development for special education teachers mirror those faced by their general education colleagues. These initiatives are often characterized by varied, disconnected topics with little or no follow-up, and generally do not address working with students with disabilities (Norman, Caseau, & Stefanich, 1998; Stodden, Galloway, & Stodden, 2003).

The workshops could also allow teachers and mentors to examine practices of other teachers through vignettes, or case studies. This is one activity recommended by Ball and Cohen (1999) as essential for broadening teacher knowledge, and would contribute to teachers' "repertoire" as outlined by Feiman-Nemser (2001). Such analysis of practices would expose teachers to wider range of possibilities for special education practice (Conderman & Katsiyannis, 2002; Fuchs & Fuchs, 1994; Winzer, 2000).

Conclusion

The Study of Personnel Needs in Special Education found that high quality inservice, as based on teacher perception, was related positively with measures of teacher quality (SPeNSE, 2002). The proposed induction program for special education teachers is comprehensive and practical for the school administrator to implement. It includes weekly IEP team meetings, a dual mentoring program with

flexible components, and a series of professional development workshops centered around problems of the field, personal issues, and critique of others' teaching. It combines information from both the general and special education literatures on teacher induction and uses knowledge that is specific to special education, including problems faced by teachers and students. The program provides new teachers with plans for working with specific students, outlets to "learn in and from practice" both general and special education, and opportunities to engage in persistent debates and problems in the field. It attempts to lay the foundation for engagement in communities of practice, involving both general and special education teachers in working together for the success of all students.

References

Archwarnety, T., & Katsiyannis, A. (2000). Academic remediation, parole violations, and recidivism rates among delinquent youths. *Remedial and Special Education, 21*, 161-170.

Ball, D. L., & Cohen, D. K. (1999). Developing practice, developing practitioners: Toward a practice-based theory of professional education. In G. Sykes (Ed.), *Teaching as the learning profession: Handbook of policy and practice* (pp. 3-32). San Francisco: Jossey-Bass.

Bergert, S., & Burnette, J. (201). Educating exceptional children: A statistical profile. Washington, DC: Office of Educational Research and Improvement (ERIC Document Reproduction Service No. ED452649).

Boe, E. E., Cook, L. H., Bobbitt, S. A., & Terbanian, G. (1998). The shortage of fully certified teachers in special and general education. *Teacher Education and Special Education, 21*, 132-149.

Conderman, G. J., & Katsiyannis, A. (2002). Instructional issues and practices in secondary special education. *Teacher Education and Special Education, 23*, 167-179.

Council for Exceptional Children. (2001). New study on special education working conditions. *CEC Today, 8,* 1.

Feiman-Nemser, S. (2001). From preparation to practice: Designing a continuum to strengthen and sustain teaching. *Teachers College Record, 103*, 1013-1055.

Fuchs, D., & Fuchs, L. S. (1994). Inclusive schools movement and radicalization of special education reform. *Exceptional Children, 60*, 294-309.

Gersten, R., Keating, T., Yovanoff, P., & Harniss, M. K. (2001). Working in special education: Factors that enhance special educators' intent to stay. *Exceptional Children, 67*, 549-567.

Griffin, C. C., Winn, J. A., Otis-Wilborn, A., & Kilgore, K. L. (2002). *New teacher induction in special education*. Center for Personnel Studies in Special Education. Retrieved October 7, 2002, from the World Wide Web: www.coe.ufl.edu/copsse/index.htm

Littrell, P. C., Billingsley, B. S., & Cross, L. H. (1994). The effects of principal support on special and general educators' stress, job satisfaction, school commitment, health, and intent to stay in teaching. *Remedial and Special Education, 15*, 297-310.

Malian, I. M., & Love, L. L. (1998). Leaving high school: An ongoing transition study. *Teaching Exceptional Children, 30*(1), 4-10.

Mastropieri, M. A. (2001). Is the glass half full or half empty? Challenges encountered by first-year special education teachers. *The Journal of Special Education, 35*, 66-74.

McKenzie, R. G. (1995). Special educators as secondary content teachers: How justifiable? *Teacher Education and Special Education, 18*, 275-280.

Nichols, A. S., & Sosnowsky, F. L. (2002). Burnout among special education teachers in self-contained cross-categorical classrooms. *Teacher Education and Special Education, 25*, 71-86.

Norman, K., Caseau, D., & Stefanich, G. P. (1998). Teaching students with disabilities in inclusive science classrooms: Survey results. *Science Education, 82*, 127-146.

President's Commission on Excellence in Special Education. (2002). *A new era: Revitalizing special education for children and their families*. Washington, D.C.

Pugach, M. C. (1992). Uncharted territory: Research on the socialization of special education teachers. *Teacher Education and Special Education, 15*, 133-147.

Sabornie, E. J., & deBenttencourt, L. U. (1997). *Teaching students with mild disabilities at the secondary level*. Upper Saddle River, NJ: Merrill.

Shoho, A. R., Katims, D. S., & Meza, P. (1998). The alienation of special education teachers. *ERS Spectrum, 16*(4), 18-23.

Singer, J. D. (1993). Are special educators' career paths special? Results from a 13-year longitudinal study. *Exceptional Children, 59*, 262-279.

Sires, C., & Tonnsen, S. (1993). Special education: A challenge for principals. *NASSP Bulletin, 77*, 8-11.

SPeNSE. (2002). *Recruiting and retaining high-quality teachers*. Study of Personnel Needs in Special Education. Retrieved March 31, 2003, from the World Wide Web: www.spense.org

Stodden, R. A., Galloway, L. M., & Stodden, N. J. (2003). Secondary school curricula issues: Impact on postsecondary students with disabilities. *Exceptional Children, 70*, 9-26.

Stough, L. M., & Palmer, D. J. (2003). Special thinking in special settings: A qualitative study of expert special educators. *The Journal of Special Education, 36*, 206-222.

Wagner, M. M., & Blackorby, J. (1996). Transition from high school to work or college: How special education

students fare. *Special Education for Students with Disabilities, 6*(1), 103-120.

Wehmeyer, M., Sands, D. J., Knowlton, H. E., & Kozleski, E. (2002). *Teaching students with mental retardation: Providing access to the general curriculum.* Baltimore, MD: Paul H. Brookes.

Whitaker, S. D. (2000). Mentoring beginning special education teachers and the relationship to attrition. *Exceptional Children, 66*, 546-566.

Winzer, M. A. (2000). The inclusion movement: Review and reflections on reform in special education. In M. Winzer & K. Mazurek (Eds.), *Special education in the 21st century: Issues of inclusion and reform* (pp. 5-26). Washington, D.C.: Gallaudet University.

Wisniewski, L., & Gargiulo, R. M. (1997). Occupational stress and burnout among special educators: A review of the literature. *Journal of Special Education, 31*, 325-346.

Zigmond, N. (2003). Searching for the most effective service delivery model for students with learning disabilities. In H. L. Swanson, K. Harris, & S. Graham (Eds.), *Handbook of learning disabilities* (pp. 110-122). New York: Guilford Press.

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. Contributors acknowledge by virtue of their submission to the journal that they will consent to have their work available internationally through the EBSCO portal, as per agreement with the MWERA.

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

Multi-context Use of Language: Toward Effective Thinking and Planning for Curriculum

Anne D'Antonio Stinson
William Chandler
M. Virginia Epps
Melissa Frieberg
University of Wisconsin-Whitewater

Abstract

The flexible, multi-contextual use of language is essential to integrated learning and thinking. Likewise, learning and thinking in an integrated way is essential to multi-dimensional teaching. This study examines the ways pre-service secondary teachers define their subjects. Students enrolled in a secondary reading methods course were asked to provide multiple definitions for a predetermined list of critical vocabulary terms common to multiple disciplines. We used these definitions to measure participants' level of sophistication with regard to the multi-context use of language. Participants' responses illuminated implications for curriculum development in secondary schools including the need for models for pre- and in-service teachers of authentic integrated curriculum.

Introduction

Central to every teaching/learning situation is the preparation and presentation of content. Teachers and other instructional specialists work at great lengths to organize learning experiences that will insure student engagement and bring about significant intellectual change. One way of accomplishing this goal has been to foster recognition of real and perceived linkages between topics and content. This practice has been described in schools and professional development workshops using various terms such as connected curriculum, integrated curriculum, cross-disciplinary curriculum, and interdisciplinary curriculum. Implicit in each of these terms is the idea that curriculum is as multi-dimensional as the students' lives. Thus, to make meaningful connections to their lives, students must experience multi-dimensional, integrated curriculum. Unfortunately, the reality of integrated curriculum is often the trivialization of content by practices that not only fail to make these connections, but also suppress the processes of meaningful content acquisition. This is the case if one assumes that curriculum is integrated for writing if any writing is done, for math if any calculations are performed, for science if any natural phenomenon is mentioned, and/or for art if any visual is created, regardless of developmental or educational appropriateness.

