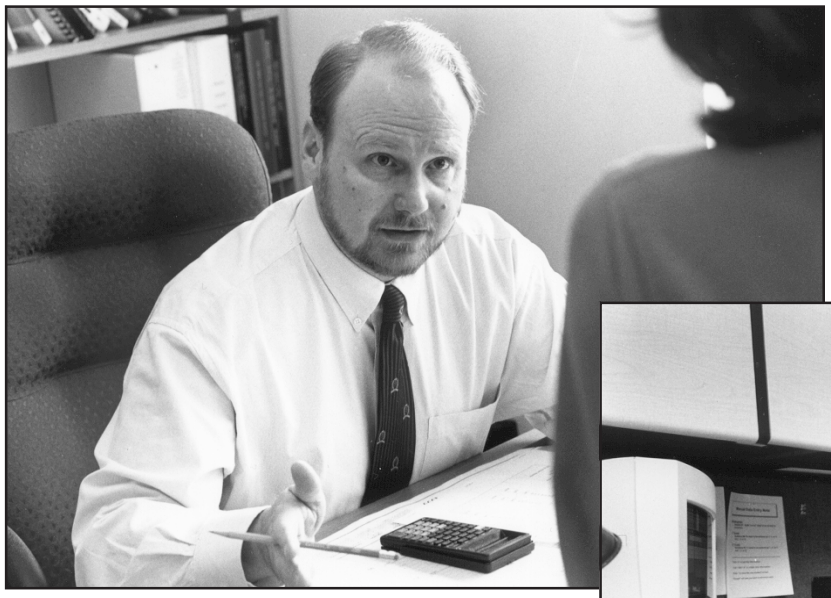
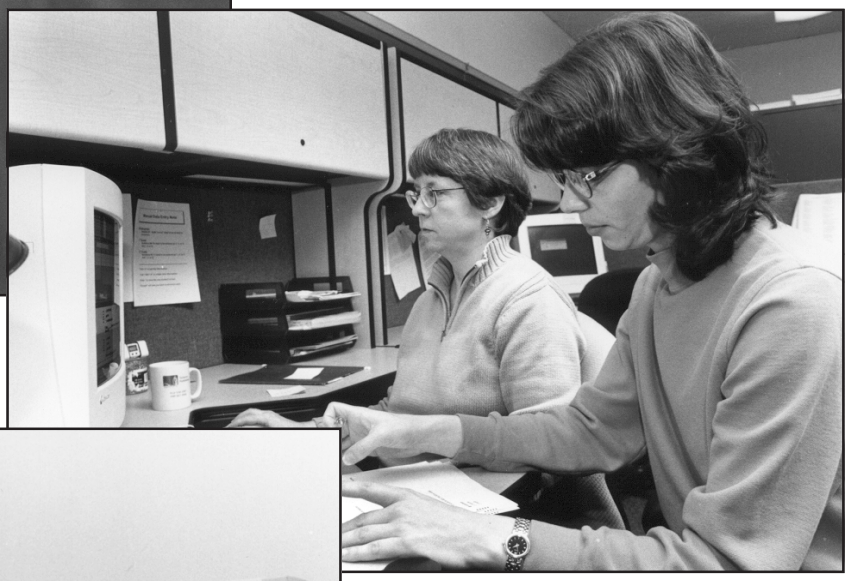
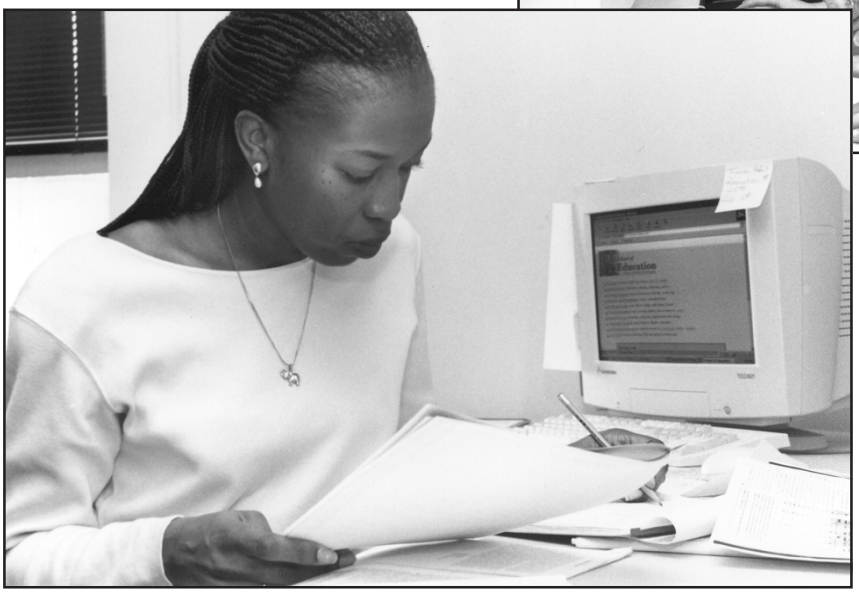
Volume 12, No. 2 Spring 1999

MID-WESTERN EDUCATIONAL RESEARCHER

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







Indiana Center for Evaluation
Indiana University

On the Cover

Evaluation is increasingly expected of agencies within the educational and human service sectors. However, few of these agencies possess the experience or resources necessary to conduct efficient and meaningful evaluations of their programs.

To address this need, the Indiana Center for Evaluation was established in the fall of 1996 as a collaborative venture between the Indiana University School of Education and Junior Achievement of Central Indiana. Under the direction of Dr. Kim Metcalf, the Center's primary mission is to promote and support systematic program evaluation, particularly in the educational and nonprofit sectors. Drawing upon the expertise of faculty and staff from a variety of disciplines, the Center seeks to assist those organizations moving toward organized investigation of their programs to enhance quality. As a result, a focus of the Center's work is on helping its clients develop, implement, and maintain on-going, structured program evaluation for continuous improvement.

The Center is an independent agency, devoted to the application of multiple research methods to the practical evaluation of services and programs. Its clients have reflected a variety of private, public, and governmental agencies, primarily in education and social services.

Among the range of the Indiana Center for Evaluation's current projects are a 15-year longitudinal evaluation of more than 1.5 million participants in Junior Achievement programs in central Indiana; a multi-year evaluation of the impact of the controversial Cleveland Voucher Program for the Ohio Department of Education; and a broad-scale evaluation of the Anchor Schools program for migrant students in Florida, Georgia, Tennessee, and South Carolina for the U.S. Department of Education.

Pictured on the cover (from top): Kim Metcalf, Director; Rebecca Gross, Administrative Assistant to the Director and Amy Craig, Office Support Staff; JaDora Sailes, Junior Achievement Post-Doctoral Fellow.

Information for Contributors to the Mid-Western Educational Researcher

The *Mid-Western Educational Researcher* accepts research-based manuscripts that would appeal to a wide range of readers. All materials submitted for publication must conform to the language, style, and format of the *Publication Manual of the American Psychological Association*, 4th ed., 1994 (available from Order Department, American Psychological Association, P.O. Box 2710, Hyattsville, MD 20784).

Four copies of the manuscript should be submitted typed double-spaced (including quotations and references) on $8^1/_2$ x 11 paper. Only words to be italicized should be underlined. Abbreviations and acronyms should be spelled out when first mentioned. Pages should be numbered consecutively, beginning with the page after the title page. Manuscripts should be less than 20 pages long. An abstract of less than 100 words should accompany the manuscript.

The manuscript will receive blind review from at least two professionals with expertise in the area of the manuscript. The author's name, affiliation, mailing address, telephone number, e-mail address (if available), should appear on the title page only. Efforts will be made to keep the review process to less than four months. The editors reserve the right to make minor changes in order to produce a concise and clear article. The authors will be consulted if any major changes are necessary.

Manuscripts should be sent with a cover letter to:

Deborah L. Bainer, *MWER* Co-Editor
1680 University Drive, Ohio State University at Mansfield, Mansfield, OH 44906

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

MID-WESTERN EDUCATIONAL RESEARCHER

Volume 12, Number 2 Spring 1999 ISSN 1056-3997

MWER Publication Address

Deborah L. Bainer
The Ohio State University, Mansfield
1680 University Drive
Mansfield, OH 44906
Phone: (419) 755-4287
Fax: (419) 755-4367
email: bainer.1@osu.edu

MWERA Membership Information

Jean W. Pierce
Dept. EPCSE
Northern Illinois University
DeKalb, IL 60115
Phone: (815) 753-8470
Fax: (815) 753-9250
email: P30JWP1@mvs.cso.niu.edu

The Use of Tests of Statistical Significance

Thomas R. Knapp
Ohio State University

Abstract

This article summarizes the author's views regarding the appropriate use of significance tests, especially in the context of regression analysis, which is the most commonly-encountered statistical technique in education and related disciplines. The article also includes a brief discussion of the use of power analysis after a study has been carried out.

Although statistical significance tests have come under repeated attacks for several years, most recently in psychology by Jacob Cohen (1994), Frank Schmidt (1996), and others, there are times when they should be used and there are times when they should not be used. What follows is an attempt to identify those times as far as educational research is concerned.[1]

A brief history of the controversy, 1970-1998

In 1970 there appeared a book edited by sociologists Denton Morrison and Ramon Henkel, entitled *The Significance Test Controversy*. That book consisted of chapters written by people on both sides of the issue, but most of the authors were "con", i.e., they had little or nothing good to say about significance tests. Several of those chapters had originally appeared elsewhere in books or as journal articles, and some of the comments were downright nasty. In his chapter, for example, Paul Meehl characterized the researcher who uses significance tests as "a potent but sterile intellectual rake who leaves in his merry path a long train of ravished maidens but no viable scientific offspring" (Meehl, 1970, p. 265).

For the next couple of decades things were relatively quiet, except for the occasional raising of a few new voices (e.g., Carver, 1978). Significance tests continued to be used by researchers who felt they were warranted and continued to be eschewed by researchers who felt they were not. Then in the 90s, prompted by articles written by Cohen (1990, 1994) and Schmidt (1992, 1996), the controversy was rekindled. It led to the creation of a task force in psychology to deal with the matter and to the publication in 1997 of another entire book devoted to the "pros" and "cons" of significance testing, edited by Lisa Harlow, Stanley Mulaik, and James Steiger, entitled *What If There Were No Significance Tests?* (See Levin, 1998 and Thompson, 1998 for reviews of, and reactions to, that book.) Schmidt had advocated the discontinuation of **all** significance tests in favor of confidence intervals around obtained effect sizes, and the discontinuation of **all** narrative literature reviews in favor of meta-analyses for pooling results across studies. At the time of the writing of this article—Autumn, 1998—the APA Task Force had not issued its final report, but its interim report in 1997 suggested that Schmidt's extreme positions would not be supported.

The situation in educational research has closely paralleled the recent developments in psychology. Starting in 1993 with an entire issue of the *Journal of Experimental Education* devoted to the topic of significance testing (again, "pros" and "cons", but mostly "cons"—see esp. Carver, 1993 and Thompson, 1993), there appeared subsequent articles by Thompson (1996), Robinson and Levin (1997), and others, culminating in a debate on the topic at the April, 1998 annual meeting of the American Educational Research Association in San Diego.

The position taken here

This writer takes a very simple approach to the controversy. If there is a hypothesis to be tested and if a statistical inference is warranted (for a probability sample drawn from a well-defined population), then significance testing should be used. (The terms "hypothesis testing" and "significance testing" are regarded as interchangeable, but see Huberty, 1993 concerning the distinctions that are sometimes made between the two.) If there is no hypothesis to be tested but a statistical inference is warranted, then interval estimation (constructing a confidence interval around a point estimate) should be employed. If a statistical inference is not warranted (when the obtained data are for a full population or for a non-probability sample), whether or not there is a hypothesis to be tested, descriptive statistics should suffice.

One can often get hypothesis testing "for free" by using interval estimation (if the hypothesized parameter is not in the confidence interval, reject it), but there are situations where that is not the case (see Dixon and Massey, 1983, p. 93). When dealing with percentages, differences between percentages, or ratios of percentages, for example, the standard errors for the hypothesis-testing approach and the interval-estimation approach may differ considerably (see Knapp and Tam, 1997). For odds ratios associated with 2x2 contingency tables the significance test is straightforward, whereas the determination of the corresponding confidence interval is extremely complicated (see Fleiss, 1981, pp. 71-75).

Regression analysis

It is indeed curious that the adversaries in the significance testing controversy rarely use examples involving regression analysis (Steiger and Fouladi, 1997 is a notable

exception), which is the statistical technique that is most commonly used in the behavioral sciences.[2] There are many textbooks (e.g., Cohen and Cohen, 1983; Darlington, 1990; Marascuilo and Levin, 1983; Pedhazur, 1997; Stevens, 1996) and monographs (e.g., Achen, 1982; Berry, 1993; Berry and Feldman, 1985; Breen, 1996; Fox, 1991; Hardy, 1993; Iversen, 1991; Jaccard, Turrisi, and Wan, 1990; Jaccard and Wan, 1996; Langbein and Lichtman, 1978; Lewis-Beck, 1980; Newbold and Bos, 1985; Schroeder, Sjoquist, and Stephan, 1986)[3] that treat regression analysis. Hypothesis testing is given much greater emphasis than interval estimation in those sources. Most never even mention confidence intervals or devote very little space to their use (despite the fact that such intervals are routinely provided in the output of certain computer programs), suggesting that significance testing is the preferred approach. Of all of these authors, the only one who provides any sort of extended discussion of the advantages and disadvantages of confidence intervals vs. significance tests is Achen (1982), and he doesn't take a stand on one approach in preference to the other. Most users of regression analysis apparently are content with testing hypotheses concerning correlation coefficients (simple and multiple), regression coefficients (standardized or unstandardized), intercepts, and the like.

Some comments regarding observed power

There has recently been a disturbing tendency (disturbing to this writer and to a few others—see, for example, Goodman and Berlin, 1994, and Zumbo and Hubley, 1998) in some textbooks, journal articles, and computer programs to report the "observed power" for a study (see, for example, Munro, 1997 and the output for some of the analysis of variance programs in SPSS). Power is, or at least should be, an a priori concept. Researchers know (or should know), GOING INTO a study, the probability of getting a statistically significant finding (given the alternatively hypothesized effect size, the specified alpha level, and the sample size), i.e., the probability of rejecting a false null hypothesis in favor of a true alternative hypothesis. What some people are arguing for these days is the calculation of the obtained effect size (that's fine) and the determination of the corresponding "observed power" (that's not), COMING OUT OF a study. The rationale goes something like this: I'm willing to take the obtained **sample** effect size as a good estimate of the **population** effect size, see what power I had for that effect size for the sample size I drew, and determine what sample size I would need in my next study in order to have the power I want. That sort of reasoning seems terribly convoluted and an inappropriate use of power analysis as an aspect of statistical inference. Those who are interested in a counter-argument regarding the concept of "observed power" are urged to read the articles by Falk, Hogan, Muller, and Jennette (1992) and by Taylor and Muller (1995) and come to their own conclusions about the defensibility of that concept. The first of those articles is a substantive article concerning an experiment involving a fixed sample size (a priori power was not involved in its determination) of 26 people randomly assigned to two treatments, for which the research hypothesis is null, i.e., the theoretical position is that there is no treatment effect. (They found none and the study was terminated before the originally anticipated date.) The second article is a methodological article that advocates the calculation of obtained power for the Falk experiment for varying effect sizes close to null, and the construction of one-sided confidence intervals around those powers AND one-sided confidence intervals for the associated sample sizes.

Steiger and Fouladi

In defending their preference for interval estimation in multiple regression analysis (they also advocate the reporting of observed power), Steiger and Fouladi (1997) give the example of a confidence interval for the squared multiple correlation coefficient. The obtained R^2 in a sample of 45 observations on six variables (five independent and one dependent) was .40, which was statistically significant at the .001 level; the limits of the 95% confidence interval for the population R^2 were .095 and .562. They claim that the inference provided by the interval estimate is much more informative, albeit less impressive, than the inference provided by the significance test. That may be, but the price that was paid to get it (computationally complex calculations that are not included in standard statistical packages—but are available from Steiger and Fouladi) may not be worth it. This writer personally prefers the significance test, for a **given** null hypothesis, a **given** alternative hypothesis, a pre-specified alpha, and a sample size that is appropriate for a **given** desired power. Cohen's well-known and readily-available power book (Cohen, 1988) contains all of the necessary formulas and tables. There are also several readily-available software packages for carrying out such analyses.

Conclusion

This article has tried to summarize when significance tests (hypothesis tests) should be used and when they should not. Traditional regression analysis is one of the contexts in which tests of statistical significance appear to be most defensible and for which the corresponding interval estimation procedures are either not appropriate or are unnecessarily complicated.

It could be that many educational researchers are "closet Bayesians". They would like to be able to determine the probability that the null hypothesis is true, given the data, but in classical statistical inference that is not possible, so they must settle for the probability of getting the data (or something even more extreme), given that the null hypothesis is true (see Cohen, 1994). That's when they get frustrated and are prone to making all sorts of mistakes in interpreting significance tests. But the cure for this is not the abandonment of significance tests; the cure is to use them properly and interpret them properly OR to come out of the closet and become a Bayesian (see Pruzek, 1997 and Berger, Boukai, and Wang, 1997 regarding those alternatives).

Footnotes

[1] It might be argued that educational research is just like psychological research, sociological research, or research in any of the other social sciences, but many years ago Gowin (1972) claimed that it is (or at least should be) distinctive. Education is primarily interventionist. Our society doesn't have to develop various curricula, pay some teachers more than others, etc., but it has chosen to do so. It is therefore appropriate that controlled experiments and large correlational studies be carried out in order to determine to what extent such things "work".

[2] In their summary of statistical techniques used in reports of studies published recently in the *American Educational Research Journal*, the *Educational Researcher*, and the *Review of Educational Research*, Elmore and Woehlke (1998) indicated that multiple regression analysis was used in 148 out of 1906 articles (7.8 %), but if you add to that the 99 articles that used bivariate correlation, the 70 articles that used a t test, the 221 articles that used the analysis of variance or covariance (all of which can be subsumed under regression analysis—see, for example, Cohen, 1968) the total is 538 out of 1906 (28.2 %).

[3] These monographs were all categorized under the "Regression" grouping in a recent Sage University Paper.

References

Achen, C.H. (1982). *Interpreting and using regression*. Beverly Hills, CA: Sage.

Berger, J.O., Boukai, B., and Wang, Y. (1997). Unified frequentist and Bayesian testing of a precise hypothesis. *Statistical Science*, *12*, 133-148.

Berry, W.D. (1993). *Understanding regression assumptions*. Newbury Park, CA: Sage.

Berry, W.D., and Feldman, S. (1985). *Multiple regression in practice*. Beverly Hills, CA: Sage.

Breen, R. (1996). *Regression models*. Thousand Oaks, CA: Sage.

Carver, R.P. (1978). The case against statistical significance testing. *Harvard Educational Review*, *48*, 378-399.

Carver, R.P. (1993). The case against statistical significance testing, revisited. *Journal of Experimental Education*, *61*, 287-292.

Cohen, J. (1968). Multiple regression as a general data-analytic system. *Psychological Bulletin*, *70*, 426-443.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd. ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1990). Things I have learned (so far). *American Psychologist*, *45*, 1304-1312.

Cohen, J. (1994). The earth is round ($p<.05$). *American Psychologist*, *49*, 997-1003.

Cohen, J., and Cohen, P. (1983). *Applied regression/correlation analysis for the behavioral sciences* (2nd. ed.). Hillsdale, NJ: Erlbaum.

Darlington, R.B. (1990). *Regression and linear models*. New York: McGraw-Hill.