Curriculum that connects to students' lives results in a more significant development of knowledge, skills, and dispositions than will traditional, departmentalized, content-area instruction. When substantial integration of curriculum is present, students perceive and respond to concepts across disciplinary boundaries rather than as elements of discrete subjects, and communicate these concepts through multiple representational forms. They exhibit flexible thinking and their work reflects multiple perspectives. Those outside of the classroom begin to hear a different kind of conversation about what is being learned. Importantly, however, while all of this synthesis is occurring, the parent disciplines maintain their integrity.

Background to the Study

In 1997 two of the authors organized a graduate level summer workshop entitled "Math and the Arts." This three-week seminar was intended to provide an engaging setting in which in-service teachers could thoughtfully consider how the arts could be employed in the teaching of mathematics. The course was team taught by three content area specialists, one from each of the following: math, music, and art. The format was dialogic, rather that didactic. While the faculty provided content expertise from the different disciplines, they also sought to model dialogism in order to break down, as Bakhtin (1981) suggests, the barriers created by context-specific languages. Ten in-service teachers signed up for this graduate course. The seminar resulted in intense, productive, mediating discussion between faculty and students (as well as among students and among faculty) and some practical curricular applications that the in-service teachers have subsequently reported to the instructors. We (the current group of authors) have continued the discussion of integrating disciplinary subject matter, in part as a result of the recognition that any integration of content should be grounded in more than the production of quick-and-easy, simplistic, hands-on learning activities.

A part of our extended discussion focuses on discipline-specific vocabulary. Despite differences between disciplines, there are many commonly used terms whose definitions differ, to varying degrees, across the disciplines. We decided to explore the potential of these shared terms as means for grounding or planning and organizing meaningful integrated curriculum development. This study investigated how pre-service teachers know and understand words as part of their

own area of study, as well as how they might extend into other academic disciplines.

Review of the Literature

Curriculum is a manifestation of how humans think. Curriculum, as it is formulated for any of its applications (schooling, job training, and even the passing on of family cooking traditions), is an attempt to order a set of ideas so that those ideas can be captured and that understanding can be developed by the receiving individual or "learner." The varied terms educators associate with curricular design give evidence to methods of ordering what is to be learned. The term "sequential curriculum" suggests a gradual progress toward learning goals through a series of learned skills, while "spiral growth curriculum" suggests planned repetition as a way of working toward learning goals. Even the term "thematic curriculum" suggests a collection of concepts based on links associated with a chosen topic of study. However, when curriculum is practiced in real-life settings, the single course of study occurs as a gestalt; the ideas communicated are greater than the sum of their parts. There is a holistic presence about the things that are taught. Within the educational system (and for our purposes) this concept is known as integrated curriculum. The term describes a curricular format that attempts to frame the world's reality as interacting and interdependent. By moving beyond more narrow, discipline-centered orientations it is intended that deeper and more personal learning will occur.

Dewey framed a view of the world as "an impressive and irresistible mixture of sufficiencies, tight completenesses, order, recurrences which make possible prediction and control, and singularities, ambiguities, uncertain possibilities, processes going on to consequences as yet indeterminable" (Dewey, 1958, p. 47). To navigate this complexity Dewey encouraged reflective thinking. Dewey urged that reflection should be grounded in an education of experience (Dewey, 1997). Powerful knowing would result from an interaction with real-world problems. Since the situational nature of the world is multidimensional and affords viewing from a number of perspectives, his proposal describes both a need for and a practice of curricula that is integrated.

Goodman's (1968) description of how an individual comes to understand the world also offers insight into an integrated curriculum format. For Goodman every time a person is confronted with a perceptual phenomenon there is an opportunity for interpretation. Nothing that is experienced need be interpreted as it has been; instead, each encounter requires a renegotiation of object and setting by the individual. Worldmaking involves the continuous constructing of a point of view that presents the sum of previous interpreted experiences. The association of personal beliefs to the construction of knowledge is also central to the work of Greeno (1989) regarding situated cognition. While he suggested that critical thinking "has to do with whether individuals think reflectively" (p. 139), he also recognized the role of contextuality in learning. One's beliefs and subsequent actions are best established within a context. An integrated curriculum attempts this by setting at least a mental context for that which is to be learned.

Currently, two additional conditions exist that give currency to the practice of integrated curriculum. One is the theoretical work of Howard Gardner. The second is the standards-based movement that has grown out of the Goals 2000 efforts and continues with No Child Left Behind. Gardner (1993) has noted that we all have propensities for at least seven basic ways of knowing. As this theory has been played out among teachers and curriculum developers, these "multiple intelligences" realize avenues of access for learners. Curriculum is thus organized in ways that facilitate learning for each student. While this is not exactly an integrated curriculum organization, it does lead educators to recognize that content has multiple dimensions and that learning and knowing can be personalized to the intellectual aptitudes of the learner. Additionally, the current rush to standards has had an impact on curriculum development. Within recent years, in response to Federal mandates, professional organizations have developed "standards," or statements of what learners should know at the conclusion of their mandatory education. Within each of these collections of standards is a call for "making connections between . . . disciplines" (e.g., National Standards for Arts Education, 1994, p.72). As a result, curriculum developers are seeking ways to incorporate the skills of analysis and synthesis into the disciplines of the curriculum so as to bring about deep and reflective knowing across the disciplines.

Integrated curriculum is not a new concept; however, there are barriers that prevent it from becoming a substantial reality in classrooms. Curriculum must be organized in ways that encourage more than the linking of knowledge across disciplines. Integrated curriculum is an educational endeavor that seeks to engage learners in the mediation of meaning. Vygotsky (1986/1934) described that word meanings evolve "with the various ways in which thought functions" (p. 217). Understanding core terms and their specific meaning within multiple contexts is central to being able to synthesize meaning across disciplines.

Methodology

Purpose of the Study

The current study represents an effort to investigate how pre-service teachers operationalize the use of language in multiple contexts. Our intent was to investigate how pre-service teacher education students organized and articulated their own working definitions for these terms. This investigation provided insight into the difficulty those students will have when they, in essence, repeat the task in their own curriculum development activities as in-service professionals. We decided to focus on how shared vocabulary might provide an important stepping-off point for meaningful cross-curricular instruction. The following question guided our

research: Do students perceive words as having discipline-specific meanings and are these meanings associated with the participants' major or minor area of study?

Data Collection

Initially, we each gathered a list of words which we individually believed had both specific or "special" meaning in our respective disciplines (art, language arts, math, and science), as well as different special meanings in one or more other content areas and/or in general usage. We combined these lists to form an alpha-list of 160 vocabulary terms. Once this list was complete, each of us reviewed the list with the following direction: "Eliminate any words that you think *will not* have content-specific meaning across two or more disciplines." All words eliminated by at least one team member were dropped from further consideration. The resulting list contained eighty words, one half of the original list.

We then reviewed the eighty-word list in order to select the best words, i.e., those words that promised to elicit the richest and most varied responses to our request for multiple definitions. The art specialist among us selected eighteen words, the language arts specialist selected thirty-three, the math specialist selected fourteen, and the science specialist selected twenty-seven words. A short list of the eleven words selected by at least three of the four content area specialists resulted. We randomly arranged these eleven words for use in the survey instrument. The survey instrument listed the following eleven words in this order: rhythm, balance, space, set, interaction, intensity, analogous, pattern, density, composition, and contrast.

Participants

Participants in this study were enrolled in one of two sections of a required content area literacy course taught by one of the researchers. The students were pre-service teachers in secondary education who either held baccalaureate degrees or were seniors who anticipated graduation within one or two semesters. Each of the participants had a major in one of the following content areas: art education, business education, English/language arts, English as a second language, foreign language, health, math, music education, physical education, one of the natural sciences, or social studies. Seventy-six pre-service teachers of the 85 enrolled in the course participated in this study.

The Survey Instrument

The survey instrument asked students to provide information about their major(s), minor, gender, and degrees held or expected. No other demographic information was collected.

The instrument provided space for three separate definitions for each of the eleven terms. A researcher instructed the participants to complete the demographic information and then to think of as many definitions for each of the terms as they could and record one in each block. While students were free to elect not to have their definitions included in our data pool, they were required to provide the definitions for use in a class discussion of effective content-area vocabulary instruction.

Coding

Schloss & Smith (1999) suggest that categories of content will emerge as a result of an initial review of the data (p. 190). The review of our data suggested several categories of response (see Table 1). We each prepared a master list of definitions to use as a reference during the coding process; each content-area specialist provided discipline-specific definitions for all eleven words. We then independently coded

Table 1
Categories and codes for analysis of definitions.

A. Usage	01	General
	02	Specific, i.e., definition associated with a discipline, but not the student's major or minor
	03	Specific, i.e., definition associated with a discipline which is the student's major
	04	Specific, i.e., definition associated with a discipline which is the student's major
	05	Technical, i.e., definition given is tightly restricted to a discipline. E.g., mass/volume, g/ml, or g/cm^3, but not the student's major or minor
	06	Technical, i.e., definition given is tightly restricted to the discipline of the student's major
	07	Technical, i.e., definition given is tightly restricted to the discipline of the student's minor
	08	Slang
B. Part of Speech	01	Adjective
	02	Noun
	03	Verb
C. Concept Accuracy	01	Correct; definition given is consistent with accepted usage in at least one discipline or in general usage
	02	Incorrect; definition given is not consistent with accepted usage in at least one discipline or in general usage
	03	Incomplete or partially correct; definition too vague to recognize

the raw data for usage (general, specific, or technical), association with the participants' majors and minors, parts of speech, and concept accuracy.