Dixon, W.J., and Massey, F.J. (1983). *Introduction to statistical analysis* (4th ed.). New York: McGraw-Hill.

Elmore, P.B., and Woehlke, P.L. (April, 1998). Twenty years of research methods employed in *American Educational Research Journal*, *Educational Researcher*, and *Review of Educational Research*. Paper presented at the Annual Meeting of the American Educational Research Association, San Diego.

Falk, R.J., Hogan, S.L., Muller, K.E., and Jennette, J.C. (1992). Treatment of progressive membranous glomerulopathy. *Annals of Internal Medicine*, *116*, 438-445.

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

Fox, J. (1991). *Regression diagnostics*. Newbury Park, CA: Sage.

Goodman, S.N., and Berlin, J.A. (1994). The use of predicted confidence intervals when planning experiments and the misuse of power when interpreting results. *Annals of Internal Medicine*, *121*, 200-206.

Gowin, D.B. (1972). Is educational research distinctive? In L.G. Thomas (Ed.), Philosophical redirection of educational research (Chapter I, pp. 9-25). The Seventy-first Yearbook of the National Society for the Study of Education. Chicago: The University of Chicago Press.

Hardy, M.A. (1993). *Regression with dummy variables*. Newbury Park, CA: Sage.

Harlow, L.L., Mulaik, S.A., and Steiger, J.H. (Eds.) (1997). *What if there were no significance tests*? Mahwah, NJ: Erlbaum.

Huberty, C.J. (1993). Historical origins of statistical testing practices: The treatment of Fisher versus Neyman-Pearson views in textbooks. *Journal of Experimental Education*, *61*, 317-333.

Iversen, G.R. (1991). *Contextual analysis*. Newbury Park, CA: Sage.

Jaccard, J., and Wan, C.K. (1996). *LISREL approaches to interaction effects in multiple regression*. Thousand Oaks, CA: Sage.

Knapp, T.R., and Tam, H.P. (1997). Some cautions concerning inferences about proportions, differences between proportions, and quotients of proportions. *Mid-Western Educational Researcher*, *10*, 11-13.

Langbein, L.I., and Lichtman, A.J. (1978). *Ecological inference*. Beverly Hills, CA: Sage.

Levin, J.R. (1998). To test or not to test H_0. *Educational and Psychological Measurement*, *58*, 313-333.

Lewis-Beck, M.S. (1980). *Applied regression*. Beverly Hills, CA: Sage.

Marascuilo, L.A., and Levin, J.R. (1983). *Multivariate statistics for the social sciences*. Monterey, CA: Brooks/Cole.

Meehl, P.E. (1970). Theory testing in psychology and physics: A methodological paradox. In D.E. Morrison and R.E. Henkel, *The significance test controversy*. (Chapter 26, pp. 252-266). Chicago: Aldine.

Morrison, D.E., and Henkel, R.E. (1970). *The significance test controversy*. Chicago: Aldine.

Munro, B.H. (1997). *Statistical methods for health care research* (3rd. ed.). Philadelphia: Lippincott.

Newbold, P., and Bos, T. (1985). *Stochastic parameter regression analysis*. Beverly Hills, CA: Sage.

Pedhazur, E.J. (1982). *Multiple regression in behavioral research* (2nd. ed.). New York: Holt, Rinehart, and Winston.

Pruzek, R.M. (1997). An introduction to Bayesian inference and its applications. In L.L. Harlow, S.A. Mulaik, and J.H. Steiger (Eds.), *What if there were no significance tests*? (Chapter 11, pp. 287-318). Mahwah, NJ: Erlbaum.

Robinson, D.H., and Levin, J.R. (1997). Reflections on statistical and substantive significance, with a slice of replication. *Educational Researcher*, *26*, 21-26.

Schmidt, F.L. (1992). What do data really mean? Research findings, meta-analysis, and cumulative knowledge in psychology. *American Psychologist*, *47*, 1173-1181.

Schmidt, F.L. (1996). Statistical significance testing and cumulative knowledge in psychology: Implications for the training of researchers. *Psychological Methods*, *1*, 115-129.

Schroeder, L.D., Sjoquist, D.L., and Stephan, P.E. (1986). *Understanding regression analysis*. Newbury Park, CA: Sage.

Steiger, J.H., and Fouladi, R.T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L.L. Harlow, S.A. Mulaik, and J.H. Steiger (Eds.), *What if there were no significance tests*? (Chapter 9, pp. 221-257). Mahwah, NJ: Erlbaum.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences* (3rd. ed.). Hillsdale, NJ: Erlbaum.

Taylor, D.J., and Muller, K.E. (1995). Computing confidence bounds for power and sample size of the general linear univariate model. *The American Statistician*, *49*, 43-47.

Thompson, B. (1993). The use of statistical significance tests in research: Bootstrap and other alternatives. *Journal of Experimental Education*, *61*, 361-377.

Thompson, B. (1996). AERA editorial policies regarding statistical significance testing: Three suggested reforms. *Educational Researcher*, *25*, 26-30.

Thompson, B. Review of What if there were no significance tests? *Educational and Psychological Measurement*, *58*, 334-346.

Zumbo, B.D., and Hubley, A.M. (1998). A note on misconceptions concerning prospective and retrospective power. *The Statistician*, *47* (Part 2), 385-388.

The Value of Multimethod Qualitative/Quantitative Research Methodology in an Educational Program Evaluation: A Case Study

Catharine C. Knight
University of Akron
Walter J. Kuleck
The Hennepin Group

Abstract

This paper presents the case of an evaluation of a literacy-based public school classroom intervention that had been implemented without initial regard for subsequent evaluation, to examine the value of a multimethod qualitative and quantitative approach. We show how this evaluation challenge was addressed using both techniques, and how they complement each other to create richer understandings. Given the small number of participants in and the brevity of the intervention, the finding of quantitative results that support the very positive qualitative results is encouraging. Thus, the multimethod qualitative/quantitative approach appears to have been a useful and informative one in such an applied setting.

Introduction

There are situations in this imperfect world when it is necessary to evaluate a program or intervention under circumstances that would seem to make such an evaluation impractical. In this paper we explore one such situation concerning a literacy-based intervention in urban elementary schools. It uses a case study approach in an exploration of the use of multimethod qualitative/quantitative techniques that might be useful if applied where evaluation had not been foreseen but was later required.

In natural settings such as public schools or municipal entities an intervention is often designed and implemented without incorporating explicit evaluative elements. This is regretfully so although, with some foresight, mechanisms for evaluation could have been planned to have been an integral part of the intervention. For example, often in such situations a baseline of the target dependent variables is neither established nor are initial conditions documented. Without baseline data, it is difficult *post hoc* to determine what changes the intervention has effected. Moreover, independent and confounding variables are rarely controlled such that the effects of the intervention can be distinguished from other influences on the parameters of interest. Consequently, it can be a challenging and even frustrating process to credibly evaluate the intervention's effects.

In one case experienced personally by one of the authors in 1977, nearly three billion dollars had been invested by the Law Enforcement Assistance Administration in projects and interventions in police departments and municipalities. When Congress some years later required documentation of the effects of these programs, a plausible evaluation could not be done because forethought had not been given to evaluation during the design of the intervention. For example, *a priori* measurements of desired outcomes (where they had even been made explicit!) rarely had been taken. Thus, it was difficult or impossible to demonstrate what outcomes could be legitimately attributed to a given program or intervention.

Qualitative research techniques permit retrospective studies with a rigor similar to those studies carried out concurrently with the intervention (Bogdan and Biklen, 1998). Consequently qualitative researchers may not see the lack of measurements of initial conditions or the omission of an *a priori* evaluation plan as a concern. In contrast, quantitative research depends on some type of inferred gain due to an independent variable (Newman and Benz, 1998), in order to perform statistical analyses and find mathematical significance. However, sponsors and funding agencies may more often be interested in clear, unambiguous quantitative data than understanding process and nuance when considering their policy decisions (Bogdan and Biklen, 1998). Therefore, for them the lack of quantitative evaluative components may be a major issue.

In this paper we will present the case of an evaluation of a literacy-based public school classroom intervention that had been implemented without initial regard for subsequent evaluation, as an example of the value of both the qualitative and quantitative approaches in an applied setting. It will show how this evaluation challenge was addressed using both qualitative and quantitative techniques. Finally, the use of this multimethod model will be examined to ascertain the differential information afforded by its qualitative and quantitative components.

The Challenge

We evaluated an intervention consisting of an educational writing process known as Classroom Publishing or Bookmaking (Marzollo, 1991). Classroom Publishing had

been introduced in two Midwestern urban elementary schools, one a predominately low income inner-city school (Urban School) and the other a science magnet inner-city school (Science Magnet). At Urban School, 278 students in nine classes from grades one through four were introduced to Classroom Publishing. A more widespread intervention had been in place at Science Magnet, where 582 students in eighteen classes from grades kindergarten through five participated. The length of the intervention prior to the evaluation had been brief: one and a half school years.

Classroom Publishing is intended to improve student's writing, reading and language skills. The process is highly contextual in nature, immersing its participants in a literacy-rich environment. It encourages children to develop their creative and expressive skills as they typically write and illustrate books for an audience beyond their classroom (Domsky, 1990; Vacca, Vacca, and Gove, 1995). Some proponents believe that the process of Classroom Publishing generalizes to broader thinking strategies and to subjects beyond language and communication (Zemellman, Daniels, and Hyde, 1993).

The challenge was to evaluate the efficacy of Classroom Publishing in the target urban schools without the benefit of an initial provision within the intervention for evaluation. Thus, *a priori* measures and baseline were not available. Consequently, a triangulation approach (McMillan, 1996;Newman and Benz, 1998) was designed that employed both qualitative and quantitative strategies. The combination of methods was constructed to extract the greatest practical meaning from the data to get the clearest possible understanding of the value and effectiveness of the intervention.

Considerations

In the past the preferred approach used to analyze the results of an intervention has been primarily quantitative (Shaker, 1990), in which descriptive and inferential statistics are applied to data that can be quantified.

An alternative approach to the analysis of the results of an intervention is qualitative, in which data are viewed from a qualitative rather than quantitative perspective. Descriptions of pre- and post-intervention conditions are contrasted, with inductive reasoning applied to judge what relations among them seem reasonable, logical and appropriate, i.e., credible.

Both quantitative and qualitative approaches have their adherents, bodies of knowledge and places in research and evaluations. In a sense they reflect alternative perspectives on the nature of reality (Firestone, 1987). In this specific evaluation, the lack of appropriate pre-intervention measures as well as limited control imposed by field conditions restricted the applicability of the quantitative approach, as preferred as it might be by the intervention's sponsor. The quantitative approach was used in a limited way due to the circumstances described; it then became an adjunct to qualitative methods. Both perspectives were anticipated to complement each other.

Evaluation Issues

Given these considerations, an evaluation for this classroom intervention was constructed. According to Newman and Deitchmnan (1983), the keys to a "good evaluation" are in answering, among others, these questions (selected due to their particular relevance to this study):

1. What are the purposes of the evaluation?
2. Can these purposes be assessed?
3. What potential effect will the evaluation have on the project?
4. Do the evaluations reflect the concerns and interests of all the interested parties associated with the project?

In this case, both the project implementers and the schools' teachers and administrators were consulted in answering these questions before this evaluation was initiated. All parties agreed that the purpose of the evaluation was to build a case, if possible, for continued funding for the intervention. All parties believed that the basis for this case should be the improvement in academic accomplishment of the student participants, and that improvement could be gauged. Finally, all agreed that the continued funding of the project was in the best interests of the students based on their anecdotally-based perceptions of the evaluation's effects. The input from administrators and teachers contributed to the guidelines for developing the evaluation plan in order to maximize evaluative credibility (Newman and Deitchman, 1983). However, the positive bias of these parties towards the program was clear.

The resulting evaluation plan was to provide evidence of the intervention's effectiveness in support of grant-seeking proposals to provide continued funding for the Classroom Publishing process. Thus, in response to Newman and Deitchman's (1983) questions above:

This was to be an objectives-based study per the taxonomy of Stufflebeam and Webster (1983), where the purpose is to determine whether objectives have been achieved.

Since it was determined that quantitative methods alone would be insufficient (given the lack of intervention-specific pretest measures) to evaluate the comprehensive effects of Classroom Publishing on student thinking and behavior, a range of qualitative data-gathering procedures were designed.

The evaluation would likely have a significant effect on the future of the project as continued funding would be at least in part affected by the findings of this evaluative study.

Finally, the qualitative strategies were designed to gather data from teachers, parents, students and administrators, to ensure that all four groups of stakeholders were represented.

Primary Concern: Validity

In constructing an evaluation, Hedrick, Bickman, and Rog (1993) assert that it is important to plan an analysis carefully to ensure an efficient study that answers the critical research questions investigated. Indeed, Newman and Deitchman (1983) assert "One has to look at the objectives and identify the most relevant methods of measuring them in terms of practicality and validity. The validity must be considered not just in terms of tests but also in terms of credibility to the community involved… if the objectives or criteria are not perceived to be relevant to people with different interests, as in the multiple stakeholder approach, there is no way of achieving credibility" (p.294). In this evaluation case, a predominately qualitative approach was anticipated to be useful in that it likely accessed the richest content; the qualitative data were to be complemented and ideally supported by the available quantitative data. As McMillan (1996) points out, "If the results of several methods of collecting data agree, the finding is judged to be credible" (p.251).

Further, from the quantitative perspective this study was a clear example of ex post facto research, where "the investigators decide whether one or more preexisting conditions have caused subsequent differences between subjects who experienced different types of conditions (the phrase ex post facto means 'after the fact')" (McMillan, 1996, p. 185). McMillan goes on to emphasize that the researcher should select subjects to be as similar as possible except for the independent variable(s) being studied. Because the participants in the classroom publishing process had already been determined, the researchers in this study had to find "control groups" as similar as possible to those participants. However, in compensation, ex post facto research does provide an opportunity to study effects in a natural setting (Wiersma, 1995), increasing the perceived credibility of the results and their acceptance by practitioners. After much analysis and reflection, we selected a combination of quantitative and qualitative methodes to evaluate the effectiveness of Classroom Publishing in these two urban schools. The constituent groups affected by or concerned with the intervention were identified: students, teachers, and parents. A collection of methods to assess qualitative data from these groups was designed. Then, the available quantitative data were identified and data matching and analysis procedures prepared. Unfortunately, the available quantitative data very were limited, not designed for this study and subject to significant confounding, and not necessarily of high validity in this context.

Methodology

Four types of data, three qualitative and one quantitative, were assembled for this evaluation plan. First, a series of brief, tightly structured group interviews was held for the teachers who participated in the Classroom Publishing process to capture their views and ideas toward the value and effectiveness of this process. In this method, the facilitator asks a probe question and then supports the discussants' process. In this particular variation the facilitator records the group's responses (in this case on 3"x5" cards placed in view of the group) and then organizes them into related clusters or narratives as appropriate (Kuleck and Knight, 1988). The nature of the process used creates consensus. The probes used in the teachers' focus groups were:

1. What did you like about Classroom Publishing?
2. What did you **not** like about Classroom Publishing?
3. What would you change?

Second, a brief, tightly focused questionnaire was administered to parents of participating children to gather qualitative data about parental views toward the value and effectiveness of Classroom Publishing. The parent questionnaire probes used were:

1. Did you see your child's finished book? What did you think about it?
2. Do you think Classroom Publishing helped your child with reading, writing and language? Why or why not?

Third, participating children were invited to comment in writing on their Classroom Publishing experience. Two classes did so, one with letters to the Classroom Publishing instructor and the other by writing and publishing a book containing their feedback.

Fourth, standardized test results were used to determine what effect participation in Classroom Publishing had on student writing skills, using standardized data from the State Fourth Grade Proficiency Test. For one school, classes that participated in Classroom Publishing could be contrasted with those that did not. For the other school, because all the classes in one grade participated, a demographically matching school that did not participate in Classroom Publishing was identified for comparison purposes.

Evaluation Results

Qualitative Data

Qualitative data: Teachers

Data were collected and recorded using structured group interviews with teachers whose students participated in Classroom Publishing. Sixteen teachers from Science Magnet (of the eighteen participating in the intervention) and nine teachers from Urban School (of ten) attended the sessions. The procedure used was an iterative, interactive real-time synthesis of the typological analysis and constant comparison protocols described by LeCompte and Preissle (1984). At the end of each session, the process notes and their emergent patterns were reviewed with the teacher group to ensure the accuracy of the data and their consensus with regard to them.

The teachers from both schools were extremely enthusiastic about the use of Classroom Publishing. The focus

Table 1
Teacher Qualitative Data

Basic Skills

- "Classroom Publishing fosters (the) ability to write complete sentences."
- "Classroom Publishing helps prepare for standardized reading tests" (mentioned in both groups).
- "Classroom Publishing reinforces skills, enforces rules, e.g., grammar" (also mentioned in both groups).
- "Classroom Publishing extended their vocabulary."
- "Classroom Publishing fostered comprehension skills."
- "Through Classroom Publishing, kids learned to edit."

Motivation, pride and self esteem

- "Kids couldn't believe their own progress."
- "Classroom Publishing fostered self-esteem through accomplishments."
- "Classroom Publishing demonstrated everybody can do something—well" (mentioned in both teacher groups).
- "Classroom Publishing so motivated some kids they are on their third book." (also mentioned in both groups).

Development of thinking processes and strategies

- "Classroom Publishing made kids organize thoughts."
- "Classroom Publishing fostered sequencing: beginning, middle, end."
- "Classroom Publishing fostered thinking through to conclusion."
- "Classroom Publishing helped to develop understanding of fantasy vs. reality."
- "Classroom Publishing helped develop understanding of the Main Idea."
- "Classroom Publishing helped kids learn the parts of books; they now begin a book by checking out the publisher, illustrator, etc."
- "Classroom Publishing's influence extended even to math, e.g., calculating the percentage of pages that were pictures."
- "Classroom Publishing-fostered skills carried through to creative writing and social studies."

group responses from both schools were pooled under three domains: basic skills, motivation, and thinking skills.

Basic Skills

Consistently, one of the stated objectives of Classroom Publishing was to help students develop their basic skills. The teacher groups indicated that the process was broadly successful in doing so. Representative responses are found in Table 1.

Motivation, pride and self esteem

An often-mentioned outcome of Classroom Publishing in the teacher groups might be termed motivation, pride and self-esteem. Classroom Publishing was reported to be so motivating that children elected to spend time on this process to the extent of producing two and even three books during the school year, well beyond what was expected. Some representative comments are also found in Table 1.

Development of thinking processes and strategies

Perhaps most important from the teachers' views were those related to the results Classroom Publishing appeared to have on the development of thinking processes and strategies. Examples of teacher comments are found in Table 1.

Further, Classroom Publishing's influence appeared to generalize; typical teacher responses included: "Classroom Publishing's influence extended even to math, e.g., calculating the percentage of pages that were pictures" as well as "Classroom Publishing-fostered skills carried through to creative writing and social studies."

Summary of teacher qualitative data

Teachers indicated that the greatest value of Classroom Publishing may not be in developing language skills—as vital as they are—but in supporting more general cognitive development, e.g., providing a supportive environment to enhance the development of concepts such as text structure, as evidenced by the proper use of paragraphing. More generally, Classroom Publishing appeared to develop children's overall thinking and problem solving skills while fostering teamwork and building goal orientation. Enhanced self-esteem in children was thus an unsurprising outcome.

In addition, Classroom Publishing was seen by teachers to provide a valuable means for constructive self-expression, allowing children to begin to expand their inner life to encompass the world and others around them.