Data Analysis

As indicated above, two questions guided our investigation of how pre-service teachers operationalize the use of language in multiple contexts. We hypothesized that the answers to each of our questions would be affirmative.

Question #1 Do students perceive words as having discipline-specific meaning?

Sub-question a: Do the students submit multiple definitions for each of the given words? The participants were asked to provide as many definitions (up to three) for each of the terms as they could think of. In practice, they provided from none to three definitions. We have provided the definition counts for the two words that the research team felt to be most informative for the purposes of this article: *pattern* and *balance* (see Tables 2 and 3). Of the two, *balance* appears to be the more productive, resulting in 147 definitions as opposed to 113 definitions for *pattern*. As a second indicator of the greater productivity of *balance* when compared to *pattern*, note that for *balance*, 71% of the participants produced more than one definition while only 47% of the participants produced more than one definition for *pattern*. The answer to this sub-question then becomes affirmative; yes, the participants did record multiple definitions for each term. However, the degree of productivity varied and it is important to note that this sub-question does not consider the issues of redundancy and accuracy.

Sub-question b: Are the definitions given related to a specific discipline? We hypothesized that the participants would provide definitions linked to at least one specific discipline. Tabulation of the researchers' coding showed unexpected variability; even so, some trends became evident (see Tables 4 and 5). Fifty-nine percent of the definitions for *pat-tern* were associated with a specific discipline while 55% of the definitions for *balance* were associated with a specific discipline. Therefore, the answer to sub-question b must be an unenthusiastic or weak "yes."

Sub-question c: Are these definitions generally accepted by at least one discipline? Because we initially hypothesized that the first definition the participants gave would be associated with the discipline of their major or minor, we expected to see definitions that would generally be accepted by at least one discipline among the definitions submitted. Again, an unexpected level of variability among the researchers' coding confounded analysis and interpretation; however, this variability in itself became instructive and stimulated worthwhile discussion of the problem of multi-context meanings of key concepts in the respective disciplines of the arts, mathematics, literacy, and the natural sciences. Tables 6 and 7 show that only 34% and 45% of the definitions for *pattern* and *balance,* respectively, were coded by at least three of the four coders as being consistent with accepted knowledge in at least one discipline; therefore, the answer to the third sub-question must be "no."

Question #2 Are the first definitions reported tied to the participants' major or minor?

The coders classified only 2% of the responses for *pattern* as "first entry tied to major." For *balance*, only 14 % were convincingly classified as "first entry tied to major." The answer to question two also must be "no."

Findings and Implications

Implications for Further Research

We had hypothesized that secondary education students would identify terms first with their major or minor areas of study when giving definitions; additional definitions would demonstrate their ability to cross over to another content area or to general usage. However, we found little to sup-

Table 2
Definitions provided for pattern

Number of definitions per participant	Number of participants	Total definitions produced	Percent of the participants
0	3	0	4%
1	37	37	49%
2	32	64	42%
3	4	12	5%
Total	76	113	100%

Table 3
Definitions provided for balance

Number of definitions per participant	Number of participants	Total definitions produced	Percent of the participants
0	1	0	1%
1	19	19	28%
2	37	74	49%
3	17	51	22%
Total	76	146	100%

port our hypothesis due to the form the students' definitions took. In the current study, students were asked to generate definitions for the indicated terms used within two or more content disciplines. Often those definitions tended to be vague and general in nature rather than tied to a specific discipline. Although there were some definitions that were able to be identified with specific disciplines, they usually did not completely match the accepted definition nor did they necessarily match with the students' discipline. One possible way to better test this hypothesis in a future study would be to provide several correct (according to different disciplines) definitions listed and ask students to list them in the order of their own primary understanding of the term.

As noted in the methodology section, we developed a coding system for this study to allow us to analyze the definitions provided by students for content area usage and con-

Table 4
Frequency of category assignation for pattern

Type of definition	Number of coders	Codes	Number of definitions	Percent of definitions
General	3 or more	1	46	41%
Special	4	2, 3, or 4	12	11%
	3	2, 3, or 4	24	21%
	2	2, 3, or 4	31	27%
Technical	4	6 or 7	0	0%
	3	6 or 7	0	0%
	2	6 or 7	0	0%
Total			113	100%

Table 5
Frequency of category assignation for balance

Type of definition	Number of coders	Codes	Number of definitions	Percent of definitions
General	3 or more	1	65	44%
Special	4	2, 3, or 4	10	7%
	3	2, 3, or 4	28	19%
	2	2, 3, or 4	43	29%
Technical	4	6 or 7	0	0%
	3	6 or 7	0	0%
	2	6 or 7	0	0%
Total			146	99%

Table 6
Concept accuracy or consistency with accepted knowledge for pattern

Number of coders	Codes	Number of definitions	Percent of definitions
4 of 4	1 (consistent)	13	11%
3 of 4	1 (consistent)	26	23%
≥2 of 4	2 (inconsistent)	16	14%
≥2 of 4	3 (vague)	52	46%
mixed	1,1, 2, & 3	6	5%
Total		113	99%

Table 7
Concept accuracy or consistency with accepted knowledge for balance

Number of coders	Codes	Number of definitions	Percent of definitions
4 of 4	1 (consistent)	33	23%
3 of 4	1 (consistent)	32	22%
≥2 of 4	2 (inconsistent)	20	14%
≥2 of 4	3 (vague)	39	27%
mixed	1,1, 2, & 3	22	15%
Total		146	101%

cept accuracy. However, due to the incomplete or inaccurate definitions given by the participants, and partially due to an incomplete or inaccurate understanding of other definitions by the researchers themselves, this coding did not result in reliable findings; there was no inter-rater reliability. We believe two factors contributed to this breakdown. Even with definitions provided by the research team, the researchers themselves had difficulty identifying (1) the level of "correctness" consistently across disciplines, and (2) which discipline (if any) a definition belonged to. For example, many participants defined the term "pattern" as a guide used for a sewing project. We were forced to ask ourselves, is this an accurate definition in the field of family and consumer education or a term of general usage? A future study, as indicated above, should start with clear definitions of terms for various disciplines, and those doing the analysis of the student responses should have clear agreement on what defines each code.

Finally, a future study might ask students to do more than provide or chose definitions. It might also ask students to (1) decide what is the key or most salient part of the definition of the term across all disciplines or (2) identify consistencies or inconsistencies in the definitions across disciplines. These efforts may lead to a better picture of the readiness or ability of pre-service teachers to engage in cross-disciplinary or integrated curriculum instruction.

Implications for Practice

As teacher educators look to create opportunities for pre-service teachers to develop curriculum and instructional strategies for integrating content disciplines, we must stress the importance of the nature and integrity of each discipline. Without thoughtful application in the classroom, integration may become artificial or inaccurate, and may force associations among concepts. Connections across disciplines may be tenuous and the content of one or more disciplines may not be in line with grade level standards or curriculum requirements. Efforts to correct these problems are hampered by students', teachers', and a community's perceptions of disciplines as distinct and mutually exclusive. Our research points out that pre-service secondary teachers have difficulty conceptualizing terms as being used in multiple disciplines and are unable to distinguish nuances of those definitions that apply more to one field than another. Similarly, we ourselves, as faculty from different disciplines, were often in disagreement over the "correctness" of definitions from areas other than our own. Much work needs to be done in the area of teaching through integrated curriculum.

First, we propose that faculty, at both the post-secondary and secondary school levels, need to spend more time in dialog about the essence of concepts and skills to be taught and how those concepts or skills can be illustrated, applied, understood, or extended through various disciplines. The scheduling structure of many middle schools allows team planning time that should allow those types of discussions to take place, if that time is properly used. However, the departmentalized schedule of high schools and universities does not provide the opportunity or structure to allow, much

less encourage, these discussions. Schools operating under different models, such as charter or alternative schools, especially need to build in this feature since they are the types of schools most likely to be using some form of cross-disciplinary teaching. The discussion not only builds a basis for instruction, but also helps participants to generate language that is inclusive and clear.

Second, cross-disciplinary teaching requires the use of clear and inclusive language to connect ideas and concepts as they really exist, not as we isolate them for the purpose of teaching a discipline. For secondary teachers and their students, who have been educated to see disciplines as separate, efforts need to be made to point out the consistencies, similarities, ties, and fine discriminations among content areas through language, images, symbols, and actions. Too often teachers expect students to make these generalizations or distinctions with little practice or guidance, or may not recognize them as correct when students do. Integrated curriculum does not require the use of language that is so general as to encompass aspects of many disciplines at the same time, nor is it desirable to do so. Such efforts would lose the beauty and the essence of both language and the discipline. What is required is an understanding of how the nuances of language provide insight and meaning within and across disciplines and how language in the context of one discipline can help illuminate another.