Finally, teachers reported that Classroom Publishing is a process that reached students wherever they may be in their cognitive and intellectual development. Those students with a need to develop organized thinking are stimulated to develop systematic thought processes. Those children demonstrating higher order thinking found those skills enhanced and, more importantly, apparently generalized to other subject areas, including math and science.

Qualitative data: Parents

Parent data were collected via a questionnaire sent home with children from two classes in each school. At Science Magnet, 23 of 36 questionnaires were returned (two of the 23 parents noted that they had not seen their child's book and had no comment). At Urban School, 17 of 45 parents responded. Recognizing that the parent questionnaires were likely returned by parents more involved in their child's schooling and thus likely inclined favorably to the Classroom Publishing process, their comments were uniformly positive.

Among the outcomes often cited by the parents were improved reading, writing and language skills. Parents found that Classroom Publishing provided new outlets for self-expression and creativity. The parents (as had the teachers)

noted that students found Classroom Publishing encouraged the constructive expression of inner thoughts and feelings; typical parent responses included: "I think it helps her to express herself and get enjoyment out of reading" and "It revealed inner thoughts expressed for the first time."

The parents also reported enjoyment from their children's products and expressed pride in them. One parent evidently shared her daughter's book more widely: "I think it was very good. I like her part about (company name) and so did my managers."

Children's enhanced self esteem as a product of accomplishment was also a noteworthy outcome as cited by parents: "...I think this is a great way to promote self esteem." "(child's name) is very excited about writing his book." "...he wants to write more books. He also reads more."

Perhaps two parents summed it up best: "I would like (child's name) involved in more projects similar to this one" and "...it will help a lot of other kids like it helped my child." Examples of typical parent comments may be found in Table 2.

Table 2
Parent Qualitative Data

Improved Reading, Writing and Language Skills

- "...it helps her to read and understand what she is reading."
- "Writing a book involved a use of spelling words known and learning new ones. Reading vocabulary was increased with a carry over to their daily tasks. Speaking skills were practiced whenever the book was read to friends, teachers, family and once to the school assembly."
- "...she did the work and used her imagination and that showed me some of the skills learned from good teaching in other words it really went in and stayed in."

Self-Expression and Creativity

- "My thought was that the book was very creative and she used a lot of thinking in order to create her book."
- "The story line was a surprise."
- "Her book was well laid out, and took some imagination for the story. The best I can remember, the spelling was correct. But I am not sure whether a lamb and a coyote will marry and live happily after ever."

Parental Enjoyment of and Pride in Children's Products

- "The finished book will be a treasure to be remembered now and in the years to come."
- "...I enjoyed reading it and have kept it for a keepsake."
- "Yes I saw his book. I think it is wonderful."
- "...I'm very proud of ____'s book and can't wait to show it off."
- "I think it was a wonderful short story."
- "I think it was great."

Qualitative data: Students

Two classes responded to a request for feedback about Classroom Publishing. The group of students from Science Magnet responded in the form of letters to the Classroom Publishing instructor. The other class, from Urban School, actually wrote and published a book containing their feedback! Examples from both of these sources are found in Table 3. One student's comment succinctly summarized the Classroom Publishing process:

> I like Classroom Publishing because it helps me in reading and writing. I like it when we first started making our stories. We started out with a story map to help us organize our ideas. Then we took our journal home and started writing our stories. When mine was finished I brought back my story to school and had (the instructor) help with my spelling. Then when I was done I wrote it on good paper and went over the words with black marker. I started illustrating my story. Then I bound it together and made a book.

It is clear that the students who responded validated the comments and opinions of their teachers and parents. The

Table 3
Student Qualitative Data

Letters to the Classroom Publishing Instructor

- "I liked it because we go to draw and make up our own words. I am happy because I was the Author and Illustrator."
- "I had fun making a book and publishing a book. You also taught me to do I what I couldn't do. Now I can make (it) through this grade because I remembered what you said."
- "This book publishing Program is great! My parents loved my book. My Mom asked me if she could take it to work. So if the Program could go on, I would be very happy to write a book again."
- "I didn't know how to write a book that good until you came to our class. I like to make books now."
- "I used to not even write a paragraph so you should know how good and helpful you were to me."
- "I used to hate writing stories. Now I can use my imagination. When we had to follow the chart, each time my mind would open wider and wider... Know (*sic*) when I go to my grandmother's I write all the time."

From the Classroom Publishing Book

- "I like that it helps my writing, drawing and creativity."
- "I learn a lot an (*sic*) I know a lot about indenting and paragraphs."
- "My mother was very proud of how I improved on my paragraphs. I liked illustrating and writing the book."
- "I like classroom publishing because it helps me read, wrote, learn how to draw, spell and work hard."

children's enthusiasm and energy for Classroom Publishing were undoubtedly responsible for the enthusiasm seen from the teachers and parents for the process.

Quantitative Data

While the qualitative data were strong in their endorsement of the process, there still remained the question of whether Classroom Publishing did in fact contribute to objectively measurable improvements in basic skills, particularly writing. The research staff of the City School District was consulted to determine possible alternatives for developing quantitative data. After finding that very little quantitative data were available, the evaluators and research staff determined that the only available measure of Classroom Publishing's effect, if any, would be the standardized State Fourth Grade Proficiency Test as administered by the City School District.

Fortunately, the State Test included specific sections to assess writing and reading skills. Only the scale scores from these sections were made available to be used as a quantitative measure of student writing, reading and (by inference, language) skills. Unfortunately, only the fourth grade would be represented at each school, comprising but 19 of the 278 student participants at Urban School (one classroom out of a total of 58 fourth graders) and 102 of the 582 at Science Magnet. The choice was then not which measures and procedures to use, but whether to proceed with what was available or abandon the quantitative side of the model altogether. The former course was chosen in order to strengthen the evaluation and satisfy the sponsor's needs for compelling data.

The two schools participating in Classroom Publishing, while both inner city schools, differed significantly from each other. "Science Magnet" is a magnet school; parents must apply for their children to attend. "Urban School" is a more typical inner city school, adjacent to a large County Metropolitan Housing Authority facility. We may reasonably infer that the students at Science Magnet benefited from a higher level of parental involvement when compared with that of Urban School. As parental involvement is an important contributor to student success, Science Magnet's relatively higher level of student success is not surprising. Consequently, for the purposes of this evaluation we deemed it advisable to consider the results of the two schools separately.

Classroom Publishing was implemented across the entire Fourth Grade at Science Magnet. Consequently, at Science Magnet, it was not possible to compare the performance of students who participated in Classroom Publishing with those who did not participate. Consequently, another school, Computer Magnet, was suggested by staff of the City School District to be highly comparable to Science Magnet with regard to demographics and geographic location. For example, the current poverty rate (as measured by participation in the School Lunch Program) at Computer Magnet is 88.41%, while the corresponding rate for Science Magnet is 90.46%.

When comparing the writing performance results from the Fourth Grade Proficiency Test writing scale from the two magnet schools, we found that 70% of the Science Magnet students (who had experienced Classroom Publishing) achieved proficiency (defined as a score of 4 or better) compared with 57% of those at Computer Magnet (who had not).

Table 4
Expected Writing Score Chi-Square Test between Schools from the Fourth Grade Proficiency Test (1996 Science Magnet c.f. 1996 Computer Magnet)

Classes	Non-Proficient Students	Proficient Students	Percentage Passed
Observed Pass-Fail Distribution for Classroom-Publishing Classes (Science Magnet, 1996)	31	71	70%
Expected Pass/Fail Distribution based on No-Classroom-Publishing Classes (Computer Magnet, 1996)	43.86	58.14	57%

Chi-Square Probability (Yate's Correction) = .08

Table 5
Expected Reading Score Chi-Square Test between Schools from the Fourth Grade Proficiency Test (1996 Science Magnet c.f. 1996 Computer Magnet)

Classes	Non-Proficient Students	Proficient Students	Percentage Passed
Observed Pass-Fail Distribution for Classroom-Publishing Classes (Science Magnet, 1996)	21	81	80%
Expected Pass/Fail Distribution based on No-Classroom-Publishing Classes (Computer Magnet, 1996)	38.76	63.24	62%

Chi-Square Probability (Yate's Correction) = .02

Table 6
Expected Writing Score Chi-Square Test within School from the Fourth Grade Proficiency Test (1996 c.f. 1995 Test Results, Science Magnet)

Classes	Non-Proficient Students	Proficient Students	Percentage Passed
Observed Pass-Fail Distribution for Classroom-Publishing Classes (Science Magnet, 1996)	31	71	70%
Expected Pass/Fail Distribution based on No-Classroom-Publishing Classes (Science Magnet, 1995)	26.52	75.48	74%

Chi-Square Probability (Yate's Correction) = .46

Table 7
Expected Reading Score Chi-Square Test within School from the Fourth Grade Proficiency Test (1996 c.f. 1995 Test Results, Science Magnet)

Classes	Non-Proficient Students	Proficient Students	Percentage Passed
Observed Pass-Fail Distribution for Classroom-Publishing Classes (Science Magnet, 1996)	21	81	80%
Expected Pass/Fail Distribution based on No-Classroom-Publishing Classes (Science Magnet, 1996)	26.52	75.48	74%

Chi-Square Probability (Yate's Correction) = .39

Table 8
Expected Writing Score Chi-Square Test within School from the Fourth Grade Proficiency Test (1996 Test Results, Urban School)

Classes	Non-Proficient Students	Proficient Students	Percentage Passed
Observed Pass-Fail Distribution for Classroom-Publishing Classes (Urban School, 1996)	13	6	32%
Expected Pass/Fail Distribution based on No-Classroom-Publishing Classes (Urban School, 1996)	12.67	6.33	33%

Chi-Square Probability (Yate's Correction) = .13

Table 9
Expected Reading Score Chi-Square Test within School from the Fourth Grade Proficiency Test (1996 Test Results, Urban School)

Classes	Non-Proficient Students	Proficient Students	Percentage Passed
Observed Pass-Fail Distribution for Classroom-Publishing Classes (Urban School, 1996)	7	12	63%
Expected Pass/Fail Distribution based on No-Classroom-Publishing Classes (Urban School, 1996)	6.33	12.67	67%

Chi-Square Probability (Yate's Correction) = .612

As Table 4 shows, the χ^2 probability of .08 approaches significance. The corresponding reading proficiency percentages as shown in Table 5 were 80% at Science Magnet and 62% at Computer Magnet, with a χ^2 probability significant at the .02 level.

As a cross check, comparisons were then made between the observed frequencies of Science Magnet students (all who had experienced Classroom Publishing) who were found proficient and non-proficient in writing and reading in the Classroom Publishing class, and the predicted frequencies based on the 1995 proficiency testing for Science Magnet fourth graders. With this comparison, the χ^2 probabilities were not significant as shown in Tables 6 and 7. This comparison was chosen to keep the school constant as compared with the Science Magnet and Computer Magnet pairing; however the two groups were comprised of students from different years, confounding what effect the Classroom Publishing intervention might have had.

Comparisons were also made between the observed frequencies of Urban School students class found proficient and non-proficient in writing and reading in the Classroom Publishing class, and the predicted frequencies based on the three non-participating classes. Here the χ^2 probabilities were not significant, as shown in Tables 8 and 9. This comparison was chosen to complement the previous comparisons, in this case by keeping the year of testing and the school constant. However, the two groups were comprised of students from different classes, again confounding what effect the Classroom Publishing intervention might have had.

Summary of Quantitative Data

If Classroom Publishing were effective, we would expect to see the results of the State Fourth Grade Proficiency Test's Writing and Reading Scales to favor participating students over non-participating students. To test this, we selected two groups of non-participating students to compare with the participants from Science Magnet's fourth grade:

1. Computer Magnet fourth grade students
2. Science Magnet students from the prior year's fourth grade

Further, the students from the Urban School's two non-participating fourth grade classes were compared with the students from the Urban School fourth grade class that did participate.

The results of the analyses of both the writing and reading scale score data from the State Fourth Grade Proficiency Test revealed a significant result for the comparison of Science Magnet's Fourth Grade (all of whom participated in the Classroom Publishing process) and Computer Magnet's Fourth Grade (none of whom participated). However, when the previous year's fourth grade class at Science Magnet was used, the contrast was not significant. Likewise, the within-school comparison between the participating (in Classroom Publishing) Urban School Fourth Grade class and the two non-participating classes failed to approach significance. Therefore the quantitative data can be described as encouraging but not conclusive.

Summary of Classroom Publishing Evaluation Results

The qualitative data gathered from the teachers, students and parents in the first through fifth grade classes that participated indicate that Classroom Publishing appears to contribute to the writing competence and cognitive development of students. This appears to be true for a wide range of cognitive ability and basic skills. Thus, those students with a need to develop organized thinking were stimulated to develop more systematic thought processes. Those children demonstrating higher order cognitive processing found those skills enhanced and, more importantly, apparently generalized to other subject areas, including math and science.

Therefore, the qualitative approach yielded a positive evaluation of this Classroom Publishing intervention. However, the sponsors of the intervention were also interested in documenting quantified skill improvements that could be attributed to Classroom Publishing, in support of grant proposals. The quantitative data, writing and reading scale scores from the State Fourth Grade Proficiency Test, showed significant differences between the magnet inner-city school's participating Fourth Grade and the non-participating Fourth Grade in another, roughly matched magnet inner-city school. In the other school, comparisons of writing scale scores between the one class in which Classroom Publishing was introduced and the school's two non-participating Fourth Grade classes failed to reach significance. These results are mixed, but sufficiently encouraging to be considered supportive by the intervention's sponsors.

Discussion

Meeting the Evaluation Challenge

When Classroom Publishing in kindergarten through fifth grade classes at these two urban inner-city schools was instituted, evaluative components had been omitted. In order to construct a credible and useful evaluation of the process, we developed a set of post hoc, qualitative data-gathering procedures supported by available quantitative assessments, in this case the writing and reading assessment scales of the State 4th Grade Proficiency Examination. Each method was developed to be accessible and valid. In addition, while one site could be assessed "within school," the second site required a suitable comparison school. Another magnet school in the district served this purpose.

Hence, though the Classroom Publishing intervention in these urban schools seemed at first to be difficult to evaluate, a useful set of complementary strategies was eventually developed. The use of qualitative methods made it possible for the evaluators to ask the question, "Does Classroom Publishing work"? The quantitative methods made it possible for the evaluators to ask the question "How **well** does Classroom Publishing work?"—at least insofar as a measurable effect on standardized test scores may be concerned.

Qualitative Strategy and Credibility

Primary in any evaluation are issues of validity and credibility. With respect to the qualitative methods used, "There are no set standards for evaluation of the validity of a field research's conclusions, but this does not decrease the need to consider carefully the evidence and methods on which conclusions are based… Individual items of information can be assessed in terms of at least three criteria:

1. How credible are the informants?
2. Were statements made in response to the researcher's questions, or were they spontaneous?
3. How does the presence or absence of the researcher and/or the researcher's informant influence the actions and statements of group members?" (Becker, 1958 p.341-2)

The "informants" used, teachers, parents and students, were those best able to credibly describe the effects of Classroom Publishing in these classrooms. Further, we believe that the use of structured group interviews and open-ended questions helped insure that the responses of those from whom data were gathered were open and spontaneous. The consistency of the data indicates that the presence of the evaluator was not a significantly confounding influence (or, implausibly, that the evaluator identically influenced the focus groups and questionnaire respondents), and moreover that the triangulation (Newman and Deitchman, 1983) sought was achieved. Therefore, we concluded that the qualitative methods chosen appear to have allowed a meaningful evaluation of the Classroom Publishing intervention in these two Midwest urban schools.

Quantitative Reinforcement of Qualitative Findings

Though the quantitative data provided ambiguous results, it can be argued that the significant comparisons—that between the two magnet schools in the same year—were the most plausible. This is in view of the comparatively limited number of participating fourth grade students at Urban School (19 vs. 38 non-participating), which made the intra-school comparisons more problematic. Given the small number of participants in and the brevity of the intervention, the finding of quantitative results that support the very positive qualitative results is encouraging. Thus, the multimethod qualitative/quantitative approach in this applied setting appears to have been a useful and informative one. This result contrasts with the Head Start evaluation study that was reported by Bogdan and Biklen (1998), where qualitative methods, used in conjunction with quantitative measures, demonstrated that the qualitative results were misleading.

The Value of Multimethod Models: Issues and Opportunities

The plausibility of qualitative data can be increased by using as many appropriate sources and types of qualitative data that can be practicably gathered. This multimethod "triangulation" approach (Newman and Benz, 1998) can result in meaningful and credible results, leading to useful and practical conclusions. If the spectrum of available data sources or data gathering modalities is restricted, credibility will suffer and the utility of the conclusions will be diminished. Thus, it is incumbent on those using qualitative methods to cast their net as widely as possible. In this evaluation case, each of the stakeholder groups—teachers, parents, students and intervention sponsors—was given its opportunity to contribute to the evaluation.

Similarly, the greater the opportunity to assess quantitative data, the greater support they can give to the credibility of the qualitative findings. Of course, the inverse construction may also obtain; qualitative methods may be used to reinforce (or, in the case of the Head Start study that was reported by Bogden and Biklen (1998), repudiate), quantitative methods.

From a theoretical perspective, Newman and Benz (1998) point out that qualitative and quantitative methods are neither antithetical nor mutually exclusive. Rather, they are complementary sides of the same coin, an "interactive continuum." In this case, the evaluators endeavored to use the qualitative results to build a "theory": that Classroom Publishing facilitated the development of writing and reading skills. This "theory" was then tested using a quantitative model. While the quantitative results were only partially statistically significant, they did provide useful support for the qualitative-based "theory." The next step, of course, would be to refine the "theory" on the basis of the quantitative results, developing hypotheses to be tested qualitatively, and so on through the cycle of refinement (Newman and Benz, 1998).

In the final analysis, a useful and credible evaluation was constructed although the applied intervention was not initially designed with evaluation in mind. The sponsors of this Classroom Publishing intervention were reassured that Classroom Publishing was more than a "feel good" exercise and could pursue funding sources with both confidence and credibility.

References

Becker, H. S. (1958). Problems of interference and proof in participant observation. *American Sociological Review*, *23*, 652-660.

Bogdan, R. C., and Biklen, S. K. (1998). *Qualitative research for education: An introduction to theory and methods* (3rd ed.). Boston: Allyn and Bacon.

Domsky, D. (1990). On the write track: The writing process at work in a first grade classroom. *Writing Notebook: Creative word processing in the classroom*, *8*(1), 41-42.

Firestone, W. (1987). Meaning in method: The rhetoric of quantitative and qualitative research. *Educational Researcher*, *16*, 16-21.

Hedrick, T. E., Bickman, L., and Rog, D. J. (1993). *Applied research design: A practical guide*. Newbury Park, CA: Sage.

Kuleck, W. J., and Knight, C. C. (1988). Supporting Decision Making, Planning and Information Management: Cognitive Theory in Organizational Practice. In *Advances in Information Processing in Organizations*, Vol. 3, pp. 263-274. New York: JAI Press.

Lecompte, M. D., and Preissle, J. (1984). *Ethnography and qualitative design in educational research*. San Diego: Academic Press.

Marzollo, J. (1991). Bookmaking made simple. *Instructor*, *9*(2), 36-42.

McMillan, J. H. (1996). *Educational research: Fundamentals for the consumer* (2nd ed.). New York: Harper Collins.

Newman, I., and Benz, C.R. (1998). *Qualitative-quantitative research methodology: Exploring the interactive continuum*. Carbondale IL: Southern Illinois University Press.

Newman, I., and Deitchman, R. (1983). Evaluationresearch: A suggested move toward scientific and community credibility. *Journal of Studies in Technical Careers*, *4*, 289-298.