Third, there is a need for models for pre- and in-service teachers of true integrated curriculum and instruction. Interestingly, institutions of higher education often provide interdisciplinary courses for students to take, but rarely are they found in schools or colleges of education. Although we sometimes combine some methods of teaching various content areas, they are generally taught as separate courses by specialists in the respective fields. We need to explore ways to structure courses, content, assignments, and expectations to support content integration.

References

Bakhtin, M. (1981). *The dialogic imagination: Four essays*. Austin, TX: University of Texas Press.

Dewey, J. (1960). *How we think*. Lexington, MA: D. C. Heath and Company.

Dewey, J. (1997). *Experience and education*. New York: Touchstone.

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Goodman, N. (1978). *Ways of worldmaking*. Indianapolis, IN: Hackett Publishing Company.

Greeno, James G. (1989). A perspective on thinking. *American Psychologist, 44*(2), 134-141.

Schloss, P. & Smith, M. (1999). *Conducting research*. Upper Saddle River, NJ: Prentice-Hall.

Vygotsky, L. (1986/1934). *Thought and language*. Cambridge, MA: The MIT Press.

The Structural Validity of the Ohio Ninth Grade Proficiency Test

Yonghee Kim
William E. Loadman
The Ohio State University

Abstract

The structural validity of the Ohio Ninth Grade Proficiency Test was investigated with the test scores of 713 eighth graders who had their first opportunity to take the test in March 1999 in a large urban school district in Ohio. The study examined a hypothetical structure of the test to determine its fit to the data. Confirmatory factor analysis was performed, and several fit indices including χ^2 statistics and the root-mean-square error of approximation were examined. The theoretical structure of the test was first found to be valid, and then the invariance of the structure was examined across males and females. Both the tests of structural fidelity and invariance were cross-validated using a sample of 735 eighth graders who took the parallel test in March 2002. The results provided evidence of the invariance of factor loadings and factor variances in both years. Factor covariances were found to be invariant in 1999, but this was not confirmed by 2002. Results consistently showed the lack of invariance of error variances.

Since excellence in education first received national attention in the early 1960s, competency testing has been a major component of policy makers' efforts to reform education (Green, 1995). In the 1980s, when the issues of excellence were again brought before the public, there was pressure on state of Ohio legislators to develop some form of competency testing of basic skills to be implemented in school curricula. Statewide administration of the Ohio Ninth Grade Proficiency Test (ONGPT) began in the fall of 1990, and the test is now required by legislation. The ONGPT was instituted as a means to hold students and school districts accountable for a minimum level of educational achievement so that students would not graduate from an Ohio high school without basic skills for lifelong learning (Ohio Department of Education, 2001).

Validation is a fundamental prerequisite to test development and proper interpretation of test scores (Messick, 1988, 1995; AERA, APA, NCME, 1999, p. 9). In particular, the nature of responsible high stakes testing mandates serious attention to quality control due to its critical role in educational decisions. The educational decisions related to high stakes testing include setting content standards and performance standards (cutoff scores), and determining rewards and consequences for performance of local education authorities (school districts) and individual educational units. The consequences for individual students are high as well with graduation requirements largely based on "passing" tests in multiple subject areas of the high stakes tests like the ONGPT.

School proficiency tests have drawn researchers' interests, and various validity aspects have been investigated thus far. Smiko, Denny, and Twing (2000) provided the content and curricular validity evidence of the Texas Assessment of Academic Skills Test by scrutinizing each step of the test

development. Behrman (1998) analyzed the content of the reading part of the New Jersey's High School Proficiency Test to determine whether the test could account for students' differential content schemata. In the work of DeMars (1998), the effect of response format test performance in the Michigan High School Proficiency Test was examined across gender. The study analyzed how males and females differentially performed in the multiple-choice items and the open-ended items. The Michigan High School Test was also reviewed and criticized by Shafter (1997) on the language section. The test was found to create panic, uniformity, and a subversion of sound language arts practice even if it was designed to foster accountability, rigor, and standards. Lane and Stone (2002) proposed a validation study examining the consequences of a state assessment and accountability programs. They argued the need for consequential validation and presented a way to model the relationship between performance gains and school, principal, teacher, and student variables. Structural fidelity and the invariance of the internal structure of a statewide proficiency test were studied as well. Powers and Jones (1992) examined the structure of the California Achievement Tests and investigated the invariance across race and sex groups.

The ONGPT has been validated from several aspects regarding the use and interpretation of test scores. For instance, the content aspect was investigated. Chandler and Brosnan (1995) and Kunicki and Osborne (1992) analyzed and compared the content domains of the proficiency test against the content domains covered in various texts adopted by Ohio teachers. Several correlation studies investigated the criterion validity of the ONGPT using classroom grades or other proficiency test scores. Gibson (1997) used classroom grades in identifying students likely to fail the ONGPT. Hull and Tache (1993) examined the correlations of the test

scores in the ONGPT and the Iowa Test of Basic Skills, and Lanese (1992) used the Cleveland District test as a predictor of the ONGPT test performance. Other variables that were examined are poverty (Gallagher, 1993), the time period between instruction and testing (Heiney, 1998), teacher attitude toward the standardized testing (Wiesenauer, 2003), and demographic variables differentiating school districts (Shroder, 2002). There are more than a few studies available validating the content and criterion aspects of the ONGPT. However, a structural validation or a model invariance test across identifiable populations has never been published for the Ohio high school proficiency test.

Valid test use and interpretation depend on the establishment of internal validity evidence that the structure of the test supports the reported scores and their uses and interpretations (Messick, 1994). The latest Standards for Educational and Psychological Testing (AERA, APA, & NCME, 1999) also recommends that evidence of the structural validity of tests should be gathered (see standards 1.11 and 1.12, p. 20). Examining the internal structure of a test or assessing the *structural fidelity* (Loevinger, 1957) determines whether the structural relations among test items parallel the structural relations of other manifestations of the trait being measured (p. 661). In other words, a measure should relate to other observed variables in a way that is consistent with theoretically derived predictions based on the test design blueprint. When one examines the relations between a measure of one construct and other observed variables indicating other constructs, one expects their empirical association to parallel the theoretically specified associations the test makers have specified. To the extent that they do, structural validity exists (Bollen, 1989, p. 188).

In addition to structural validity evidence, Messick (1993, p.81) addressed the importance of the generalizability of the structure across identified population groups. Fair test use implies that decisions based on the test will be equally appropriate regardless of an individual's group membership (Guion, 1966; Thorndike, 1971). A test should not measure ability differently for identifiable groups unless it has a purposeful intent to measure different abilities for different groups and the intent is demonstrated (Rock, Werts, & Flaugher, 1978). When there is evidence that a test measures different constructs in different gender, racial, or cultural groups, it is not appropriate to make any statistical inferences (e.g., testing for mean difference between groups of test takers) (AERA, APA, & NCME, 1999, p. 78; Byrne & Shavelson, 1987; Rock et al., 1978).

This study served two purposes. First, it evaluated the theoretical five-factor model of the ONGPT composed of writing, reading, math, citizenship, and science sections, for which separate scores are derived to obtain evidence of the structural validity of the high school proficiency test. The second purpose of this study was to determine the extent of invariance of the test structure across males and females. Even if the test structure was valid for the total population, it might have measured one construct for males and a differ-

ent construct for females; and it might have measured the same construct for the two groups but with different patterns. In each of these instances, the test would be psychometrically biased if it were interpreted and used uniformly for both groups (see Rock et al., 1978).

The advantage of examining the factorial structure and its invariance for this high school graduation test is that what we learn may provide a note of caution with respect to any plans to change the test, allowing us to consider its structural appropriateness and equity issues. These findings have ramifications not only for the state of Ohio, but also for other potential high stakes testing programs because the procedures demonstrated here could be applied by states and other high stakes tests administrators in evaluating the structural validity of the measures they use prior to reporting and using the scores for accountability. It should be noted, however, that the term 'bias' as used in this study should be carefully interpreted since the invariance test of the structure demonstrated here does not include external criteria information in the analysis. The procedure employed here could suggest a first step in a broader process of an integrated validation procedure that should include not only internal checks on the population invariance of the underlying constructs but also checks on the population invariance of their relationship with external criteria.

Method

Sample

The study analyzed test scores of eighth graders who had their first opportunity to take the Ohio Ninth Grade Proficiency Test (ONGPT) in a large urban school district in Ohio. The ONGPT is administered twice a year, in spring and fall, to students in grade 8 through grade 12. Once a student (in any grade from 8 to 12) passes all five subjects in any time of the administrations, the student is not required to take the test again. The proportion of the eighth graders passing all five subjects varies across years but approximately 45% to 50% pass the test on their first try. The students selected for the study were part of the March 1999 administration cohort. For cross-validation purpose, the March 2002 test results were also gathered and analyzed. They were the eighth graders who took the parallel test in the same school district. The numbers of students whose subscale scores were analyzed were 713 (364 males and 349 females) in 1999 and 735 (389 males and 346 females) in 2002. The sample composition and descriptive statistics are presented in Table 1.

Instrument

The Ohio Ninth Grade Proficiency Test (ONGPT) is a five-part test that assesses students' knowledge and skills in writing, reading, math, citizenship, and science. The questions in the five test areas are based on learner outcomes which were adopted by the State Board of Education in 1988 (Ohio Department of Education, 2001). Each test subject is

Table 1

Table 1
Means and Standard Deviations of Test Scores for Total, Males, and Females

Year	1999						2002					
Test	Total ($n=713$)		Male ($n=364$)		Female($n=349$)		Total ($n=735$)		Male ($n=389$)		Female ($n=346$)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Writing	4.9	1.5	4.8	1.9	5.1	1.0	4.8	1.1	4.5	1.1	5.1	0.9
Reading	208.1	27.6	206.0	27.7	210.3	27.3	206.6	30.3	204.1	30.2	209.4	30.3
Math	191.1	19.9	193.1	20.4	189.2	19.2	189.6	33.0	188.8	37.0	190.4	27.8
Citizenship	201.4	23.1	203.4	24.5	199.3	21.5	201.9	25.0	202.0	25.8	201.8	24.2
Science	197.2	17.9	199.4	19.2	195.1	16.1	197.0	22.0	197.5	23.1	196.6	20.8

composed of certain number of subscales measured by multiple choice items except writing. Each item scores 1 and the maximum possible scores for reading, math, citizenship, and science are 40, 40, 50, and 40 respectively. The writing test score, ranging from 0 to 8, is the sum of the scores of two compositions on different writing themes. Each composition is worth a maximum of four points and is assessed by two readers using a holistic rubric. The scores from the two readers are later averaged, and the average scores are summed, which becomes a maximum of 8. Anyone who receives an 'unsatisfactory' on any of the four standards of the writing composition test or a total score of less than 5.0 is considered not to be proficient. The obtained raw scores on reading, mathematics, citizenship, and science are transformed into scaled scores through common item equating applying a Rasch IRT model to ensure that students of comparable proficiency are to have the comparable scaled scores regardless of minor fluctuations in difficulty across test administrations. Maximum and minimum scaled scores slightly vary across administrations but in spring 2002, for instance, reading ranged from 72 to 280, mathematics ranged from 101 to 279, citizenship ranged from 53 to 317, and science ranged from 97 to 282. The passing requirement is 200 for each subject. The unit of analysis of the study was the subscale scores which were continuous not dichotomous. The composition of the ONGPT is provided in Table 2.

Analysis

To determine the structural validity of the Ohio Ninth Grade Proficiency Test (ONGPT), a hypothetical factorial model of the test was examined employing a maximum likelihood confirmatory factor analysis (CFA). It has been widely acknowledged that CFA is the most prevalent technique performed to assess the factorial structure of certain scales (Kelloway, 1996). In addition, recent methodological advances in the application of CFA enable one to do rigorous tests of factorial invariance across identifiable population groups (Jöreskog, 1979; Jöreskog & Sörbom, 1986).

In this study, we examined our model based on the appropriateness of the parameter estimates and its fit measures provided by a CFA computer program, AMOS 4.0 (Arbuckle & Wothke, 1999, SPSS). The ONGPT CFA model is presented in Figure 1. The schematic model hypothesizes *a priori* that: (1) there are five factors explaining a response

Table 2
Composition of the ONGPT

Test Area	Possible Max. Raw Score	Subscale	Number of Item
Writing	8	Content	(2 compositions)
		Organization	
		Language	
		Writing Convention	
Reading	40	Fiction-constructs meaning	(10 items)
		Fiction-extends meaning	(6 items)
		Nonfiction-constructs meaning	(6 items)
		Nonfiction-extends meaning	(10 items)
		Everyday/fictional	(8 items)
Math	40	Measurement	(10 items)
		Arithmetic	(12 items)
		Geometry	(6 items)
		Data analysis	(6 items)
		Algebra	(6 items)
Citizenship	50	Geography	(6 items)
		Citizenship knowledge	(10 items)
		Government	(14 items)
		Economics	(5 items)
		Law	(8 items)
		History	(7 items)
Science	40	Life science	(10 items)
		Physical science	(10 items)
		Earth and space science	(10 items)
		Nature of science	(10 items)

to the ONGPT. The five factors are Writing, Reading, Math, Citizenship, and Science, and these are intercorrelated; (2) there are 24 manifest variables, which represent the subscales of each subjects of the ONGPT. Each subscale has a nonzero loading on the pertinent subject and a zero loading on the other four subjects; and (3) there are 24 errors of measurement, each associated with a manifest variable, which are modeled to be uncorrelated with one another. To ensure the structural fidelity of the five-factor ONGPT model not only to the total group but also to the gender groups the model was each tested against the total and the gender groups.

Testing for the invariance of the internal structure of the ONGPT across subgroups was accomplished by placing constraints on particular parameters (e.g., Werts, Rock, Linn, and Jöreskog, 1977). The statistical test was conducted by comparing χ^2 values obtained from several nested models, whose parameters were gradually constrained. Fit statistics

of the ONGPT five-factor model were first obtained from the baseline model representing the best fitting model simultaneously for the males and the females with no parameters constrained to be equal. It is noted that the baseline model should be differentiated from the model tested against the total group in the previous step. The baseline model takes the gender groups into account simultaneously, thus the χ^2 and the degrees of freedom of the model are the sum of the values from the independent male- and female-group tests due to the additive characteristic of the χ^2 distribution. The χ^2 value of the baseline model was examined with regard to its difference from that of the most restricted model. The null hypothesis of the most restricted model was that all of the matrices of factor loading, factor variance/covariance, and error variance are invariant across males and females. As the χ^2 difference between the baseline model and the most restricted model was statistically significant, a follow-up process was necessary to pinpoint the source of the non-invariance of the structure. The model parameters were constrained in a gradual manner, and each of the χ^2 values of the constrained models was respectively compared with the χ^2 value of the less restrictive comparative model. The hypotheses for the χ^2 tests were: (1) factor loadings are invariant across males and females, (2) factor loadings and factor variances are invariant across males and females, (3) factor loadings, factor variances, and factor covariances are invariant across males and females, and (4) factor loadings, factor variances, factor covariances, and error variances are invariant across males and females. The last hypothesis was first tested against the baseline model to see if the ONGPT five-factor model met the most restrictive assumptions, and it was rejected at the .01 alpha level. The issue of inflated alpha level due to multiple hypothesis tests by just constraining some parameters and testing them to the same data repeatedly was handled through Bonferroni adjustments (e.g., Bollen, 1989, p. 369). The critical probability level was set to .01 to maintain an overall alpha of .05.

Results

Structural validity

Evaluation of the fit of the Ohio Ninth Grade Proficiency Test (ONGPT) hypothetical model was first based on the adequacy of the parameter estimates and then the overall fit of the model to the samples. Tables 3 through 5 provide the parameter estimates of the total and the gender data obtained from two different test administrations. The sign, dimension, and direction of the estimates were equally adequate for all groups. The directions were positive and the magnitudes were statistically significant except for the error variance. The error variances were small which did not concern us because they were errors of measurement.

One of our primary interests was the extent to which the hypothesized model of the ONGPT fit, in other words, adequately described the sample data. Table 6 summarizes the fit measures of the model. For the interpretation of the RMSEA, Browne and Cudeck (1993) developed a guide-
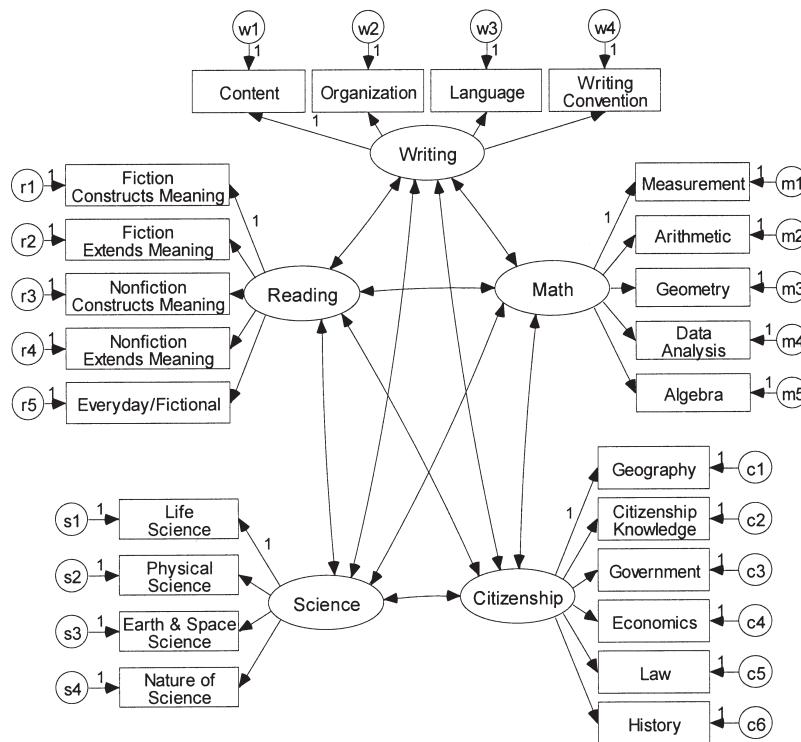
Note: Variables in small circles (w1 through s4) indicate measurement errors.