Schutt, R. K. (1996). *Investigating the social world: The process and practice of research*. Thousand Oaks, CA: Pine Forge Press.

Stufflebeam, D. L., and Webster, W. J. (1983). An analysis of alternative approaches to evaluation. In G. F. Madaus, M. S. Scriven and D. L. Stufflebeam (Eds.), *Evaluation models: viewpoints on educational and human services evaluation* (pp.23-43). Boston: Kluwer-Nijhoff Publishing.

Shaker, P. (1990). The metaphorical journey of evaluation theory. *Qualitative Studies in Education*, *3*(4), 355-363.

Vacca, J. L., Vacca, R. T., and Bove, M K. (1995). *Reading and learning to read*. New York: Harper Collins.

Wiersma, W. (1995). *Research methods in education: An introduction* (6th ed.). Boston: Allyn and Bacon.

Zemelman, S., Daniels, H., and Hyde, A. (1993). *Best practice: New standards for teaching and learning in America's schools*. Portsmouth: Heinemann.

Multimethod Analysis of Mathematics Achievement Tests

Dimiter M. Dimitrov
Kent State University

Abstract

Multimethod analysis of mathematics achievement tests is illustrated by combining psychometric and statistical methods in the analysis of results from the California Achievement Test-Mathematics administered to seventh-graders from North-East Ohio.

Taken into account were the category objectives and thinking skill levels defined for the two parts of the test, Computations and Concepts and Applications. The goal is to provide educational analysts results they can use in making informed decisions about teaching mathematics within local educational settings.

Data related to validity, reliability, scaling, norming, and equating are commonly provided with nationally standardized mathematics achievement tests (see, e.g., CTB/ McGraw-Hill, 1986). However, the results reported for local student populations are usually limited to classical item parameters and descriptive statistics of students' scores on such tests. Additional test data at state and district levels may provide research analysts information they can use to further support their decisions about teaching mathematics in local educational environments.

The purpose of this paper is to provide information that may help in making informed decisions based on CAT-M results, by combining Item Response Theory (IRT) and statistical methods in the analysis of results from the California Achievement Test-Mathematics (CAT-M) administered to seventh-graders from North-East Ohio. This study addresses a number of questions:

1. Which IRT model fits the CAT-M data for the target population?
2. How does the CAT-M work at different ability levels?
3. Does the average item difficulty change across different category objectives and thinking skill levels of the CAT-M?
4. Is the relative standing of students the same across different CAT-M items?
5. How many items are needed per CAT-M category objective and thinking skill level in order to obtain given reliability?
6. How can students' abilities be predicted from CAT-M scores?

Method

Results from the CAT-M (CTB/McGraw-Hill, 1985) of 4135 seventh-graders from a large urban area in North-East Ohio were used. The two parts of the CAT-M, Computation Test and Mathematics Concepts and Applications Test, were analyzed separately. The Computation Test included 50 items grouped by one factor, **Category Objective** (CO), with 10 levels: (1) Subtract fractions, (2) Multiply whole numbers, (3) Multiply decimals, (4) Multiply fractions, (5) Divide whole numbers, (6) Divide decimals, (7) Divide fractions, (8) Integers and percents, (9) Subtraction of whole numbers and decimals, and (10) Addition of whole numbers, decimals, and fractions (CTB/McGraw-Hill, 1986).

The Concepts and Applications Test included 55 items grouped by two factors. The first factor, Category Objective (CO), has six levels, (1) Numeration, (2) Number Sentences, (3) Number Theory, (4) Problem Solving, (5) Measurement, and (6) Geometry. The second factor, Thinking Skill (TS), has three levels, (1) Recall and recognition, (2) Inference, and (3) Evaluation.

The IRT analysis included the calculation of (a) data fit statistics, (b) item and test characteristics,(c) students' ability scores, and (d) descriptive statistics for test scores of students with different abilities. The computer programs RASCAL (Assessment Systems Corporation, 1995a) and XCALIBRE (Assessment Systems Corporation, 1995b)were used for the IRT analysis, while SPSS (SPSS Inc., 1997) and MicroFACT (Waller, 1995) were used for the statistical analysis.

A two-way unbalanced ANOVA was conducted for the Concepts and Applications Test with two fixed factors, CO and TS, with the dependent variable being the IRT difficulty of the items. It was performed through the SPSS procedure MANOVA/METHOD = SEQUENTIAL. Of special interest was the interaction between the two factors in order to see if the difference between the average item difficulties of different category objectives varied across the three thinking skill levels.

To answer the research question related to the prediction of students' abilities on CAT-M scores, a regression analysis was conducted with the independent variable being the test score and the dependent variable being the ability score. The ability scores of all 4135 students were calculated XCALIBRE.

Generalizability theory study (G-study) and related decision study (D-study) were conducted for the CAT-M tests by the use of the GENOVA program (Crick and Brennan, 1983). For the Computation Test, students (S) were the object of measurement and items (I) represented a random facet nested within the fixed facet Category Objective (CO). Thus,

the appropriate G-study design in this case was the partially nested design S x (I:CO) (see, e.g., Shavelson and Webb, 1991, p. 75). With the Concepts and Applications Test, a G-study was conducted for the partially nested design S x (I:TS), with items nested within the fixed facet Thinking Skill (TS).

Related D-studies were conducted with both the S x (I:CO) and S x (I:TS) designs for the estimation of the GT coefficients $E\rho^2$ and Φ. The **generalizability coefficient**, $E\rho^2$, is analogous to the reliability coefficients in classical test theory. It is suitable for decisions about the relative standing of students on the test scale. The **index of dependability**, Φ, introduced by Brennan and Kane (1977) as a generalizability index for absolute decisions, is suitable for criterion-referenced analysis and decisions (see, e.g., Shavelson and Webb, 1991, pp. 83-97).

Results

The IRT assumption about unidimensionality of the data was tested using MicroFACT (Waller, 1995), which performs the iterated principal factor analysis on tetrachoric correlations for binary response data. The results indicated the presence of a dominant factor underlying the students' performance on each test. For the Computation Test, 36.72% of the total variance was explained by the first factor versus 1.54 % explained by the second factor. For the Concepts and Applications Test, this ratio was 42.46 % versus 0.48% in favor of the first (dominant) factor.

The results of the IRT analysis showed that the one-parameter IRT (Rasch) model did not fit the CAT-M data. The RASCAL χ^2 fit statistic indicated misfit of 44 items from the Computation Test and 45 items from the Concepts and Applications Test, with χ^2 values of those items exceeding the critical value, $\chi^2(19)= 30.14$, at the level of significance $\alpha = .05$.

For data fit of the 2- or 3-parameter IRT models, XCALIBRE reported a standardized residual statistic for each item. This statistic is normally distributed and values in excess of 2.0 indicate misfit with a type I error rate of 0.05. The results showed that the data did not fit the 2-parameter IRT model. Standardized residuals in excess of 2.0 for 8 items from the Concepts and Applications Test and 20 items from the Computation Test were found. For each test, the data fit the 3-parameter IRT model because none of the standardized residuals exceeded 2.0.

The internal consistency reliability of each test was found to be 0.90. The information curves of the two tests are given in Figure 1. The average amount of information provided by the Computation Test was found to be 9.31 versus 7.39 provided by the Concepts and Applications Test. Thus, for the local population of seventh-graders, the Computation Test provided more accurate estimates of students' abilities as compared to the Concepts and Applications Test (see, e.g., Allen and Yen, 1979, pp. 262-267). This is especially true for students with ability scores between 0.0 and

Table 1
Item Parameter Estimates for the Computation Test

	Parameter					Parameter			
Item	*a*	*b*	*c*	PC	Item	*a*	*b*	*c*	PC
1	.43	-2.55	.14	87	26	.43	-1.37	.14	75
2	.65	-2.11	.14	89	27	1.45	.80	.20	42
3	.46	-2.00	.14	94	28	.37	.44	.17	54
4	.77	.07	.15	56	29	1.77	1.37	.12	23
5	1.64	.56	.17	45	30	1.96	1.14	.09	23
6	.67	-.90	.13	73	31	.75	-1.36	.13	81
7	.54	-.37	.13	63	32	1.00	-1.13	.12	81
8	.74	-.53	.14	68	33	.76	-.66	.11	69
9	.91	.28	.13	50	34	.95	-.58	.12	70
10	.94	.35	.14	50	35	.89	-.34	.13	65
11	.53	-2.41	.14	88	36	.86	-1.26	.13	82
12	1.65	.45	.14	46	37	.87	-1.57	.14	86
13	1.76	.80	.10	33	38	.97	-.79	.13	75
14	1.86	.42	.13	45	39	.93	-.29	.13	64
15	1.83	.67	.11	37	40	1.07	1.62	.13	23
16	.67	-1.52	.13	82	41	2.11	1.11	.13	28
17	.62	-1.80	.13	85	42	2.31	1.20	.17	30
18	.78	-1.34	.13	82	43	1.81	1.28	.17	28
19	.79	-1.13	.13	79	44	2.16	1.30	.12	23
20	.71	-.42	.12	65	45	1.48	1.12	.19	34
21	.47	-2.61	.14	88	46	.83	2.44	.16	22
22	.66	-1.17	.14	78	47	.84	1.49	.12	27
23	.71	-.61	.14	69	48	.34	-.09	.15	59
24	.91	.49	.16	48	49	.93	1.80	.16	26
25	.78	.39	.15	50	50	1.09	2.51	.09	11

Note: Used was the 3-parameter IRT model, with *a* = discrimination parameter, *b* = difficulty parameter, and *c* = "guessing"

	Parameter					Parameter			
Item	*a*	*b*	*c*	PC	Item	*a*	*b*	*c*	PC
1	.68	-3.00	.18	97	29	.91	-.94	.15	77
2	.58	-2.32	.17	89	30	.72	-.18	.18	63
3	.83	-2.50	.17	95	31	.70	-.18	.19	63
4	.93	-.41	.18	68	32	.87	.10	.16	56
5	.76	-1.70	.18	87	33	.89	.23	.17	54
6	.83	-1.81	.17	89	34	.46	.78	.19	49
7	.70	-2.01	.17	89	35	.97	-.19	.16	62
8	.76	-1.51	.18	85	36	1.26	1.09	.18	35
9	.51	-1.88	.18	84	37	.93	-.54	.17	70
10	.61	-.97	.18	75	38	.99	.64	.18	45
11	.54	-.79	.18	71	39	1.02	-.51	.16	70
12	.65	-.60	.18	70	40	.58	.10	.17	57
13	.76	-1.39	.18	83	41	.73	.67	.18	47
14	.67	-1.38	.17	82	42	.77	1.03	.17	78
15	.74	-.35	.18	66	43	.51	1.44	.17	79
16	.70	-1.10	.18	78	44	.81	-.16	.16	61
17	.56	-.72	.17	70	45	.86	.69	.21	47
18	.80	-.95	.16	76	46	1.49	.73	.14	38
19	.33	.99	.21	51	47	.90	.62	.16	45
20	1.13	-.58	.18	73	48	1.20	.91	.20	40
21	1.00	-.79	.18	76	49	.86	.56	.14	45
22	.64	-.24	.17	63	50	.88	1.61	.15	27
23	.55	.33	.19	54	51	.99	.28	.17	52
24	.56	.13	.20	58	52	1.04	2.17	.13	19
25	.62	.42	.18	52	53	.88	1.15	.16	35
26	.83	-.40	.18	67	54	1.05	1.64	.17	28
27	.79	-.74	.18	73	55	1.20	2.05	.12	17
28	.94	-1.52	.17	87					

2.0 on the logit scale, i.e. students above the average and below the top on the ability range of the target population. Beyond this interval, both tests do not work particularly well.

Table 1 provides estimates of *a* (discrimination parameter), *b* (difficulty parameter), and *c* ("guessing parameter") for the Computation Test. The table also shows the percent of correct answers (PC) for each item, based on 4135 students. The item difficulties were spread without any big gaps within the logit interval (-2.61 to 2.51). The item discrimination power varied within the relatively large interval (0.37 to 2.31). The "guessing" parameter, *c*, was quite small in magnitude and variability. This indicates that, for each item, there is small probability for students with low ability to answer the item correctly. The same pattern of findings was observed for the item parameter estimates of the Concepts and Applications Test (see Table 2).

Table 3

Test Score Means and Standard Deviations by Eight Ability Levels of the Students

Computation Test			Concepts and Applications Test		
Ability Interval			Ability Interval		
From - To	*M*	*SD*	From - To	*M*	*SD*
Below P_5 (-2.10)	11.39	3.25	Below P_5 (-1.87)	14.81	3.95
P_5 - P_{10} (-1.49)	15.75	2.35	P_5 - P_{10} (-1.34)	20.09	2.17
P_{10} - P_{25} (-.63)	19.65	2.42	P_{10} - P_{25} (-.67)	25.22	2.19
P_{25} - P_{50} (.08)	25.27	2.24	P_{25} - P_{50} (.02)	31.59	2.30
P_{50} - P_{75} (.77)	31.13	2.35	P_{50} - P_{75} (.70)	38.39	2.08
P_{75} - P_{90} (1.33)	37.50	2.29	P_{75} - P_{90} (1.41)	44.39	1.81
P_{90} - P_{95} (1.64)	42.15	1.55	P_{90} - P_{95} (1.83)	48.43	.92
Above P_{95}	45.54	1.75	Above P_{95}	51.51	1.27

Note: Given in parentheses are the values of the percentiles P_5, P_{10}, ..., P_{90}, P_{95} on the ability scale (in logits).

Table 4

Generalizability Study of the S x (I:CO) Design for the Computation Test and the S x (I:TS) Design for the Concepts and Applications Test

		Computation Test		Concepts and Applications Test	
Source of Variation	Variance Component	Estimated Variance Component	Percentage of Total Variance	Estimated Variance Component	Percentage Of Total Variance
Students (S)	σ_S^2	.0290	14	.0219	9
Items (I)	σ_I^2	.0288	14	.0624	26
S x (I:CO), E[a]	$\sigma^2_{S\times(I:CO),E}$	.1491	72		
S x (I:TS), E[b]	$\sigma^2_{S\times(I:TS),E}$			.1530	65

[a] For the Computation Test , with Category Objective (CO) fixed facet.
[b] For the Concepts and Applications Test, with Thinking Skill (TS) fixed facet.

Table 3 shows means and standard deviations of CAT-M scores for students at eight ability levels. Boundaries of the ability intervals are the percentiles P_5, P_{10}, P_{25}, P_{50}, P_{75}, P_{90}, and P_{95} on the ability scale (in logits).

Table 4 shows results from the G-studies conducted for the Computation Test, with the S x (I:CO) design, and for the Concepts and Applications Test, with the S x (I:TS) design. With each of the two designs including a fixed facet, the variance due to interaction between subjects and items is inseparable from the variance due to random error in each of the variance components $\sigma^2_{S\times(I:CO),E}$ and $\sigma^2_{S\times(I:TS),E}$. It should be noted, however, that the "guessing" part of the random error variance was relatively small (see the *c*-values in Tables 1 and 2). For the Computation Test, the variance component $\sigma^2_{S\times(I:CO),E}$ accounted for the largest part of the total variance, 72% . Hence, the relative standing of students on the computation scale changes a great deal across items. This was also true for the Concepts and Applications Test where the variance component $\sigma^2_{S\times(I:TS),E}$ also explained the largest part, 65%, of the total variance. Table 5 shows D-study results about relations between number of items and reliability coefficients $E\rho^2$ and ϕ. For relative decisions with the Computation Test, for example, a reliability of .90 or above ($E\rho^2 \geq .90$) requires at least six items within each category objective of the test. Similarly, for absolute decisions with the Concepts and Applications Test, a reliability of .90 or above ($\phi \geq .90$) requires at least 30 items per thinking skill level of the test.

Table 6 shows results from the 6 x 3 two-way ANOVA, using the item difficulty as the dependent variable and the fixed factors CO and TS of the Concepts and Applications Test as independent variables. The non-significance of the main effects, CO($F(5,39) = 2.06$, $p = .092$) and TS($F(2, 39) = 1.49$, $p = .237$), indicates that the average item difficulty is the same across all category objectives and, separately, across all thinking skill levels. The significance of the interaction between the two factors, CO x TS($F(6,39) = 2.62$, $p = .031$), shows that the difference between the average item difficulties of the category objectives varies across the thinking skill levels of the test.

Table 5

Decision Study of the S x (I:CO) Design for the Computation Test and the S x (I:TS) Design for the Concepts and Applications Test

Computation Test			Concepts and Applications Test		
Number of Items	$E\rho^2$	Φ	Number of Items	$E\rho^2$	Φ
1	.661	.620	6	.720	.646
2	.796	.766	10	.811	.753
3	.854	.830	15	.866	.821
4	.886	.867	20	.896	.859
5	.907	.891	25	.915	.884
6	.921	.907	30	.928	.901
7	.932	.919	35	.938	.914
8	.940	.929	40	.945	.924
9	.946	.936	45	.951	.932
10	.951	.942	50	.955	.938

Table 6

Unbalanced 6 x 3 (CO x TS) ANOVA design with Dependant Variable the Item Difficulty for the Concepts and Applications Test

Source	SS[a]	df	MS	F	p-value
Model	29.34	13	2.26	2.23	.027
Category Objective (CO)	10.41	5	2.08	2.06	.092
Thinking Skill (TS)	3.02	2	1.51	1.49	.237
CO x TS	15.91	6	2.65	2.62	.031
Within + Residual	39.47	39	1.01		
Total	68.82	52	1.32		

[a] SEQUENTIAL Sums of Squares Source via SPSS (Windows, v. 6.1).

Regression analysis was conducted in an attempt to find a simple model for predicting students' abilities on CAT-M scores. Students with ability scores beyond the interval bounded by ±3.0 on the logit scale, representing about 1% of the 4135 students for each CAT-M test, were excluded from the regression analysis in order to avoid the "outliers" effect. Figure 2 represents an edited SPSS output from the simple linear regression analysis conducted for the Computation Test. The Multiple R of 0.97 indicates an extremely high positive correlation between observed and predicted ability scores of the students. Also, $R^2 = 0.94$ shows that 94% of the differences in the ability scores of the students are explained by differences in their test scores. The regression equation in Figure 2 provides simple and significant prediction of the abilities on test scores. Its graphical representation is given in Figure 3. Almost identical regression results were found for the Concepts and Applications Test (see Figure 4). With this test, 97% of the students' ability variance was explained by the test score variance and, again, the simple linear regression provided highly significant prediction of the abilities on test scores (see, also, Figure 5).

```
                        COMPUTATION TEST
Dependent variable: ABILITY
Multiple R            .96770
R Square              .93643
Adjusted R Square     .93642
Standard Error        .27260
                     Analysis of Variance:
                DF   Sum of Squares      Mean Square
Regression       1        4471.8224       4471.82237
Residuals     4085         303.5487           .07431
F =   60179.44220       Prob > F =   .0000
-------------------- Variables in the Equation --------------------
Variable           Parameter   Standard     T for H0:
                    Estimate     Error     Parameter=0   Prob > |T|
TEST SCORE          .119996     .000489      245.315       .0000
Constant          -3.414913     .014586     -234.118       .0000
------------------------------------------------------------------------
Regression equation: ABILITY = (.120)(TEST SCORE) - 3.415
```

Figure 2. Edited SPSS output from the simple linear regression of ability scores on test scores for the Computation Test.

Discussion

Along with the standard information about CAT-M results, provided to local educational analysts, there are additional findings that should be taken into account for the target population of seventh-graders. In the context of the research questions in this study, several findings are important.