Figure 1. The Ohio Ninth Grade Proficiency Test Five-Factor Model

Table 3

Estimates and Standard Errors for the Ohio Ninth Grade Proficiency Test Five-Factor Model for Total Group from 1999 and 2002

1999

Factor loading matrix

	Writing	Reading	Math	Citizenship	Science
W1	1				
W2	.32 (.02)				
W3	1.02 (.01)				
W4	.46 (.03)				
R1		1			
R2		.58 (.03)			
R3		.51 (.02)			
R4		1.06 (.04)			
R5		.53 (.03)			
M1			1		
M2			1.41 (.06)		
M3			.53 (.03)		
M4			.52 (.03)		
M5			.67 (.03)		
C1				1	
C2				3.08 (.02)	
C3				4.44 (.03)	
C4				1.26 (.01)	
C5				1.89 (.01)	
C6				2.30 (.02)	
S1					1
S2					1.12 (.06)
S3					1.00 (.05)
S4					.92 (.05)

Factor variance/covariance matrix

	Writing	Reading	Math	Citizenship	Science
Writing	.24 (.01)				
Reading	.39 (.04)	3.30 (.26)			
Math	.35 (.04)	2.34 (.19)	3.41 (.29)		
Citizenship	.12 (.01)	.88 (.08)	.91 (.08)	.33 (.04)	
Science	.28 (.03)	2.19 (.17)	2.36 (.18)	.79 (.07)	2.21 (.20)

Error variances

w1	w2	w3*	w4	r1	r2	r3	r4	r5	m1	m2	m3
.01	.09	.00	.10	1.60	1.12	.74	1.88	.95	2.31	2.24	1.08

m4	m5	c1	c2	c3	c4	c5	c6	s1	s2	s3	s4
1.13	1.09	.74	1.96	3.15	1.14	1.50	1.19	1.84	2.13	1.78	1.71

2004

Factor loading matrix

	Writing	Reading	Math	Citizenship	Science
W1	1				
W2	.49 (.03)				
W3	1.01 (.01)				
W4	.51 (.03)				
R1		1			
R2		.66 (.03)			
R3		1.05 (.04)			
R4		.86 (.04)			
R5		.66 (.03)			
M1			1		
M2			1.36 (.08)		
M3			.52 (.04)		
M4			.55 (.04)		
M5			.67 (.05)		
C1				1	
C2				3.44 (.02)	
C3				3.94 (.03)	
C4				1.85 (.01)	
C5				2.22 (.02)	
C6				1.70 (.01)	
S1					1
S2					.87 (.04)
S3					.73 (.03)
S4					.80 (.03)

Factor variance/covariance matrix

	Writing	Reading	Math	Citizenship	Science
Writing	.24 (.01)				
Reading	.89 (.08)	2.52 (.19)			
Math	2.35 (.18)	3.16 (.26)	4.17 (.30)		
Citizenship	.33 (.03)	.40 (.04)	1.09 (.09)	.30 (.04)	
Science	.12 (.01)	.45 (.05)	2.70 (.19)	.70 (.06)	4.68 (.34)

Error variances

w1	w2	w3**	w4	r1	r2	r3	r4	r5	m1	m2	m3
.01	.12	.00	.20	1.23	.83	1.23	1.12	.99	1.75	2.28	1.16

m4	m5	c1	c2	c3	c4	c5	c6	s1	s2	s3	s4
1.01	1.05	.65	1.90	2.88	1.38	.87	1.02	1.95	2.47	1.86	1.52

Note: The actual values of w3* in 1999 and w3** in 2002 are .001 and .003, respectively. Factor loading and Factor variance/covariance estimates are all statistically significant at the .05 alpha level. Standard errors of estimation are presented in parenthesis.

Table 4

Estimates and Standard Errors for the Ohio Ninth Grade Proficiency Test Five-Factor Model for Gender from 1999

Male

Factor loading matrix

	Writing	Reading	Math	Citizenship	Science
W1	1				
W2	.35 (.04)				
W3	1.01 (.01)				
W4	.50 (.04)				
R1		1			
R2		.51 (.03)			
R3		.47 (.03)			
R4		.96 (.05)			
R5		.53 (.04)			
M1			1		
M2			1.49 (.08)		
M3			.55 (.04)		
M4			.51 (.04)		
M5			.70 (.04)		
C1				1	
C2				3.2 (.31)	
C3				4.6 (.43)	
C4				1.4 (.16)	
C5				1.9 (.20)	
C6				2.3 (.23)	
S1					1
S2					1.16 (.07)
S3					1.00 (.07)
S4					.93 (.06)

Factor variance/covariance matrix

	Writing	Reading	Math	Citizenship	Science
Writing	.24 (.02)				
Reading	.44 (.06)	4.04 (.42)			
Math	.34 (.06)	2.54 (.29)	3.38 (.39)		
Citizenship	.13 (.02)	.97 (.12)	.87 (.11)	.34 (.06)	
Science	.31 (.30)	5.57 (.27)	2.44 (.26)	.82 (.10)	2.49 (.30)

Error variances

w1	w2	w3*	w4	r1	r2	r3	r4	r5	m1	m2	m3
.01	.11	.00	.12	1.66	1.08	.66	2.11	1.06	2.08	1.92	1.13

m4	m5	c1	c2	c3	c4	c5	c6	s1	s2	s3	s4
1.09	1.11	.82	1.93	2.95	1.06	1.53	1.13	1.78	.198	1.76	1.56

Female

Factor loading matrix

	Writing	Reading	Math	Citizenship	Science
W1	1				
W2	.27 (.03)				
W3	1.01 (.01)				
W4	.39 (.03)				
R1		1			
R2		.67 (.05)			
R3		.58 (.04)			
R4		1.21 (.07)			
R5		.53 (.04)			
M1			1		
M2			1.36 (.08)		
M3			.52 (.04)		
M4			.55 (.04)		
M5			.67 (.05)		
C1				1	
C2				3.00 (.28)	
C3				4.28 (.39)	
C4				1.09 (.14)	
C5				1.97 (.20)	
C6				2.31 (.22)	
S1					1
S2					1.02 (.09)
S3					.96 (.08)
S4					.87 (.07)

Factor variance/covariance matrix

	Writing	Reading	Math	Citizenship	Science
Writing	.22 (.02)				
Reading	.32 (.05)	2.53 (.29)			
Math	.38 (.06)	2.10 (.24)	3.24 (.35)		
Citizenship	.12 (.02)	.78 (.10)	.90 (.11)	.32 (.05)	
Science	.28 (.04)	1.91 (.21)	2.20 (.25)	.74 (.09)	1.95 (.37)

Error variances

w1	w2	w3**	w4	r1	r2	r3	r4	r5	m1	m2	m3
.01	.07	.00	.08	1.19	1.15	.81	1.61	.81	2.26	2.56	1.02

m4	m5	c1	c2	c3	c4	c5	c6	s1	s2	s3	s4
1.15	1.05	.66	1.98	1.43	1.17	1.24	1.87	2.18	1.77	1.89	

Note: The actual values of w3* for male and w3** for female are .002 and .001, respectively. Factor loading and Factor variance/covariance estimates are all statistically significant at the .05 alpha level. Standard errors of estimation are presented in parenthesis.

Table 5
Estimates and Standard Errors for the Ohio Ninth Grade Proficiency Test Five-Factor Model for Gender from 2002

Male

Factor loading matrix

	Writing	Reading	Math	Citizenship	Science
W1	1				
W2	.52 (.04)				
W3	1.01 (.02)				
W4	.55 (.04)				
R1		1			
R2		.66 (.04)			
R3		1.03 (.06)			
R4		.85 (.05)			
R5		.66 (.04)			
M1			1		
M2			1.06 (.05)		
M3			.67 (.04)		
M4			.71 (.03)		
M5			.65 (.03)		
C1				1	
C2				3.51 (.29)	
C3				4.32 (.36)	
C4				1.86 (.17)	
C5				2.37 (.21)	
C6				1.75 (.17)	
S1					1
S2					.89 (.05)
S3					.70 (.04)
S4					.79 (.04)

Factor variance/covariance matrix

	Writing	Reading	Math	Citizenship	Science
Writing	.24 (.02)				
Reading	.86 (.10)	2.66 (.28)			
Math	2.39 (.25)	3.81 (.37)	4.38 (.43)		
Citizenship	.30 (.05)	.40 (.06)	1.08 (.12)	.29 (.05)	
Science	.11 (.02)	.43 (.07)	2.81 (.28)	.68 (.08)	5.14 (.51)

Error variances

w1	w2	w3*	w4	r1	r2	r3	r4	r5	m1	m2	m3
.01	.15	.00	.14	1.24	.89	1.33	1.11	1.16	1.17	2.27	1.18

m4	m5	c1	c2	c3	c4	c5	c6	s1	s2	s3	s4
.88	.96	.55	1.69	2.91	.88	1.36	1.10	2.06	2.39	1.88	1.52

Female

Factor loading matrix

	Writing	Reading	Math	Citizenship	Science
W1	1				
W2	.41 (.03)				
W3	1.00 (.01)				
W4	.44 (.04)				
R1		1			
R2		.67 (.04)			
R3		1.10 (.06)			
R4		.88 (.05)			
R5		.66 (.04)			
M1			1		
M2			.99 (.06)		
M3			.72 (.04)		
M4			.63 (.04)		
M5			.61 (.04)		
C1				1	
C2				3.30 (.32)	
C3				3.49 (.34)	
C4				1.81 (.18)	
C5				2.02 (.21)	
C6				1.62 (.17)	
S1					1
S2					.86 (.05)
S3					.76 (.05)
S4					.81 (.05)

Factor variance/covariance matrix

	Writing	Reading	Math	Citizenship	Science
Writing	.22 (.02)				
Reading	.95 (.12)	2.29 (.26)			
Math	2.29 (.25)	3.70 (.36)	3.94 (.42)		
Citizenship	.33 (.05)	.41 (.06)	1.11 (.14)	.33 (.06)	
Science	.13 (.02)	.48 (.06)	2.55 (.26)	.70 (.09)	4.17 (.44)

Error variances

w1	w2	w3**	w4	r1	r2	r3	r4	r5	m1	m2	m3
.00	.09	.00	.09	1.19	.76	1.10	1.12	.80	.74	2.17	1.11

m4	m5	c1	c2	c3	c4	c5	c6	s1	s2	s3	s4
1.15	1.17	.75	2.12	1.39	.94			1.80	2.41	1.82	1.51

Note: The actual values of w3* for male and w3** for female are .004 and .001, respectively. Factor loading and Factor variance/covariance estimates are all statistically significant at the .05 alpha level. Standard errors of estimation are presented in parenthesis.

line suggesting that a value less than 0.05 indicates *a close fit*, a value between 0.05 and 0.08 indicates *a mediocre fit*, and a value larger than 0.10 indicates *a poor fit*. The obtained RMSEA values were 0.045, 0.048, and 0.043, for the total, the male, and the female groups respectively in 1999. This was indicative of a *close* fit of the model to the data. Also, the *incremental fit measures* (Bollen, 1989, pp. 269-276) suggested to be larger than 0.95 for a good fit (see Hu & Bentler, 1999) were consistent in showing a good fit of the model to the total as well as to the gender groups (TLI: 0.951 to 0.972; IFI: 0.957 to 0.976). The AGFI values of 0.886 through 0.915 indicated a decent fit, yet they were relatively small comparing to the other fit measures, which we suspected was due to the relatively small sample size (n = 713) (see Anderson & Gerbing's simulation study, 1984). Generally, an AGFI value close to or larger than 0.90 is considered to be a good fit. A replication of the model supported the good fit to an independent sample. The RMSEA values from the 2002 data were respectively 0.046, 0.056, and 0.042 for the total and the gender groups. The other fit measures including the AGFI (0.866 to 0.913), the TLI (0.951 to 0.972), and the IFI (0.957 to 0.976) were good as well.

Model Invariance

As evidence was found supporting the structural validity of the test scores, a necessary follow-up procedure was to validate its structural invariance across males and females. The unstandardized parameter estimates of factor loadings, factor variance/covariances, and error variances in Tables 3 through 5 the relationships among variables. The reason for comparing unstandardized estimates over standardized estimates was that standardized estimates are a function of standard deviations that are sample-dependent. The tables show that the directions of the parameter estimates are equal and the magnitudes follow the same pattern for males and females. For instance, among the four writing subscales, W3 (Language) was the strongest loading on the Writing factor equally for males and females. Likewise, R3 (Nonfiction-constructs meaning), M2 (Arithmetic), C2 (Citizenship knowledge), and S2 (Physical science) were the most heavily loaded manifest variables on the relevant factors across the groups.

In Table 7, the χ^2 value of 1,415.92 (542, n = 713) from the most restricted model, whose factor loadings, variances,

covariances, and error variances was all set to be the same for males and females, was compared with the χ^2 value of 838.95 (484, n = 713) of the baseline model. The comparison yielded χ^2 difference of 576.97 (58, n = 713), which was statistically significant at the .01 alpha level. The comparison for the 2002 data resulted in a significant χ^2 value (563.67 (58, n = 735), p < .01) as well. This indicated that some equality constraints of the ONGPT model did not hold across the male and the female eighth graders who took the same test. Therefore, the individual parameters of factor loadings, factor variances, factor covariances, and error variances of the ONGPT model needed to be constrained step by step and to be examined to pinpoint the source of the invariance.

Comparisons of the nested models in Table 7 revealed that factor loadings, factor variances, and factor covariances functioned equally for the gender groups from year 1999. The χ^2 difference between the factor loading-constrained model and that of the baseline model was 34.0 (19, n = 713, p > .01). The same comparisons for factor variances ($\Delta\chi^2$(5, n = 713) = 9.26, p > .01) and factor covariances ($\Delta\chi^2$(10, n = 713) = 18.46, p > .01) showed no statistical deviation from the respective comparative models. The χ^2 (1,415 (542, n = 713)) of the error variance-constrained model, however, was distinctly larger than that of the comparative model (900.66 (518, n = 713)) and the difference ($\Delta\chi^2$ (24, n = 713) = 515.26, p < .01) was statistically significant. Analyses supported the structural invariance of the ONGPT five-factor model across the male and the female groups sampled in 1999 except the error variances. The relationship between Writing ability and writing subscale scores and between Reading ability and reading subscale scores, could be assumed the same for males and females. In addition, the variances of the five test subjects and the reciprocal relationships of those were equal for the gender groups. However, the amount of disturbances that disrupt the relation between the factors and their observed variables was not equivalent. The replicated comparisons with the data from 2002 revealed that factor loadings ($\Delta\chi^2$ (19, n = 736) = 23.47, p > .01) and factor variances ($\Delta\chi^2$ (5, n = 736) = 3.24, p > .01) were invariant. As found earlier, the test for invariant measurement errors deviated significantly from the comparative model ($\Delta\chi^2$ (24, n = 736) = 1,487.93, p < .01). Also, factor covariances were not confirmed to be invariant ($\Delta\chi^2$ (10, n = 736)

Table 6

Goodness-of-fit Statistics for the Ohio Ninth Grade Proficiency Test Five-Factor Model: Independent Fits for Total, Males and Females

Group	n	Year	χ^2	df	RMSEA	CI	AGFI	TLI	IFI
Total	713	1999	584.91	242	0.045	0.040; 0.049	0.915	0.966	0.970
Male	364		442.22	242	0.048	0.041; 0.055	0.886	0.964	0.969
Female	349		396.73	242	0.043	0.035; 0.050	0.894	0.966	0.970
Total	735	2002	622.91	242	0.046	0.042; 0.051	0.913	0.966	0.970
Male	389		533.79	242	0.056	0.049; 0.062	0.866	0.951	0.957
Female	346		390.48	242	0.042	0.034; 0.050	0.889	0.972	0.976

Note: RMSEA = root mean squared error of approximation (Steiger & Lind, 1980); CI = confidence interval at 95%; AGFI = adjusted goodness-of-fit index (Jöreskog & Sörbom, 1993); TLI = Tucker-Lewis index (Tucker & Lewis, 1973); and IFI = incremental fit index (Bollen, 1989a). Test for RMSEA is deviation from .05, representing *close fit* (Browne & Cudeck, 1993).

Table 7
Tests for the Parameter Invariance across Gender for the Ohio Ninth Grade Proficiency Test Five-Factor Model

Model Description	Comparative Models	Year	n	χ^2	df	$\Delta\chi^2$	Δdf
Baseline Model (B)	—	1999	713	838.95	484	—	—
All Parameters (FL+FV+FC+EV)	B			1,415.92	542	*576.97	58
Factor Loadings (FL)	B			872.95	503	34.00	19
FL + Factor Variances (FV)	FL			882.20	508	9.26	5
FL+FV+Factor Covarainces (FC)	FL+FV			900.66	518	18.46	10
FL+FV+FC+Error Variances (EV)	FL+FV+FC			1,415.92	542	*515.26	24
Baseline Model (B)	—	2002	736	924.26	484	—	—
All Parameters (FL+FV+FC+EV)	B			1,487.93	542	*563.67	58
Factor Loadings (FL)	B			947.73	503	23.47	19
FL + Factor Variances (FV)	FL			950.96	508	3.24	5
FL+FV+Factor Covarainces (FC)	FL+FV			984.62	518	*33.65	10
FL+FV+FC+Error Variances (EV)	FL+FV+FC			1,487.93	542	*503.31	24

Note: * $p < .01$. $\Delta\chi^2$ = difference in χ^2. Δdf = difference in degrees of freedom.

$= 33.65$, $p < .01$) this time. The invariance of the internal structure of the ONGPT across gender was not fully ensured based on the findings of non-invariant measurement errors both in 1999 and 2002 and factor covariances in 2002.