First, the Rasch and 2-parameter IRT models did not fit the data for the CAT-M with the target population. This finding suggests that the items differed in discriminating seventh-graders with different ability scores and that there were "guessing" effects, although they were found to be relatively small. The CAT-M data did fit the 3-parameter IRT model for the target population.

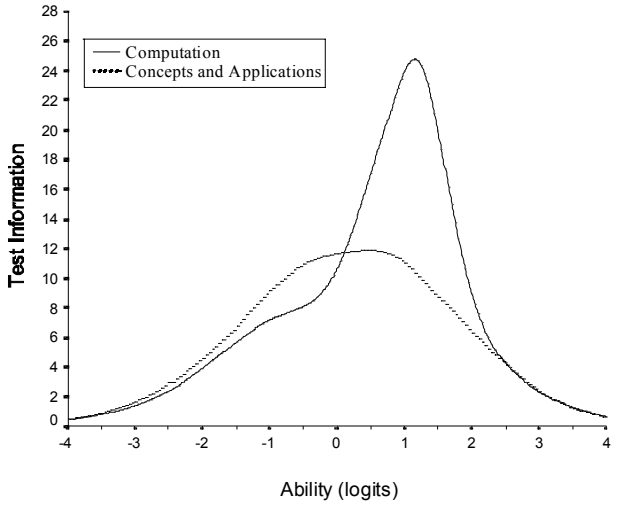


Figure 1. Test information curves for the Computation and Concepts and Applications Tests.

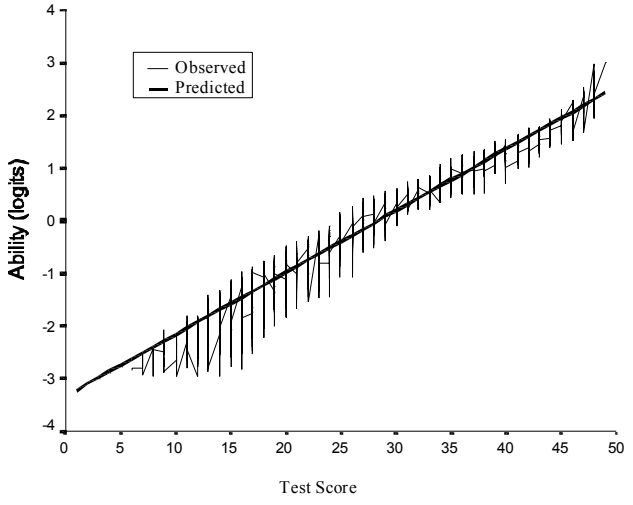


Figure 3. Simple linear regression of ability scores on test scores for the Computation Test.

CONCEPTS AND APPLICATIONS TEST

Dependent variable: ABILITY

Multiple R	.98689
R Square	.97395
Adjusted R Square	.97395
Standard Error	.16665

Analysis of Variance:

	DF	Sum of Squares	Mean Square
Regression	1	4111.2600	4111.26000
Residuals	3959	109.9474	.02777

F = 148038.79497 PROB > F = .0000

Variables in the Equation

Variable	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob > |T|
TEST SCORE	.110919	.000288	384.758	.0000
Constant	-3.836017	.010393	-369.087	.0000

Regression equation: **ABILITY = (.111)(TEST SCORE) - 3.836**

Figure 4. Edited SPSS output from the simple linear regression of ability scores on test scores for the Concepts ans Applications Test.

Second, the Computation Test provided more information and, hence, more accurate estimates of students' abilities than the Concepts and Applications Test, within the range from 0.0 to 2.0 on the logit ability scale. Beyond this interval (i.e., for students with ability below the average and for high ability students) neither test worked particularly well. The results in Table 3 show how students at eight different ability levels performed on the CAT-M.

Third, for the Concepts and Applications Test, the difference between the average difficulty of items from different category objectives varied a great deal across the thinking skill levels. Fourth, the G-study results show that the relative standing of seventh-graders on the CAT-M scale changed a great deal across different items of the test. Fifth, the D-study results provided information about the number of items required to obtain desired reliabilities for both relative and absolute (criterion-related) decisions. Sixth, the regression analysis provided a simple and highly significant model for the prediction of students' abilities on CAT-M scores.

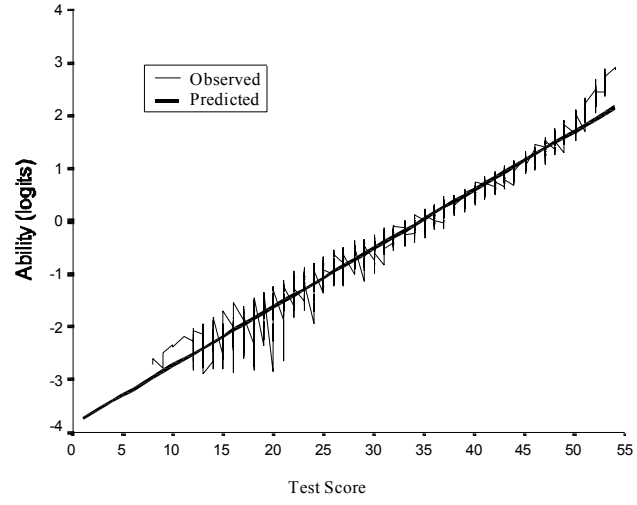


Figure 5. Simple linear regression of ability scores on test scores for the Concepts and Applications Test.

In conclusion, reports and interpretations of results of local student populations on nationally standardized mathematics are commonly based on descriptive statistics of test items and student total scores. The analysis illustrated in this article may help local educators and test analysts in interpreting test results by taking into account the ability levels of the students and the interaction between test factors such as item difficulty, category objectives, and thinking levels. In general, it provides valuable feedback for making informed decisions about teaching mathematics within local educational settings. Future research in this area will focus on relationships between psychometric and cognitive characteristics of the items. Also, one can apply the multimethod approach in the analysis of results from science, language, and other standardized tests administered to students representing large local populations.

References

Assessment Systems Corporation. (1995a). *User's Manual for RASCAL Rasch analysis program* (Windows version 3.5). St. Paul, MN: Author.

Assessment Systems Corporation. (1995b). *User's Manual for XCALIBRE marginal maximum-likelihood estimation program* (Windows version 1.0). St. Paul, MN: Author.

CTB/McGraw-Hill. (1985). *California Achievement Tests*. Forms E Level 17. Del Monte Research Park, Monterey, CA: Author.

CTB/McGraw-Hill. (1986). *California Achievement Tests*. Forms E and F. (Technical Bulletin No. 2). Del Monte Research Park, Monterey, CA: Author.

Crick, J., E., and Brennan, R. L. (1983). *Manual for GENOVA: A generalized analysis of variance system*. ACT Technical Bulletin, No. 43. The American College Testing Program. Iowa City, IA.

Shavelson, R. J., and Webb, N. M. (1991). *Generalizability theory: A primer.* SAGE. Newbury Park, CA.

SPSS Inc. (1997). *SPSS (Windows version 7.5): User's guide*. Chicago, IL: Author.

Waller, N. G. (1995). MicroFACT 1.0. *A Microcomputer factor analysis program for dichotomous and ordered polytomous data*. Assessment Systems Corporation. St. Paul, MN.

Book Review

Conducting Survey Research in the Social Sciences

John M. Linacre
University of Chicago

Newman, Isadore & McNeil, Keith. (1998) *Conducting Survey Research in the Social Sciences*. Lanham, Maryland: University Press of America.

Summary

This books outlines a constructive step-by-step approach to survey research, presented as 23 questions concerning issues to be addressed in the formulation, administration and reporting of surveys. Answers to these questions are presented as points to be considered and lists of readings. Good, practical exemplars are lacking.

Text

"The purpose of this text is to present basic concepts and general guidelines for those who are interested in conducting a survey" our authors state in their Preface. Do they succeed in their purpose?

There are numerous components that comprise the survey process. To simplify matters, the book presents scenarios involving three prototypical researchers, a psychologist, an administrator and a curriculum evaluator. This is a good approach. Would that our authors had walked us step-by-step through three actual survey projects to their successful completion—but no, they never rise to the challenge. Instead, they give us platitudes, "Once he (the psychologist) feels comfortable ..."

In fact, the book seems to regard a survey as an academic exercise performed to meet some requirement (such as a dissertation), rather than as a serious endeavor to accumulate knowledge. Graduate students, rest assured: diligently, no *slavishly,* follow the eight steps laid out in this book, and your Committee will be impressed! Your sheepskin is as good as inscribed. But your dissertation will languish on the Library shelf along with thousands of others, never to be read again.

Let us, however, imagine that we have a serious intent to gain useful knowledge. We need help and turn to this slim volume to provide it. There are 65 pages of concepts and guidelines, and a further 40 pages of examples and reference material. The 65 pages are divided into 8 chapters and presented as answers to 23 questions. This makes the text a brisk read and information easy to find.

So what information is provided? Two chapters and six appendices are devoted to specifying the research question and writing the survey instrument; three chapters and 2 appendices to defining the target population, selecting a sample and collecting data; one chapter and 4 appendices to writing the report; and two chapters and one appendix to support resources. This understates the role of resource material in this book. Each of the 23 questions, e.g., "Question 14: What survey procedure should I use?", is provided with a supplementary reading list. In the case of Question 14, the "answer" is 1½ pages long and the additional reading list another page. Indeed to resolve the many issues raised in the "answer" the serious researcher will need to refer to the reading list. Fortunately, the 23 separate reading lists are condensed into a 3½ page general reference list slipped between the Chapters and the Appendices. One book mentioned frequently in the lists is Rossi et al.'s (1983) *Handbook of Survey Research*. Plan to have ready access to that volume.

Where is our book most lacking? "At the present stage of development of the survey method, ... question wording [is] the Number One problem" (Payne, 1951, p. 4–5). In the chapter entitled, "Develop the survey", our authors provide three question exemplars:

(i) "Have you ever had the problem of not being able to stop smoking?"

(ii) "How often have you had the problem of anxiety?"

(iii) An interview question, "Since your hospitalization coverage doesn't cover any of the problems that we have discussed, how much will you be willing to pay to try to solve the problem?"

Asking convoluted, syntactically and semantically dubious questions such as these is certain to make the respondents' answers uninterpretable and the results of the survey unreproducible.

Where this book shines, however, is in its sparkling collection of 27 pithy quotations, sprinkled throughout the text. Only 2 or 3 were familiar to me. Though several didn't relate to their contexts, all were memorable!

Is this book worse than other similar works? No. Is it a useful starting point? Yes. *"All therefore whatsoever they bid you observe, that observe and do; but do not ye after their works: for they say, and do not."* (Matt.)

References

Payne, S. L. (1951) *The Art of Asking Questions.* Princeton: Princeton University Press.

Rossi, P. H., Wright, J.D., & Anderson, A. B. (1983) *Handbook of Survey Research.* New York: Academic Press.

The Status of High School Scheduling in Illinois

Donald G. Hackmann
Iowa State University

Abstract

The purposes of this descriptive study were to determine the types of scheduling models being used in Illinois high schools, identify scheduling trends, and determine reasons for adopting scheduling changes. The principals of Illinois' 635 public high schools were surveyed, and a 100% response rate was achieved after two rounds of questionnaires were mailed and telephone surveys were conducted of remaining nonrespondents. Nearly three-fourths of Illinois high schools utilize a traditional daily-period schedule, but trends indicate increasing numbers of schools are adopting or considering block-of-time models. Principals implementing block-of-time schedules noted a variety of reasons for this change, including the following: increasing student electives, improving the quality of education, implementing varied instructional strategies, increasing time for learning, and improving the building climate.

An issue being debated by many of the nation's high school faculties is the effectiveness of their approaches to scheduling the school day. Some teachers may cherish a uniform, unchanging, daily teaching routine, while others may want instructional creativity they are provided under highly dynamic, flexible scheduling approaches. Some scheduling models may readily lend themselves to a school's staffing and grouping needs but may be extremely rigid, thereby restricting the most effective instructional uses of time. Other models, while providing instructional flexibility for faculty and students, may actually be logistical nightmares for administrators who are attempting to efficiently schedule large numbers of class offerings.

Ubben and Hughes (1997) define the schedule as a "plan to bring together people, materials, and curriculum at a designated time and place for the purpose of instruction. Its basic purpose is to coordinate the requirements laid down by previously reached decisions regarding curriculum, instruction, grouping, and staffing" (p. 216). An effective schedule provides teachers with the ability to make instructional decisions based upon the needs of their students, without being hampered by barriers such as too little or too much time allocated for instruction or rigid inflexibility.

In 1994 the National Education Commission on Time and Learning reported, "The degree to which today's American school is controlled by the dynamics of clock and calendar is surprising" (p. 7). Faced with concerns arising from such reports as *A Nation at Risk* (National Commission on Excellence in Education, 1984) and *Prisoners of Time* (National Education Commission on Time and Learning, 1994), teachers and administrators have gained an increased understanding of the connection between effective time usage and maximized learning opportunities. There is a renewed interest in identifying schedules that effectively facilitate academic growth, and many of the nation's secondary schools have adopted or are actively considering new scheduling models.

Secondary Scheduling Models

Scheduling approaches for high schools can take a variety of forms, depending on the unique needs of each school, and can be divided into the following categories: a) daily period schedules; b) block schedules, including alternating-day models and semester schedules; c) modular/flexible scheduling; and d) combination models. Each model is briefly described below.

Daily Period Schedules

In this scheduling approach, the school day is separated into six, seven, eight, or more equal divisions of time, known as "periods," with each period lasting approximately 42-55 minutes in length. Canady and Rettig (1995b) report that the average daily period length is 51 minutes. Under a daily schedule, students are typically provided 3-5 minutes of time to move from class to class. Frequently, delivery of instruction strictly adheres to departmental classifications: for example, language arts concepts are presented within the English curriculum and science concepts are the property of the science department (Hackmann and Valentine, 1998). The daily period schedule has been the secondary school model of choice for the majority of the 20th century.

The primary advantage of the daily period schedule is that the school routine normally remains unchanging, each day throughout the entire school year. This routine facilitates the acclimation of students into the school environment, as well as providing for ease of lesson preparation for teachers.

The effectiveness of the daily period schedule has been questioned in recent years. Critics assert that the school day is excessively fragmented, that students have little time for in-depth study of subject matter, and that it is difficult for teachers and students to make connections across subject matter lines (Canady and Rettig, 1995a). Additional disad-

vantages include the following: 42-55 minute periods reinforce the use of the lecture method and restrict instructional creativity; with students loads of 160 or more and up to five preparations daily, teachers find it difficult to personalize instruction; and both students and teachers feel the school day moves at an extremely hectic pace (O'Neil, 1995).

Block Schedules

Block-of-time schedules divide the school day into larger timeframes, providing the opportunity to hold class sessions ranging between 85-100 minutes in length, with fewer classes meeting each day and correspondingly fewer class changes (Cawelti, 1994). Larger blocks of time provide flexibility for a variety of learner-centered activities, and teachers are encouraged to reduce their reliance on the lecture method. The Coalition of Essential Schools (Sizer, 1986) advances the metaphors of "student as worker" and "teacher as coach," noting that students should be actively—not passively—involved in the learning process, with teachers guiding their development instead of primarily engaging in direct instruction.

Although block-of-time models can vary greatly in format, two approaches are most commonly used: the alternating-day schedule and the 4x4 semester model. With the alternating-day schedule, also called the A/B schedule, students receive instruction in one-half of their courses on alternate days and complete these courses in one academic year. For example, a student will enroll in classes 1-4 and 5-8 in an alternating-day arrangement (Hackmann, 1995a).

Two variations of this model include: a) the six-block A/B model, in which students enroll in a total of six classes, three each day; and b) the seven-block A/B model, in which students enroll in seven classes, classes 1-3 and 4-6 on alternating days, and the seventh class (usually referred to as a "skinny") on a daily basis. This seventh class usually retains the format of the daily period scheduled classes; for example, if block classes meet for 90 minutes, the duration of the "skinny" is typically 45 minutes.

In the 4x4 semester plan, students complete four classes each semester, for a total of eight courses per year (Edwards, 1995). One variation is the 3x3 plan, in which students enroll in three blocked classes each semester and one "skinny" which meets the entire year, for a total of seven courses. Another variation is the trimester model, also called the Copernican plan (Carroll, 1989), in which students typically enroll in nine blocked classes, three each trimester.

Advocates of the block format assert that, in addition to providing greater instructional flexibility, block-of-time models promote active student participation in learning, improve the quality of teacher-student interaction, reduce students' daily course loads and teachers' daily teaching loads, provide increased support for interdisciplinary instruction, improve the building climate, and promote in-depth instruction (Buckman, King, and Ryan, 1995; Hackmann, 1995a; Wilson, 1995). Schools using block scheduling report numerous positive student outcomes, including decreased disciplinary referrals and suspensions (Buckman et al., 1995; Carroll, 1994; Hackmann, 1995a, O'Neill, 1995), improved attendance (Buckman et al., 1995; Hackmann, 1995a), increased Advanced Placement course enrollments (Edwards, 1995), increased content mastery (Carroll, 1994), and improved grades (Buckman et al., 1995; Edwards, 1995; Stumpf, 1995).

There is relatively little literature citing disadvantages of block scheduling, but anecdotal data indicate that some teachers express concerns over retention of academic content over a two-day period (in the case of alternating-day models) or an entire year (with semester models), adolescents' ability to maintain attention levels during larger blocks, and teachers' abilities to maintain content coverage (Lindsay, 1998). Another disadvantage is the potential need to hire additional staff, since the amount of teacher preparation time frequently increases (Hackmann, 1995b). Some schools do not have the financial resources to absorb this expense. From the students' perspective, block scheduling can be a negative experience when teachers use only the lecture method in the classroom, either because they have not trained in new teaching models or they are unwilling to modify their teaching styles. Canady and Rettig (1995a) note that a minimum of five days of staff development are necessary to provide teachers with effective strategies for teaching in large blocks of time; if financial resources are not available to support this training, they do not advocate a change to block scheduling. Block-of-time models have not been implemented in some schools due to teachers' resistance to change and concerns over the financial costs of increased staff and training.

Modular/Flexible Scheduling

With the modular scheduling approach, the school day is divided into numerous small "modules" of 10, 15, or 20 minutes, and classes are flexibly scheduled according to the number of modules deemed necessary for content instruction (Trump and Baynham, 1961). With this tremendous flexibility, courses could be scheduled in a seemingly infinite variety of formats: some could meet in daily periods; others in alternating-day blocks; or a course might meet in varying lengths of time throughout the year, for example, in both 45- and 90-minute formats, depending on the planned learning activities. Modular schedules appeared on the high school scene in the late 1950s and are generally credited to J. Lloyd Trump, but the scheduling approach began to lose popularity in the early 1970s and is relatively uncommon today. In its *Breaking Ranks* publication, the National Association of Secondary School Principals (NASSP) (1996) recommended that schools develop flexible scheduling models to permit varied instructional uses of time. Consequently, it is likely that, in addition to block

scheduling, school faculties may consider modular scheduling or other methods to provide the flexibility that is lacking in traditional scheduling models.

The primary advantage of highly flexible scheduling types is they "avoid the necessity of giving equal time to unequal subjects" (George and Alexander, 1993, p. 371). Extended time can be scheduled for core academic classes, such as language arts, social studies, math, and science, and elective courses may have time allocations reduced. Flexible models also provide the ability to adapt the time allocation to the planned learning activity. For example, a lecture activity may be scheduled for 30 minutes, while a student experiment, cooperative learning activity, or field trip might be planned for a 120-minute module.

Two major disadvantages of modular/flexible scheduling are apparent. First, classes can meet in a large variety of formats, so schedules can conceivably change each week. Consequently, developing a master schedule of all courses can provide a logistical challenge for administrators and teachers, and course conflicts are highly likely to occur. Second, the highly flexible course schedules means that student schedules will also be highly flexible. On given days, students may have unsupervised time modules between classes that could vary in length between 15 and 75 minutes. Not all students may use this unscheduled time for academic learning, which could create a supervision problem for the faculty.

Combination Models

As high school faculties continue exploring their scheduling options, some have fashioned approaches that include features of more than one model. Combination models usually fall into three arrangements: a) daily period schedule with some blocks, b) daily period schedule with interdisciplinary blocks, and c) combination alternating day/daily period schedules.

Daily period schedules with some blocks are, for all practical purposes, a traditional daily schedule with a small number of double-period blocked classes contained within. Occasionally teachers of laboratory classes, such as biology or chemistry, or college-level Advanced Placement classes may request that their courses have additional time scheduled to allow for experiments or an increased amount of academic content.

Daily period schedules with interdisciplinary blocks allow teachers to work as an interdisciplinary team, usually at the freshman level (Hackmann and Waters, 1998). The NASSP (1996) recommends that high school faculties should integrate the curriculum "to the extent possible" so that students can make connections between the disciplines (p. 11). Building upon the interdisciplinary teaming concept long advocated by middle school educators (George and Alexander, 1993), some schools schedule the freshman subjects of language arts, social studies, mathematics, and science in one large interdisciplinary block, empowering these shared teachers to divide their block into any appropriate timeframes for instructing their classes. The remaining courses are scheduled in a daily period format.

Combination alternating day/daily period schedules include features of both models in the weekly schedule. For example, an eight-class scheduling model might be configured with four 90-minute classes on Wednesdays (periods 1-4) and Thursdays (periods 5-8), but schedule all eight classes in a daily period format on Mondays, Tuesdays, and Fridays.

The perceived advantage of combination models is they allow faculties to select the best features of each scheduling option: they permit occasional larger blocks of time for instructional creativity, provide for variety within the school day, and permit a degree of flexibility not available in the more rigid approaches. One disadvantage is that a combination model may be selected as a compromise stance between two competing models identified by two factions within a faculty and, as such, may actually not fully satisfy the needs of any teachers.

A nationwide survey of 3,380 high schools (Cawelti, 1994) disclosed that 23% of responding schools were either fully or partially utilizing block scheduling in 1994, and another 15.4% planned to implement this scheduling approach the following year. Cawelti's research results predict that at least 38% of the nation's high schools—if not more—should now be utilizing alternatives to the traditional daily schedule. Rettig and Canady (1996) estimated that more than 50% should either be using or considering some form of block scheduling during the 1996-1997 school year.

This article describes a descriptive study conducted to determine the scheduling options being implemented in the state of Illinois. The study addressed the following research questions: a) What types of scheduling models are employed in Illinois public high schools, and with what degree of frequency? b) What trends are occurring with respect to the various scheduling types? and c) What reasons do principals provide for adopting, or choosing not to adopt, changes in their scheduling models?

Method and Procedures

In January 1997 principals of the 635 public high schools in the state of Illinois were mailed questionnaires in an effort to determine the scheduling models presently being used within the state and to examine scheduling trends. A cover letter was included with the questionnaire, outlining the research questions and requesting participation in the study. The questionnaire consisted of 42 short-answer and open-ended questions and was divided into three sections. The first section, consisting of eight questions, queried such school data as number of faculty, percentage of students receiving free lunches, number of Advanced Placement courses, and the scheduling of interdisciplinary teaming within the school. In the second section, containing 16 ques-

tions, respondents were asked to describe in detail their scheduling models being used in the 1996–1997 academic year, including the number and length of classes, starting and ending times, and use of advisory periods. The principals were also asked the year they implemented their scheduling model, if they intended to change models within the next year, and if they had a plan to evaluate the effectiveness of their current schedule. The third section, consisting of 18 questions, was completed by respondents whose schools had implemented new scheduling models in 1992-1993 or later, or planned to implement a new schedule in the 1997-1998 academic year. In addition to describing their new scheduling models in detail, respondents were also asked to explain their reasons for adopting new models, explain the activities in which their faculties engaged during the change process, and to note the involvement and influence various groups had in the decision to make the scheduling change. Finally, all respondents were given an opportunity to share any additional comments concerning their school's scheduling issues. Usable responses were received from 210 principals, for a 33.1% response rate.

In April 1997 a cover letter and shortened questionnaire were mailed to non-respondents. The survey was shortened to five questions in an effort to reduce the amount of time required for completion and to improve the response rate. The short questionnaire asked principals to describe their current schedule and year of implementation, note if they intended to adopt a new schedule in the 1997-1998 school year, and to describe the new schedule. Principals were also asked to provide any additional comments concerning their schedules. An additional 292 surveys were returned, for a combined total of 402 responses (63.3%).

In May 1997 telephone interviews were conducted with administrators of the remaining 133 schools; they were asked the five questions contained in the shorted survey. Finally, 10 telephone interviews were conducted in August 1997 to obtain information from schools that had not yet selected their scheduling model for the 1997-1998 academic year. As a result of these four data collection waves, responses were obtained from a total of 635 schools (100.0%), and the data reported encompasses the entire population of Illinois public high schools.

A two-phase data analysis procedure was undertaken after all data were collected. Responses to Questions 1-4 (current scheduling model, year of implementation, intent to adopt a new model, and description of 1997-1998 schedule) were entered into the computer data file, Microsoft Excel 97. Descriptive statistics including central tendency measures, totals, and response proportions were calculated for each of these items. Data for the open-ended questions, which related to adoption or rejection of scheduling approaches, were examined separately by the researcher, who subsequently developed a coding strategy based on thematic similarities among responses. Occurrence of coded passages were summed across cases to convey proportional representation of the themes among respondents.

Results

Scheduling Models Employed in Illinois

Respondents provided data concerning the scheduling models used during the current year (1996-1997) and noted the models their schools would utilize for the following year (1997-1998). Data are, therefore, reported for both of these academic years.

Illinois principals reported that their high schools have adopted a variety of approaches to configuring the instructional day, but over 94% of schools utilize either daily period or block schedules. In 1997-1998 the three most commonly used scheduling configurations used in Illinois high schools, by order of preference, were the seven-period daily (209 schools; 32.9%), eight-period daily (188; 29.6%), and eight-block alternating-day schedule (108; 17.0%). A few schools (25 in 1996-1997, 33 in 1997-1998) are using modular schedules or daily period models that contain some larger blocks. Some schools have implemented interdisciplinary blocks for freshman students (9 in 1996-1997, 13 in 1997-1998) that were recommended in *Breaking Ranks* (NASSP, 1996). Table 1 categorizes the models used during the 1996-1997 and 1997-1998 academic years into various forms of daily period schedules, alternating-day models, semester block plans, modular/flexible models, and daily schedules that include block formats.

Table 2 indicates that, when grouped by scheduling type, daily schedules were the preferred model in Illinois but decreased somewhat over the two years. In 1996-1997, 497 of 635 schools (78.3%) used daily period scheduling, and 462

Table 1
Scheduling in Illinois Public High Schools

Scheduling Model	1996-1997 Number of Schools (Percentage)	1997-1998 Number of Schools (Percentage)
Six-period daily	47 (7.4%)	42 (6.6%)
Seven-period daily	235 (37.0%)	209 (32.9%)
Eight-period daily	190 (29.9%)	188 (29.6%)
Nine-period daily	21 (3.3%)	21 (3.3%)
Ten-period daily	4 (0.6%)	2 (0.3%)
Six-block alternating day	4 (0.6%)	4 (0.6%)
Seven-block alternating day	1 (0.2%)	1 (0.2%)
Eight-block alternating day	94 (14.8%)	108 (17.0%)
3 x 3 semester block	1 (0.2%)	1 (0.2%)
4 x 4 semester block	8 (1.3%)	17 (2.7%)
Combination alternating day blocks/daily period	5 (0.8%)	9 (1.4%)
Modular/flexible schedule	10 (1.6%)	11 (1.7%)
Daily schedule with some blocks	6 (0.9%)	9 (1.4%)
Daily schedule with freshman/interdisciplinary blocks	9 (1.4%)	13 (2.0%)
TOTAL	**635** (100.0%)	**635** (100.0%)

Table 2
Scheduling Totals in Illinois Public High Schools, by Model

Scheduling Model	1996-1997 Number of Schools (Percentage)	1997-1998 Number of Schools (Percentage)
Daily period schedule	497 (78.3%)	462 (72.8%)
Block schedule (predominately)	113 (17.8%)	140 (22.0%)
Modular/flexible schedule	10 (1.6%)	11 (1.7%)
Daily schedule (predominately) with some blocks	15 (2.4%)	22 (3.5%)
TOTAL	**635** (100.0%)	**635** (100.0%)

schools (72.8%) in 1997-1998. Block schedules increased slightly over the two academic years. In 1996-1997, 113 schools (17.8%) used block scheduling, increasing to 140 schools (22.0%) in 1997-1998. Modular/flexible scheduling was used in 10 schools (1.6%) in 1996-1997 and 11 schools (1.7%) the following year. Combination models, primarily daily period schedules with some double-period blocks, were used in 15 schools (2.4%) in 1996-1997 and 22 schools (3.5%) the next year.

Table 3
Illinois Public High School Scheduling Trends by Year of Implementation of New Schedule

	Number	Percent
Implemented in 1993-1994	32	5.0%
Implemented in 1994-1995	52	8.2%
Implemented in 1995-1996	51	8.0%
Implemented in 1996-1997	55	8.7%
Implemented in 1997-98	63	9.9%
Considering implementation in 1998-99	119	18.7%

Scheduling Trends

The number of schools annually implementing changes in their schedules within the each of the past five years demonstrated a trend of gradual increase, from 32 schools in the 1993-1994 academic year to 63 schools in 1997-1998. (See Table 3.) This number was greater for the 1998-1999 school year: 119 principals reported their faculties were considering scheduling changes for 1998-1999. Forty-eight of 53 within this group (90.6%) were considering block-of-time scheduling: 19 noted "some form of block schedule," 11 indicated an alternating-day model, 10 described 4x4 semester schedules, six listed combination block/daily models, one noted a five-block semester model, and one listed a trimester Copernican schedule.

Reasons for Scheduling Changes

Principals who had adopted scheduling changes within the past five years or who were making changes for the 1997-1998 year were asked to respond to an open-ended question in which they listed the reasons their faculties chose to make schedule changes. This question was included only in the original questionnaire, of which there were 210 respondents. Principals in 71 of the 210 schools (33.8%) indicated they had or were making scheduling changes. Several respondents provided multiple responses to this question. Responses were separated by type of schedule implemented (block or daily period), tallied, and then categorized into appropriate groupings. Fifty-nine schools had adopted or were adopting block scheduling models, and 51 provided responses to the question. Each of these schools had switched from a daily period schedule. Twelve schools had switched from one daily period schedule to another (for example, moving from seven periods to eight periods), and six provided responses to the question.

As noted in Table 4, the schools adopting block scheduling formats listed numerous reasons for this change, including the following: providing course flexibility for students (24 responses), improving the quality of the students' educational experiences (22), improving instructional strategies (20), providing increased time for learning (19), and improving the school climate (10), improving the curriculum (7), meeting staffing needs (7), modeling themselves

Table 4
Reasons for Making Scheduling Changes

Reason	Block Schedule (N = 59)	Daily Schedule (N = 12)
Course Choices (increased student electives/courses, eliminate studyhalls)	24	4
Quality Education (deeper learning, higher expectations, improved achievement, graduation rate, increased graduation requirements, better prepare/meet student needs, improve success of middle range or bottom 1/3)	22	0
Instructional strategies (improved strategies, hands-on, innovation, flexibility)	20	0
Time (increased time for learning, time on task, time for homework, individualized attention, practice)	19	2
School Climate (improve teacher/student relationships, discipline, responsibility, organization, attendance, reduce number of daily classes)	10	0
Curriculum (revision, integrated learning, technology, changes in assessment practices)	7	0
Staff (utilization/efficiency/needs, forced change)	7	0
Success of Others (through visits, conference attendance, school consolidation/deactivation)	6	0
Enrollment Growth	0	3
Administrative Decree	3	0
No response	8	6

Note: Totals exceed 100% due to multiple responses.

after the success of other schools (6), and change resulting from an administrative decree of either the principal or superintendent (3).

Schools making slight modifications to their daily period models (adding or eliminating a period) indicated three reasons: providing course flexibility for students (4), responding to enrollment growth (3), and providing increasing time for learning (2).

Twelve principals of high schools using daily period schedules provided additional comments concerning their choice of models. Each respondent provided reasons why their schools were not adopting block scheduling. These answers were grouped into the following categories: the faculty rejected the block scheduling concept (4 responses), the school was unable to hire the additional staff needed for a block schedule (3), the faculties were awaiting results concerning other schools' experiences with block scheduling (2), their daily period schedules were successful (2), there was no reason to change (1), and the faculty was concerned how block scheduling might affect students who transferred into or out of the school in mid-year (1).

Conclusions and Discussion

This study of Illinois public high school scheduling models discloses that nearly three-fourths of the state's schools operated under daily period schedules during the 1997-1998 school year. Fewer than 27% were using scheduling models that incorporate block scheduling components; this percentage is below the 38% reported in Cawelti's (1994) survey and Canady and Rettig's (1995b) projected 50%. This finding within the state of Illinois should not be generalized to suggest that the national projections of Cawelti and Canady and Rettig were overestimated, because other states exceed the 50% mark. For example, in 1995-1996, 55.8% of North Carolina secondary schools were using block scheduling, up from 1.62% in 1992-1993 (Department of Public Instruction, 1997). This study likely indicates that Illinois educators are taking a more cautious approach with the shift to block-of-time scheduling than educators in other states.

Approximately 5-10% of Illinois high schools implement modifications to their schedules annually, and trends indicate that this percentage will remain constant in the immediate future. A small number of schools have adopted models incorporating features of both daily period and block scheduling. Based upon principals' responses, it is likely that schools will continue to experiment with variations of alternating day and semester block models tailored to solving perceived problems that are unique to individual schools.

Schools making adjustments to daily period schedules (for example, changing from six periods to seven periods) stated three reasons for their changes: to increase student course choices, respond to enrollment growth, and/or to provide increased instructional time. In contrast, school personnel adopting block scheduling provided a variety of reasons, primarily to provide increase student course choices, improve the quality of education, improve instructional strategies, and increase time for learning.

This study has implications for high school faculties, both within the state of Illinois and in other states. The following recommendations are presented as faculties consider changes in their scheduling models:

Teachers should be directly involved in all discussions concerning scheduling modifications. Principals of schools rejecting block scheduling consistently noted their faculties were not ready to teach in larger timeframes and that they were waiting to determine if other schools were successful with this approach. Three schools indicated their move to block scheduling was a unilateral decision of the superintendent or principal, even though the faculties did not support the new models. Fullan (1993) notes that change will not be sustained and institutionalized without the involvement and buy-in of those affected. Therefore, it is likely that teachers will not be committed to modifying their instructional strategies when the change is forced upon them, and changes in instructional practices are critical for the successful implementation of block scheduling (Canady and Rettig, 1995a).

Faculties contemplating scheduling models that deviate significantly from established models should be aware of potential problems they may create for transfer students. How can students who transfer during the academic year be effectively scheduled into schools whose scheduling models are dramatically different from their former schools? The nation's schools do not exist in vacuums and, with the high mobility rate of our population, problems will occur. Consider, for example, a student enrolled in four courses in a 4x4 semester school, who transfers in March to an eight-block alternating-day school, or a school with eight daily periods. Or, the student enrolled in a school with a seven-period daily model who transfers in October to a 3x3-semester school, or a school with a modular schedule. Can school personnel smoothly transition transfer students into their schools in a manner that ensures that their academic development is not compromised? No solutions are proposed here, but it is incumbent upon all high school educators to examine this issue in context with any proposed scheduling change. In fact, personnel in schools that choose to retain their traditional schedules should also discuss this problem, because they will soon be required to accommodate students who transfer from nontraditional scheduling models, if they have not already been faced with this issue.

Faculties should be aware that the selection of scheduling models might effect on student achievement, either positively or negatively. Relatively little research has been conducted to date on the effects the various scheduling models may have on such measures of student achievement as standardized test scores, Advanced Placement test scores, and

college entrance examinations. Although principals noted a variety of reasons for changing their scheduling models, only a few discussed issues related to student achievement. For example, if a scheduling change results in less time for instruction, academic content coverage may decrease and student learning may suffer. Conversely, increased time allocations may permit greater depth of content coverage. Furthermore, changing the number of class periods may affect students' ability to enroll in and complete their preferred course selections during their high school careers.

References

Department of Public Instruction. (1997, May). *Block scheduled high school achievement: Part II*. (Evaluation brief). Raleigh, NC: State Board of Education.

Buckman, D. C., King, B. B., and Ryan, S. (1995). Block scheduling: A means to improve school climate, *NASSP Bulletin*, *79*(571), 9-18.

Canady, R. L., and Rettig, M. D. (1995a). *Block scheduling: A catalyst for change in high school*. Princeton, NJ: Eye on Education.

Canady, R. L., and Rettig, M. D. (1995b). The power of innovative scheduling. *Educational Leadership*, *53*(3), 4-10.

Carroll, J. M. (1989). *The Copernican plan: Restructuring the American high school*. Andover, MA: The Regional Laboratory for Educational Improvement of the Northeast and Islands.

Carroll, J. M. (1994). *The Copernican plan evaluated: The evolution of a revolution*. *Phi Delta Kappan*, *76*, 104-113.

Cawelti, G. (1994). *High school restructuring: A national study*. Arlington, VA: Educational Research Service.

Edwards, C. M., Jr. (1995). The 4x4 plan. *Educational Leadership*, *53*(3), 16-19.

Fullan, M. (1993). *Change forces: Probing the depths of educational reform*. London: The Falmer Press.

George, P. S., and Alexander, W. M. (1993). *The exemplary middle school* (2nd ed.). Orlando, FL: Harcourt Brace.

Hackmann, D. G. (1995a). Improving the middle school climate: Alternating-day block schedule. *Schools in the Middle*, *5*(1), 28-34.

Hackmann, D. G. (1995b). Ten guidelines for implementing block scheduling. *Educational Leadership*, *53*(3), 28-32.

Hackmann, D. G., and Valentine, J. W. (1998). Designing an effective middle level schedule. *Middle School Journal*, *29*(5), 3-13.

Hackmann, D. G., and Waters, D. W. (1998). Breaking away from tradition: The Farmington High School restructuring experience. *NASSP Bulletin*, *82*(596), 83-92.

Lindsay, J. (1998). *The case against block scheduling* [Online]. Available: http://www.athenet.net/~jlindsay/Block.shtml

National Association of Secondary School Principals. (1996). *Breaking ranks: Changing an American institution*. Reston, VA: Author.

National Commission on Excellence in Education. (1984). *A Nation at risk*. Cambridge, MA: USA Research.

National Educational Commission on Time and Learning. (1994). *Prisoners of time: Report of the National Educational Commission on Time and Learning*. Washington, DC: U.S. Government Printing Office.

O'Neil, J. (1995). Finding time to learn. *Educational Leadership*, *53*(3), 11-15.

Rettig, M. D., and Canady, R. L. (1996). All around the block: The benefits and challenges of a non-traditional school schedule. *The School Administrator*, *8*(53), 8-12.

Sizer, T. R. (1986). Rebuilding: First steps by the Coalition of Essential Schools. *Phi Delta Kappan*, *68*, 38-42.