Conclusion

The findings supported the structural fidelity of the Ohio Ninth Grade Proficiency Test (ONGPT). The internal relationships among the five test subjects with the pertinent subscales confirmed the way that the test was designed. However, the test did not show enough evidence of its structural invariance across the gender samples analyzed in the study. While factor loadings and factor variances displayed the evidence of invariance, factor covariances and measurement errors did not hold the equality for the gender groups. Even if it is known that equal measurement errors are hard to achieve most of the time (see Byrne, 1987, 2001; Werts, et al., 1977) and even if the equality of factor loading and factor variance/covariance matrices were our higher priorities, ensuring the equality of the reliability of this high school graduation test was essential considering its high stakes role in various educational applications. Given that the interpretation of test results focuses on groups as well as individuals, we caution the uniform use and interpretation of the test scores for all test takers regardless of their group characteristics. In addition, we state that it can be risky comparing averages across groups without further investigation into the source of inequality of the factor covariances and the reliability.

The results from this study may hold for a large urban school district in Ohio. In order to evaluate the generalizability of these results, we cross-validated the hypothesized model using an independent sample from a different test administration year. Based on the inconsistent findings from the two years, it will be interesting to see how the findings will conform/disconform with another sample from the same school district and it will also be valuable to know if the findings will result in the same in suburban and rural areas. Even though

we anticipate that the results of such a test would support our findings, these remain to be empirically demonstrated.

An examination of correlation patterns and a test of the invariance of the correlation patterns can enhance the proper interpretation of the test scores. However, studies like this can benefit from examinations of individual items. Techniques such as differential item functioning enable identification of how specific items function differently for different groups. In addition, examination of the external factors affecting the test scores including instruction, curriculum, student characteristics, and so forth, would improve the usefulness of the outcomes of this study.

This study implemented a sophisticated and necessary statistical approach to structural validation for a high stakes measure of proficiency in a large urban school district in Ohio. This procedure could be used more broadly, e.g., for other districts in the state, for the state as a whole, or for other entities where high stakes assessments are administered and interpreted. Given the high stakes use of the information resulting from this situation in Ohio, it is important that those responsible for the use and interpretation of the results have adequate evidence of structural validity. Otherwise, the results of such a test are open to substantial possible misinterpretation and misuse, and the inappropriate and/or misleading interpretation may not be readily identified. Use of the results across groups in a uniform way may not be warranted. The procedures followed in this study imply coherent attention and usability. These procedures, however, are a necessary but not sufficient condition for test validity, i.e., they are one of the critical aspects of the process as specified by the standards of quality test development, use, and interpretation (AERA, APA, & NCME, 1999).

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for Educational*

and Psychological Testing. Washington, DC: American Educational Research Association.

Anderson, J., & Gerbing, D. W. (1984). The effects of sampling error on convergence, improper solutions and goodness-of-fit indices for maximum likelihood confirmatory factor analyses. *Psychometrika*, *49*, 155-173.

Arbuckle, J. L., & Wothke, W. (1999). *AMOS 4.0 user's guide*. Chicago: Smallwaters.

Basturk, R. (2001). *The relationship between selected teacher and school characteristics and Ohio teacher attitudes towards standardized tests and the Ohio Proficiency Tests*. Ph.D. Dissertation, Ohio State University.

Behrman. E. H. (1998). *A critical review of New Jersey's High School Proficiency Test in reading*. (ERIC Document Reproduction Service No. ED 417 219).

Bollen, K. A. (1989a). A new incremental fit index for general structural equation models. *Sociological Methods & Research*, *17*, 303-316.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.

Byrne, B. M., & Shavelson, R. J. (1987). Adolescent self-concept: testing the assumption of equivalent structure across gender. *American Educational Research Journal*, *24*, 365-385.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chandler, D. G., & Brosnan, P. A. (1995). A comparison between mathematics textbook content and a statewide proficiency test. *School Science and Mathematics*, *95*, 118-123.

DeMars, C. E. (1998). Sex differences in mathematics and science on a high school proficiency exam: The role of response format. *Applied Measurement in Education*, *11*, 279-299.

Gallagher, M. P. (1993). *Proficiency testing and poverty: looking within a large urban district*. Paper presented at the Annual Meeting of the American Educational Research Association, Atlanta, GA. (ERIC Document Reproduction Service No. ED 368 126).

Gibson, S. D. (1997). A comparison study of previous achievement indices for two groups of ninth-grade students: Those who passed and those who failed sections of the Ohio ninth grade proficiency test. *American Secondary Education*, *26*, 2-6.

Green, B. F. (1995). The comparability of scores from performance assessment. *Educational Measurement: Issues and Practice*, *14*, 13-15.

Guion, R. M. (1966). Employment tests and discriminatory hiring. *Industrial Relations*, *5*, 20-37.

Hambleton, R. K., Jones, R. W. (1994). Comparison of empirical and judgmental procedures for detecting differential item functioning. *Educational Research Quarterly*, *18*, 21-36.

Heiney, C. J. (1998). *Effect of problem of the day on Ohio ninth grade proficiency test mathematics scores*. Masters Thesis, Salem-Teikyo University.

Hu, L. –T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*, 1-55.

Hull, M., & Tache, D. (1993). *Are Iowa of basic skills stanines predictors of success on Ohio's ninth grade proficiency test?* (ERIC Document Reproduction Service No. ED 361 367).

Jöreskog, K. G. (1979). Simultaneous factor analysis in several populations. In K. G. Jöreskog & D. Sörbom (Eds.), *Advances in factor analysis and structural equation models* (pp. 189-206). Cambridge, MA: Abt.

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of Linear Structural Relationships by Maximum Likelihood and Least Square Methods*. Mooresville, IN: Scientific Software.

Jöreskog, K. G., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Chicago: Scientific Software International.

Kelloway, E. K. (1998). *Using LISREL for structural equation modeling: A researcher's guide*. Thousand Oaks, CA: Sage Publications.

Kunicki, J. A., & Osborne, A. (1992). The Ohio ninth grade proficiency test and a school's curriculum. *Ohio Journal of School Mathematics*, *23*, 1-8.

Lane, S., & Stone, C. A. (2002). Strategies for examining the consequences of assessment and accountability programs. *Educational Measurement*, *21*, 23-30.

Lanese, J. F. (1992). *Statewide proficiency testing: Establishing standards or barriers?* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. (ERIC Document Reproduction Service No. ED 347 196).

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635-694.

Messick, S. (1988). The once and future issues of validity: Assessing the meaning and consequences of measurement. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 33-44). Hillsdale, NJ: Lawrence Erlbaum Associates.

Messick, S. (1993). Validity. In R. L. Linn. (Ed.), *Educational measurement* (p. 81). Phoenix, AZ: The Oryx Press.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessment. *Educational Researcher*, *23*, 13-23.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*, 5-8.

Ninth-Grade Tests. Retrieved June 8, 2001, from http://www.ode.oh.us/proficiency/ninth.asp

Noel, A. W. (1994). *Predictors of success on Ohio's ninth grade proficiency test.* (ERIC Document Reproduction Service No. ED 392 803).

Powers, S., & Jones, P. B. (1992). Factorial invariance of the California achievement tests across race and sex. *Educational and Psychological Measurement, 44*, 967-970.

Raivetz, M. J. (1992). Can school district survive the politics of state testing initiatives? *NASSP Bulletin, 76*, 57-65.

Rock, D. A., Werts, C. E., & Flaugher, R. L. (1978). The use of analysis of covariance structures for comparing the psychometric properties of multiple variables across populations. *Multivariate Behavioral Research, 13*, 403-418.

Shafer, G. (1997). School reform and the high school proficiency test. *English Journal, 86*, 39-41.

Smisko, A., Denny, P., & Twing, J. S. (2000). The Texas model for content and curricular validity. *Applied Measurement in Education, 13*, 333-342.

Steiger, J. H., & Lind, J. C. (1980, June). *Statistically biased tests for the number of common factors.* Paper presented at the Psychometric Society Annual Meeting, Iowa City, IA.

Thorndike, R. L. (1971). Concepts of culture-fairness. *Journal of Educational Measurement, 8*, 63-70.

Tucker, L. R., & Lewis, C. (1973). The reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*, 1-10.

Werts, C. E., Rock, D. A., Linn, R. L., Jöreskog, K. G. (1977). Validating psychometric assumptions within and between several populations. *Educational and Psychological Measurement, 37*, 863-872.

Wiesenauer, L. K. (2003). *Ohio school district performance: An analysis of demographic factors impacting proficiency test results.* Ph.D. Dissertation, Ohio State University.

The Mid-Western Educational Research Association

A gift membership has been given to you, _____

by_____

Your name is now included as a member of 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. Members pay reduced registration fees for the annual meeting held in October. The conference attracts many nationally-recognized leaders in educational research. Enjoy your membership!

Gift Membership Gift Membership Gift Membership Gift Membership

---

Gift Membership Gift Membership Gift Membership Gift 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. Fill out the bottom portion and mail it with your check to: Jean Pierce – LEPF Dept. – Northern Illinois U – DeKalb, IL 60115

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

> **Make your check payable to MWERA.**
>
> **Professional Membership—Enclose a check for $45**
>
> **Student Membership—Enclose a check for $25**

T · H · E
OHIO
STATE
UNIVERSITY

The Ohio State University
2500 Kenny Road
Columbus, Ohio 43210