Stumpf, T. (1995). A Colorado school's un-rocky road to trimesters. *Educational Leadership*, *53*(3), 20-22.

Trump, J. L., and Baynham, D. (1961). *Focus on change: Guide to better schools*. Chicago: Rand McNally.

Ubben, G. C., and Hughes, L. W. (1997). *The principal: Creative leadership for effective schools* (3rd ed.). Needham Heights, MA: Allyn and Bacon.

Wilson, C. (1995). The 4:4 block system: A workable alternative. *NASSP Bulletin*, *79*(571), 63-65.

Time Spent on Higher-Order Tasks in Two Teacher-Apprentice Options

Elizabeth A. Wilkins-Canter
Audrey T. Edwards
Eastern Illinois University

Abstract

This study investigates the process of assisting a teacher prior to student teaching. Teacher candidates in two certification programs kept a log documenting the tasks they did during their on-site experiences and the time spent on eight types of tasks. A one-way analysis for repeated measures showed that, in both programs, candidates spent a majority of their time on clerical tasks, despite the value they saw in interpersonal contact. A t *test for independent measures revealed no significant difference between constant and varied placements in time spent on teaching a whole class and on other highly valued tasks.*

Early clinical experiences are a critical component of a teacher candidate's preparation for the capstone field experience and for entry into the profession. Since the early 1980s, reform movements have generally called for improvement in the quality and quantity of field experiences for teacher candidates (Berliner, 1985). The Holmes Group (1986) and the Carnegie Forum on Education (1986) emphasized the need to develop more extensive and better clinical experiences as part of teacher preparation programs. More recently, the National Commission on Teaching and America's Future (1996) released a report containing numerous recommendations for redesigning teacher education with specific emphasis on teacher preparation and extended programs of study. Such reform movements have impelled legislatures and school boards of education to require more school experiences of teacher candidates and have promoted the establishment of professional development schools (Guyton and McIntyre, 1990).

It is widely accepted by teacher educators that the quality of early field experiences often differs from one candidate to another because of length, tasks required during classroom visits, and type of placement. Thomson, Beacham, and Misulis (1992) found that in traditional programs, longer experiences helped teacher candidates develop confidence and self-esteem as well as heighten their awareness of the profession. Longer field experiences also provide teacher candidates with a better understanding of teachers' actions, curriculum, and student behavior. Fullan (1985) reported that by extending field experiences, teacher candidates were better able to adjust to the routines of teaching. The above findings, however, rely on self-reports of attitude.

Despite these positive findings about longer field experience programs, some teacher educators have argued that what occurs during the field experience is more important than the length of the experience (McIntyre, 1983; McIntyre, Byrd, and Foxx, 1996; Ziechner, 1980). Research studies taken from traditional teacher education programs indicate that teacher candidates in early field experiences are typically engaged in a very limited and narrow range of classroom activities (Feiman-Nemser and Buchmann, 1986; Howey, 1986; Killian and McIntyre, 1986; Tabachnick, Popkewitz, and Zeichner, 1979-1980). Howey (1986) found that "many of the experiences that [teacher candidates] have in schools lie more in the direction of largely unchallenging pedestrian activities than in well-conceived activities where prospective teachers have opportunities to inquire, to experiment, and to reflect on the subtleties and complexities of the classroom" (p. 174). Would the same be true in longer, collaborative field experiences as proposed by reform efforts?

In addition, several studies of traditional programs have supported the placement of teacher candidates in varied settings during early field experiences. Garibaldi (1992) advocated that teacher candidates should be exposed to a variety of students and schools as early as their first semester in college; furthermore, they should be assigned to different schools and classrooms every semester of their program. The purpose of increasing the number and variety of sites is to provide opportunities to investigate, reflect, and solve problems in multiple communities rather than in limited contexts (Cinnamond and Zimpher, 1990; Sedlak, 1987).

In all, some researchers indicate that in traditional programs, longer field experiences improve candidates' attitudes toward teaching and feelings of competence in the classroom; by contrast, others state that even in relatively long placements the experience itself is typically low in quality. Still others advocate variety of placement. The research literature is lacking, however, in studies of newer teacher education programs that offer very extensive field experiences. In particular, research on the new field experiences does not clearly show what tasks candidates are doing and whether those tasks vary with length or variety of placement. The present study addresses proposed changes in amount and kind of early field experience. Specifically, this quantitative study tests whether increases in length and variety of placement, as proposed by reform movements, result in better-quality clinical experiences as measured by time spent on whole-class teaching and other tasks valued by both candidates and cooperating teachers.

The present research is part of a longitudinal study that has been conducted at a university in the Midwest to assess changes brought about by reform. The researchers have been studying their university education courses that have a long early clinical experience component and are taught on site in three schools through a collaborative model. Teacher candidates in these courses have reported that, while assisting a teacher in the classroom, they complete a variety of tasks ranging from photocopying to teaching a class. They ranked teaching a class as the most valuable experience, followed by tutoring or assisting a small group, assisting the class with seatwork/lab work, grading student work, preparing teaching materials (other than

photocopying), testing, observing, and copying. Cooperating teachers agreed that whole-class teaching and tutoring were the most valuable tasks and did not differ significantly in rating the other tasks. In addition, constancy of placement, as opposed to variety, was associated with greater frequency of teaching a whole class, the most highly valued task of those studied. However, the study did not measure whether the teaching was done for any appreciable length of time.

Based on the above findings, a second study was conducted to learn whether, in these field-based programs, time spent on teaching a whole class and on other highly valued tasks increased as the semester progressed. The researchers also investigated whether time allocation was affected by constancy of classroom placement. Research suggests that a disproportionate amount of time would be spent on tasks that the candidates found low in value but that time on highly valued tasks would increase as teacher candidates spent more time in the classroom. Despite Garibaldi's advocacy of variety in placement, the authors believed that placing a given candidate with the same public school teacher, rather than shifting teachers, might result in more time spent on valued tasks.

Method

The study was conducted in three schools associated with a college of education at a university in the Midwest. In response to calls for reform at both the state and national levels, the university has actively engaged in collaborating with public schools and implementing extensive field experiences prior to student teaching. Like other institutions attempting to move away from traditional models of teacher preparation, the university has designed and implemented two programs where instruction and field experience take place concurrently in a public-school setting.

Subjects were forty-nine self-selected teacher candidates, all at the junior, senior, or postbaccalaureate level, seeking certification at the secondary level; each chose to take part in one of the on-site programs rather than a traditional, campus-based section. Randomly assigning the candidates to one of the two programs was not possible because university rules allow students a choice. However, a prior survey by the authors indicated that students were basing their choices primarily on scheduling constraints rather than on preferences for cooperating teachers or even knowledge of the programs. Thus there was reason to believe that the two groups did not differ substantially.

The first of these two programs was part of a professional development school (PDS) where teacher candidates took educational psychology and a general instructional methods course taught in a block format in either a middle school or high school setting. The teacher candidates assisted a cooperating teacher in their specialty area for one class period every day throughout the semester; during the subsequent class period, the PDS students met with university faculty to learn theoretical concepts to be applied in the classroom. In the second program, each student worked with the same cooperating teacher for a six-hour block of time in completing assigned instructional modules. University faculty and teacher candidates in the module-based program met once a week for classroom instruction to discuss theory and its application in the classroom. In both programs, cooperating teachers were invited to become actively involved in teaching information alongside university faculty, thereby encouraging collaboration.

At the beginning of the semester, university faculty in both programs encouraged cooperating teachers to assign whole-class teaching and other high-level tasks such as developing an evaluation instrument in relationship to a unit or units being taught, planning for and teaching a small group of students, or assisting a member of the faculty in an extracurricular activity.

During the spring 1996 semester, the teacher candidates kept a "Time and Task Log" documenting the tasks they did during their on-site experiences and the time spent on each task to the nearest quarter hour. The logs requested data in eight main categories: (a) whole-class teaching, (b) tutoring or assisting a small group, (c) assisting the class with seatwork/lab work, (d) grading student work, (e) preparing teaching materials (other than photocopying), (f) testing, (g) observing, and (h) copying. Figure 1 is a representative example of the "Time and Task Log."

Four statistical analyses were performed on the logs. Within the PDS group, the proportion of time spent on each of the eight classroom tasks was studied, comparing three time periods within a semester to determine whether candidates were allowed to do more whole-class teaching and other valued tasks.

Name _____ School where you assisted _____
Week you assisted_____

Record the time you spend each day in the following activities.
Give all time to the nearest ¼ hour.

CIRCLE APPROPRIATE WEEK NUMBER: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16

TASKS	M	T	W	R	F
Photocopying or collating					
Preparing teaching materials (bulletin boards, graphs, models, etc.)					
Grading student work					
Assisting 1 person/small group					
Assisting class with work (seatwork, etc.)					
Giving tests to class					
Teaching the whole class					
Observing (doing no other tasks at the time)					
Other (specify):					

Figure 1. Assisting the teacher: Time and task log

The mean rating of the eight classroom tasks was also studied to determine whether candidates increasingly engaged in more valued tasks as they gained classroom experience. The mean ratings were computed as follows: Based on a previous survey of tasks that cooperating teachers and teacher candidates value, each task had a given rank (teaching a whole class, for example, had a rank of 1). For a given task, its rank was multiplied by the number of hours each subject spent on that task; that figure was then divided by total hours spent on all tasks; next, the mean for all subjects was determined. This composite rating represents the overall quality of time spent: the lower the number, the higher the overall quality. Thus, the mean rank would be lower at the end of the semester if candidates increasingly performed whole-class teaching or other highly-valued tasks. For each of the above comparisons, a one-way analysis of variance for repeated measures was performed to determine significance.

The PDS group was also compared to the module-based group to determine whether constancy of placement (the PDS group), as opposed to variety (the module-based group), was associated with more time spent on highly-valued tasks. Again, the proportion of time spent on whole-class teaching and the mean rating of the eight classroom tasks was studied. For each of these

Table 1
PDS Group: Proportion of Time Spent on Each of Eight Classroom Tasks During Three Time Periods

Tasks	1st third of semester	2nd third of semester	3rd third of semester		
1. Whole-class teaching	X=.051	X=.089	X=.087	f=1.24	p=.304
2. Assisting small groups	.091	.119	.103	f=.47	p=.631
3. Assisting class	.234	.113	.183	f=4.24	p=.023
4. Grading papers	.125	.075	.117	f=1.44	p=.252
5. Preparing teaching materials	.090	.104	.072	f=.42	p=.662
6. Giving tests	.053	.025	.065	f=1.19	p=.318
7. Observing	.303	.456	.393	f=3.79	p=.033
8. Photocopying	.023	.019	.011	f=1.36	p=.271

Table 2
*PDS Group: Mean Rating of Eight Classroom Tasks During Three Time Periods***

	1st third of semester	2nd third of semester	3rd third of semester		
PDS Program	X = 4.62	X = 4.93	X = 4.71	f=1.15	p=.482

N=17

**lower score indicates more of the most valued tasks

two comparisons, a *t* test for independent measures was performed. An alpha level of .05 was used for all statistical tests.

Results

Within the PDS group, the proportion of time spent on each of the eight classroom tasks was studied. Three time periods were compared within the semester to determine whether candidates were allowed to do more whole-class teaching and other valued tasks as they gained classroom experience. A one-way analysis of variance for repeated measures showed no significant differences among the three time periods for most of the tasks; however, significant differences were found for "assisting the class" and "observing." The proportion of whole-class teaching, the most valued task, was uniformly low.

The mean rating of the eight classroom tasks was also studied to determine whether candidates increasingly engaged in more valued tasks as they gained classroom experience. A one-way analysis of variance for repeated measures showed no significant differences among the three time periods. The three mean ratings showed uniformly frequent occurrence of low-valued tasks.

The PDS group was also compared to the module-based group to determine whether constancy of placement (the PDS group), as opposed to variety (the module-based group), was associated with a higher proportion of time spent on highly-valued tasks. Again studying the proportion of time spent on whole-class teaching and the mean rating of the eight classroom tasks, a *t* test for independent measures was performed; no significant differences were found between the two groups. The two programs were uniformly low in proportion of whole-class teaching, and mean ratings showed uniformly frequent occurrence of low-valued tasks. (See Table 3)

Discussion

The findings from this study are of particular interest when considered in relation to reform proposals advocating extended, collaborative field experiences. In agreement with previous studies of traditional programs, a large percentage of time was spent on tasks that the candidates found low in value. This finding does not reflect a change toward better-quality field experiences; unfortunately, it echoes past research from traditional teacher education programs where candidates in early field experiences engaged in a very limited and narrow range of classroom activities (Howey, 1986; Killian and McIntyre, 1986; Tabachnick, Popkewitz, and Zeichner, 1979-1980). Although low-valued tasks are part of teachers' daily responsibilities and need to be experienced, the opportunity to involve teacher candidates in tasks that are valued more highly would seem greater if candidates spent extended periods of time in the classroom, as advocated by reform efforts. Would teacher candidates, though, spend more of their time on highly valued tasks if given a more extended experience? The findings from this study raise some doubt.

The proportion of time on highly valued tasks did not increase as the teacher candidates spent more time in the classroom. Two concerns arise from this finding. First, teacher candidates in the constant placement (the PDS group) did not spend a significantly greater proportion of time on whole-class teaching and on other highly valued tasks at the end of the se-

Table 3
Comparison Between PDS Group and Module-Based Group: Proportion of Time Spent on Whole-Class Teaching

PDS group	Module-based group
N=17 X=.05	N=32 X=.05

t= -.12 df=47 p=.905

*Comparison Between PDS Group and Module-Based Group: Mean Rating of Eight Classroom Tasks***

PDS group	Module-based group
N=17 X = 4.62	N=32 X = 4.27

t= -1.21 df=47 p=.231

**lower score indicates more of the most valued tasks

mester than at the beginning. (When comparing the three time periods for each of the eight tasks, the significant differences found for "assisting the class" and "observing" might indicate that as candidates had less opportunity to assist the class, they began to observe more.) Second, despite no change in the proportion of time, the total time that candidates in constant placements spent in the classroom did, as a rule, provide a high total time at valued tasks as well as opportunity to develop confidence and awareness of the routines of teaching. However, some of the candidates never engaged in the most valued tasks during their entire extended experiences. These two issues reinforce the notion that often what occurs during the field experience is more important than the length of the experience (McIntyre, 1983; McIntyre, Byrd, and Foxx, 1996; Ziechner, 1980).

Placing a given candidate with the same cooperating teacher, rather than shifting teachers, did not result in a greater proportion of time spent on valued tasks. The overall lack of significant differences across time (within the PDS group) and across groups (PDS and module-based) suggests that task assignments are less influenced by experience and program design than by the cooperating teacher's preferences and the teacher candidate's perceived abilities. The selection and matching of cooperating teachers to teacher candidates, therefore, takes on additional importance, especially for those in constant placements (the PDS group).

This study suggests several recommendations that may help teacher education programs to change the design of their early field experiences in response to reform. First, placement with the cooperating teacher is an important aspect. University faculty need to work closely with building administrators and cooperating teachers to insure the best selection of mentor teachers. In cases where building administrators typically make the decision about placements, emphasis should be placed on establishing collaborative partnerships between schools districts and colleges of education. Second, supervision becomes extremely important when placing teacher candidates with cooperating teachers for a longer early field experience. Through closer supervision, university faculty have greater opportunity to interact with cooperating teachers to resolve concerns, teach side-by-side when additional mentoring is needed, and encourage teacher candidates to be engaged in more highly-valued tasks. Third, communication with the cooperating teacher about what is expected of the teacher candidate is vital. Information should be shared both orally and in writing as to the expectations for each early field experience. Fourth, teacher education programs should consider a combination of constant and varied placements throughout a teacher candidate's preparation for the capstone field experience. Since neither approach appeared to out-perform the other in this study, a combination of both might be best. Additional research is needed to study the possibilities of sequencing placements with variety and constancy so as to give teacher candidates the best of both. In either case, however, the greatest challenge will be to persuade cooperating teachers to give candidates more opportunities to engage in complex, responsible tasks.

References

Berliner, D. (1985). Laboratory settings and the study of teacher education. *Journal of Teacher Education, 36*(6), 2-8.

Carnegie Corporation of New York. (1986). *A nation prepared: Teachers for the 21st century*. New York: Carnegie Corporation of New York. (ERIC Document Reproduction Service No. ED 268 120)

Cinnamond, J., and Zimpher, N. (1990). Reflectivity as a function of community. In R. Clift, W. R. Houston, and M. Pugach (Eds.), *Encouraging reflective practice in education: An analysis of issues and programs* (pp. 57-72). New York: Teachers College Press.

Feiman-Nemser, S., and Buchmann, M. (1986). Pitfalls of experience in teacher education. In J. Raths and L. Katz (Eds.), *Advances in teacher education* (Vol. 2, pp. 61-73). Norwood, NJ: Ablex Publishing.

Fullan, M. (1985). Integrating theory and practice. In D. Hopkins and K. Reid (Eds.), *Rethinking teacher education* (pp. 195-211). London: Croom Helm.

Garibaldi, A. (1992). Preparing teachers. In M. Dilworth (Ed.), *Diversity in teacher education* (pp. 23-39). San Francisco: Jossey-Bass.

Guyton, E., and McIntyre, D. J. (1990). Student teaching and school experiences. In W. R. Houston (Ed.), *Handbook of research on teacher education* (pp. 514-534). New York: Macmillan.

The Holmes Group. (1986). *Tomorrow's teachers: A report of the Holmes Group*. East Lansing, MI: Author.

Howey, K. (1986). The next generation of teacher preparation programs. In T. Lasley (Ed.), *The dynamics of change in teacher education* (pp. 161-185). Washington, DC: American Association of Colleges for Teacher Education.

Killian, J., and McIntyre, J. (1986). Quality in early field experiences: A product of grade level and cooperating teachers' training. *Teaching and Teacher Education, 2*(4), 367-376.

McIntyre, D. J. (1983). *Field experiences in teacher education: From student to teacher*. Washington, DC: Foundation for Excellence in Teacher Education and ERIC Clearinghouse on Teacher Education.

McIntyre, D. J., Byrd, D. M., and Foxx, S. M. (1996). Field and laboratory experiences. In J. Sikula (Ed.), *Handbook of research on teacher education* (pp. 171-193). New York: Macmillan.

National Commission on Teaching and America's Future (1996). *What matters most: Teaching for America's future*. Washington, DC.

Sedlak, M. (1987). Tomorrow's teachers: The essential arguments of the report. In M. Sedlak (Ed.), *Reforming teacher education: The impact of the Holmes Group Report* (pp. 4-15). New York: Teachers College Press.

Tabachnick, B. R., Popkewitz, T., and Zeichner, K. M. (1979-1980). Teacher education and professional perspectives of student teachers. *Interchange on Educational Policy, 10*(40), 12-29.

Thomson, S., Beacham, B., and Misulis, K. (1992). A university and public school collaborative approach to preparing elementary teachers. *The Teacher Educator, 28*(2), 46-51.

Zeichner, K. (1980). Myths and realities: Field-based experiences in preservice teacher education. *Journal of Teacher Education, 31*(6), 45-49, 51-55.

Research Alive

The Relationship between Culture and Cognitive Style: A Review of the Evidence and Some Reflections for the Classroom

Joan Thrower Timm
University of Wisconsin Oshkosh

Abstract

This column summarizes factors in cultural experience that affect approaches to learning and problem solving. It reviews the evidence over the past thirty years on cognitive style differences in culturally diverse groups in the United States including Asian, African, Latino, and Native Americans and reports recent research findings on the Hmong. Finally, it raises some issues to consider when working with students from different cultural backgrounds.

The possibility of a relationship between cultural experience and cognitive style has been supported, challenged, or rejected by anthropologists, psychologists, and educators. Indeed, the mere idea of such a relationship has been the subject of recent controversy and much debate. The controversy has arisen primarily out of a concern about biases in Western thought in reference to cultural differences. However timely, this concern has tended to be based on assumptions that confuse concepts of so-called "intelligence" with different approaches to learning which arise out of diverse socialization practices. As a result, even the mention of cultural cognitive style is sometimes interpreted as evidence of an arrogant and Eurocentric bias in regard to non-Western populations.

The debate has arisen out of a long series of studies in the fields of cultural anthropology, psychology, and education. These studies have focused on how thinking and learning occurs in various cultural contexts. While early studies were based on the cognitive developmental concepts of Piaget, others were derived from the pioneering work of Witkin and his associates (Witkin et al., 1973) and Berry (1976) on the relationship between culture and cognitive style. The long dialogue regarding the complexities and inter-relatedness of culture and cognitive processing is beyond the scope of this paper but has recently been addressed in a comprehensive review of cultural psychology by Michael Cole (1996).

Kraemer (1973) asserted that people sharing common primary experiences develop similar styles of cognitive processing including perceiving, conceiving, and judging. The concept of diverse cognitive styles arising out of different cultural experiences has been supported by Anderson (1988):

> Because the social, cultural, and environmental milieus of ethnic and racial groups differ, one should expect these differences to be reflected in their respective cultural/cognitive styles. Much of the literature in cross-cultural research supports this contention (p. 4).

More recently Shade (1997) has concurred with this view and has stated that:

> Culture, through the mediating process called cognitive style, determines the affective and cognitive behaviors which an individual selects to meet environmental demands. As environmental psychologists have been able to suggest, situations in which individuals find themselves tend to solicit the behavioral patterns necessary for survival within the confines of that situation. As such cognitive style has a significant impact upon an individual's competent performance in various behavioral settings (p. 10).

In addition, Shade (1997) maintains that culture influences not only cognitive processing but modes of communication and social interaction as well.

Basically, the literature on cultural considerations and cognitive style falls into three main categories: (a) an array of philosophical and historical essays about the relationship of culture and cognition; (b) a wide variety of research studies reporting differences in cognitive style and interactive modes among students from diverse groups both globally and in the United States; and (c) suggestions for taking cognitive style into account in teaching. The importance for teachers to know specific ways in which cultural experience impacts cognitive style, however, generally has not been taken into account in discussions of implementing cognitive style in classroom settings. An example of this relationship between learning at home and learning at school is described later in this paper in regard to Hmong students in American schools.

Learning Style or Cognitive Style?

The term *cognitive style* needs to be differentiated from *learning style*. Because these terms have sometimes been used interchangeably, some confusion has arisen as to what degree they overlap or refer to similar or different issues.

Learning Style

The term *learning style* has been used to refer to different factors, some internal, some external, some cognitive, some emotional, some social, and some behavioral. Irvine and York (1995) consider learning styles to be "an umbrella

term encompassing three distinct substyles: cognitive, affective, and physiological" (p. 484). Curry (1990) has pointed out this problem of ambiguity in regard to the term itself. Slavin (1997) refers to "Theories of Learning Styles" but switches to the term "cognitive style" without differentiating between them (p. 136).

Kagan (1964) distinguished between an impulsive and a reflective approach to learning. Entwistle (1981) later concurred about the importance of impulsivity or reflectivity in style. Fischer and Fischer (1979) referred to style as "a pervasive quality in the *behavior* (emphasis mine) of an individual" (p.245). Shade (1989) distinguished between an analytic and a synergetic style.

Fischer and Fischer (1979) further identified and described ten different kinds of learners: the incremental learner, the intuitive learner, the sensory specialist, the sensory generalist, the emotionally involved, the emotionally neutral, the explicitly structured, the open-ended structure(d), the damaged (in self concept and social competence among other problems), and the eclectic learner.

Based on individual *preferences for different learning conditions,* Dunn and Dunn (1979) identified four parameters of learning style: environmental, emotional, sociological, and physical. These parameters, or "stimuli," were further broken down into eighteen "elements." Among these, the environmental elements were sound, light, temperature, and design (or physical arrangement of the room); the emotional elements were motivation, persistence, responsibility, and a need for structure; the sociological elements included a preference for working alone, with peers, with an adult, or a combination of these potential partners; and the physical elements referred to perceptual strengths (visual, auditory, tactile, kinesthetic), a need for "intake" (food, drink), time of day, and greater or lesser need for mobility.

Entwistle (1981) suggested that style refers to *information processing.* Similarly, Nieto described learning style as "the way in which individuals process and receive information" (1992, p. 111). The term *learning style* as used by Entwistle and Nieto in regard to information processing is synonymous with cognitive style. Gardner (1983) has suggested that culture, affect, and cognition interact and are conducive to multiple intelligences (logical-mathematical, spatial, musical, kinesthetic, and interpersonal), thus blurring the distinction between culture, style, and different *abilities*.

Perhaps the most comprehensive definition of learning style is that of the National Task Force on Learning Style and Brain Behavior (as cited in Keefe and Languis, 1983):

> Learning style is that consistent pattern of behavior and performance by which an individual approaches educational experiences. It is the composite of characteristic cognitive, affective, and physiological behaviors that serve as relatively stable indicators of how a learner perceives, interacts with, and responds to the learning environment. It is . . . molded by . . . the cultural experiences of home, school, and society (p. 1).

Another ambiguity in definition of learning style is that the differences between style, strategy, and tactic have not always been clear. Entwistle (1988) suggested that strategy refers to *consistency in* (students') *approach* to different learning situations. Snowman (1989) suggested that tactic refers to the *observable activities or habitual responses* of students in learning situations. In view of these different interpretations of what learning style means , it is clear that different educators use the term "style" to refer to different processes and that in fact they are referring to behavior, preferences for different environments, strategies, or tactics.

In concordance with the concept of learning styles, a plethora of tests were created to measure "styles." Irvine and York (1995) report that more than thirty test instruments have been constructed. Some of these tests were designed for children, while others were created for adults and applied in both educational and business settings (Gregorc, 1982, for example). Research using these tests has been extensive. According to Irvine and York, several thousand studies were conducted between the mid 1980s and 1995. Curry (1990) has questioned both the validity and the reliability of many of these instruments. Timm (1996) has pointed out an additional problem in regard to learning style instruments. The forced choice format is based on an assumption that individuals have a fixed rather than an adaptive approach to learning situations and to problem solving. A final criticism of learning style instruments has been that they have low predictive value for achievement (Irvine and York, 1995). This, however, may be a spurious concern due to the fact that there is no reason to assume that one approach over another will necessarily result in success.

In spite of these problems in definition, test assumptions, and difficulties in utilizing test results in the classroom, the concept of learning styles does offer some important considerations about the relationship between cultural experience, individuality, and learning situations.

Cognitive Style

Correctly used, the term *cognitive style* derives from cognitive theory and refers to variations in information processing, perceiving, conceptualizing, analyzing, and problem solving procedures (Timm, 1996). Evidence suggests that cultures differ in respect to these processes. Ambiguities have occurred with the term cognitive style, however, similar to those associated with learning style. For example, Kuchinskas (1979) identified cognitive style "as the way an individual acts, reacts, and adapts to the environment" (p. 269).

In this review, the term *cognitive style* is used to refer to cognitive processes. Field independence or sensitivity, communication, and social interaction modalities are specified as such. Wherever the term *learning style* appears in this review, it is the term used by the author(s) under discussion.

Another interpretation of cognitive style (which also includes social and behavioral factors) is a concept known as *field independence/dependence*, first identified and described by Witkin and his associates (Witkin et al., 1971;

Witkin et al., 1977; Witkin, 1979; Witkin and Goodenough, 1981) by means of the Embedded Figures Test and subsequently the Group Embedded Figures Test (Witkin et al., 1973). The Children's Embedded Figures Test (Karp and Konstadt, 1971) was further developed from this test. These tests require the test taker to locate or identify basic geometric shapes embedded in surrounding complex patterns. Two important aspects of these tests have generally been overlooked in the literature. First, the shapes are basic configurations and, second, the tests are language free, thus eliminating the bias of linguistics, although directions for the test may be provided in different languages.

Because many studies have reported cultural differences in field independence/dependence, it is important to clarify these terms here. Chickering (1976) described field independence/dependence as differences in ability to distinguish figure from ground (or shape from pattern) and (by logical extension) a construct from its surrounding context. Field independent learners have been reported to be adept at identifying specific aspects of a situation and at separating concepts from context. Other characteristics include a preference to work independently, intrinsic motivation, and a desire for personal recognition. Heppner and Krauskopf (1987) further reported that field independent learners persevere longer and are more self-directive in their learning than field dependent learners. Field dependent learners tend to be situation specific in their orientation to learning, and tend not to separate concepts from context. Other characteristics include a preference to work with others, a need for extrinsic motivation, an orientation toward social cues, and a sensitivity to others. Heppner and Krauskopf (1987) have also reported that field dependent learners adapt to new situations more easily than field independent learners. Recently the term *field sensitive* has been used rather than field dependent. It is important to note that field independence or dependence are *value free* designates and that they should not be confused with notions about intelligence, ability, or as predictors of academic performance. They are simply tendencies along a continuum by which individuals perceive, conceptualize, and problem solve in their approach to a learning situation.

Cultural Factors in Learning

In the definition of learning style by the National Task Force on Learning Style and Brain Behavior (cited above), reference is made to the relationship between style and cultural experiences. Guild (1994) has reported three different sources for research information about the relationship between culture and learning processes. These are: (a) observations and descriptions of learners from different cultural groups; (b) data based on test instruments administered to diverse student populations; and (c) direct discussion (including interviews). The major ways in which cultural experiences affect cognitive style have not always been made explicit in reports of students from diverse groups, however. These experiences include socialization or child rearing practices, cultural "tightness," ecological or environmental considerations, a written or oral/aural language tradition (Worthley, 1987; Bennett, 1990), and so-called "high" or "low" context cultures (Halverson, 1993).

Permissive socialization practices, which encourage individual experimentation or trying different ways of performing tasks, result in a wider flexibility of cognitive style. Strict socialization practices, with pressure to perform tasks according to traditional ways, result in less flexibility of style (Jahoda, 1980). Strict practices which focus on obedience also tend to result in an orientation to learning which is specific to the present situation (Nedd and Gruenfeld, 1976).

Cultural "tightness" refers to the degree of emphasis and value given to traditional routines. Cultural "looseness " refers to the degree of latitude given to variation in the performance of daily tasks or routines. Thus "tight" cultures tend to follow precisely various time-honored ways while "loose" cultures are less rigid and more flexible in regard to traditional procedures (Worthley, 1987).

Ecological adaptation refers to customs in relation to nature within any given culture (Berry, 1976). For example, some cultures rely on highly developed perceptual skills for survival. Cultures which depend primarily on agriculture and animal husbandry emphasize customary routines in order to survive. Child rearing practices focus on responsibility, conformity to customs, and the value of traditional ways. Cultures which depend primarily on hunting, gathering, and to some extent fishing for survival require more self-reliance and application of skills under varying circumstances. Child rearing practices, while teaching traditional methods, also tend to encourage more individual initiative.

Literate societies use written symbol systems for the transmission of knowledge. Learning is more abstract and decontextualized than in oral societies which follow a more active mode and use demonstration and role modeling in order to teach. Learning is through observation and is based on specific situations (Hvitfeldt, 1985).

In addition to the foregoing considerations, Halverson (1993) has described another factor—that of "high" and "low" context cultures. In high context cultures, learning is situationally based within a social context. Skills and procedures are demonstrated and learning depends to a large degree on observation. Learners also relate the learning process to their place in social groups and to their role in society. In low context cultures, learning is more detached from the immediate use of the information and procedures are described in verbal or written form. Learners are less oriented toward the applicability of the information being transmitted in terms of the immediate task or social situation than they are in high context cultures.

Cultural Diversity and Cognitive Styles

There is a steadily increasing body of evidence in support of the notion of different patterns in cognitive style in-

cluding field independence/dependence among students from diverse cultural backgrounds. The following review focuses on diverse groups in the United States.

Asian Americans

Differences among Asian Americans have been reported in accordance with ethnic background.

The Hmong. As an example of the relationship between the cultural factors cited above and cognitive style, Timm and Chiang (1997) have described traditional Laotian Hmong culture and the cognitive style of Laotian Hmong students in the United States. In their former rural agricultural mountain communities in Laos, the Hmong approach to learning was situation specific. Strict socialization practices emphasized obedience and adherence to time honored procedures. The culture was "tight" with little latitude in routines. Ecologically, survival depended primarily on successful crops, although there was some hunting and fishing. As part of the socialization process, children participated in agricultural work as young as four years of age (Lee, 1986). Pressure for conformity was high in Hmong social organization, based on patrilineal clans with clear lines of male authority. Social roles were delineated along gender lines.

The culture was primarily oral and formal education was rare. Few villages had a school. Knowledge was handed down from generation to generation. It has been estimated that seventy percent of Hmong refugees were non-literate when they left Laos (Takaki, 1989). Thus, learning to use a written language was a profound problem which many faced in their relocation into literate societies such as the United States. The concept of writing was not unfamiliar to the Hmong, however. There have been "at least fourteen major attempts to develop writing systems for the Hmong language over the past one hundred years" (Smalley, 1990, p. 149) but Hmong students who did attend school were instructed in either Lao or French. The Hmong who cooperated with the United States during the Vietnam War gained some literacy in English (Duffy, 1997). The Romanized version of Hmong, developed in the early 1950s by two linguists (William Smalley and Linwood Barney) and a French priest (Yves Bertrais) and known as the Romanized Popular Alphabet (RPA), has become the most widely accepted and is the script used in the United States (J. Duffy, personal communication, January 12, 1998).

Finally, Hmong culture may be described as being high context. Learning was situationally based and children received their "education" at home and in the fields where they learned through observation. Procedures were demonstrated rather than discussed.

Hmong families in the United States continue to teach their children in the traditional way by using demonstration and relying on observational learning. At the same time, however, Hmong students are encountering curricular programs in American schools which transmit information in a decontextualized, written form and emphasize a more independent approach to learning. Using the Group Embedded Figures Test (available from Consulting Psychologists Press in Palo Alto) as the test instrument to determine field independent and field dependent cognitive styles, Timm and Chiang (1997) first reported a field dependent cognitive style consistent with Hmong situation specific learning experience. In a follow-up study, Timm, Chiang, and Finn (1998) found acculturating effects of length of residency in the United States and duration of time in American schools on Hmong students' cognitive style. Covariance statistical analyses yielded significant effects for both U. S. residency and years in American schools. In other words, evidence of Hmong cultural practices was found in the cognitive and social interaction styles of these students but shifts were also found from a situation specific or field dependent style to a more field independent style associated with the number of years the students had been living in the United States and attending American schools. Gender differences were also found in the shift in style with the boys moving into a field independent mode slightly ahead of the girls. This difference may be attributed to Hmong socialization practices in regard to gender roles (Timm et al, 1998).

Prior to the studies by Timm and Chiang (1997) and Timm et al. (1998), two earlier studies reported both cognitive and interaction styles consistent with Hmong cultural experiences. Hvitfeldt (1986) reported behaviors characteristic of a field dependent style in a literacy class for non-literate and low literate Hmong adults, ranging from twenty to sixty-five years of age. These behaviors included consistent interpersonal interactions among the students, a reliance on contextual referents, and a personal relationship with the instructor. Using the Group Embedded Figures Test, Worthley (1987) reported a two-to-one ratio of field dependence over field independence among Hmong male high school and college students, ranging from seventeen to thirty-five years in age.

Other Asian students. Reid (1987) also found acculturating effects among other Asian students and reported that college ESL students who had been in the United States for more than three years were significantly more auditory in their learning style preference in comparison with students who had been in this country for shorter periods of time. Reid further reported visual learning style preferences among Korean, Chinese, and Arabic-American students in comparison with Japanese students.

In a study of learning style preferences among Chinese, Filipino, Korean, Vietnamese, and Anglo high school students, Park (1997) reported major preferences for an auditory style among Vietnamese and Chinese American students, and a minor preference among Korean, Filipino and Anglo students. Park also reported a minor visual learning style preference among the four Asian groups in contrast with Anglo students who showed a negative response to visual learning. There were also differences among the Asian groups, with the Chinese students being the most visual, followed by the Filipino and Korean, and the Vietnamese students being the least visual in their preference. Ewing and

Yong (1993) also reported a visual preference among gifted American-born Chinese students.

Park (1997) further examined these students' preferences for group or individual approaches to learning. The Vietnamese students showed the highest preference for group learning, the Filipino students showed a minor preference for it, and the Chinese, Korean, and Anglo students did not prefer it. This is an important finding because cooperative learning approaches may work well with Vietnamese and Filipino students but not so well with Chinese, Korean, and Anglo students. Park reported that high achievers across all groups preferred an individual style and that low achievers preferred group learning.

Differences in socialization practices, social interaction styles, and educational values have been reported among other Asian American groups in reference to ethnicity and length of residency in the United States. Cabezas (1981) reported differences in socialization practices in the San Francisco area among Chinese and Filipino mothers born overseas in comparison with American born mothers. Rumbaut and Ima (1988) reported that Vietnamese, Chinese-Vietnamese, and Hmong parents in San Diego placed more emphasis on school achievement than Lao and Khmer (Cambodian) parents. These value differences may be attributed to their prior cultural experience. Lao refugees in the United States have tended to come from rural areas. The more educated and urban Lao refugees relocated in France following the takeover of Laos by communist forces after the Vietnam War. Likewise, many of the Khmer refugees who settled in the states were from rural areas of Cambodia and were less educated. The more educated Khmer were massacred during the Pol Pot regime. Consistent with Rumbaut and Ima, Timm (1994) reported that although Laotian Hmong families now living in the Midwest had come from rural areas where education was minimal, they have adopted a high value for education in regard to their children in the United States.

African Americans

Ogbu (1983) described an historical, caste-dominated society along racial lines in the United States by which exploitation has extended across economic, political and social experience. It is not surprising, therefore, that African American cultural patterns include values which emphasize group unity and mutual support (Staples, 1976). Jones (1979) added spirituality, spontaneity, and a preference for oral expression. Boykin (1986) suggested that African American culture contains nine themes: spirituality, harmony or interdependence with humans and nature, movement, "verve", affect, communalism or social connectedness, personal expression, oral tradition, and a focus on "social time." These aspects suggest that students may learn better through personal relationships with the teacher, cooperative learning modes, and oral strategies. In Shade's (1997) view, African American experience has led to "survivalisms" (p. 14) or an experiential wisdom among African Americans which is not shared by non-Blacks. According to Shade, the sources of African American culture include these survivalisms, European American mainstream society, and a culture of oppression which causes anxiety, over-identification with those in power, hostility, an ability to handle contradictions, and a preoccupation with issues of freedom and equality. Shade has suggested that "the kinship system (including protection and mutual support), world view, and social interactive behaviors have the greatest impact on learning style" (p. 15) and that African American culture and social stratification "serve as the transmitters of the cognitive and affective entry behaviors which come with the child to school" (p. 24).

Shade (1997) further reported an auditory processing mode, a precociousness sensori motor capability, a socially oriented (as opposed to an object centered) modality, and a preference for an interactive learning situation among Black children. She further suggested that perception (and therefore interpretation) of visual cues is affected by cultural experience. African Americans are more likely to be field dependent when tested on the Embedded Figures Test (Shade, 1986). This field sensitive finding is consistent with Gitter, Black, and Mostofsky (1972) who reported that African Americans are sensitive to social cues and adept at interpreting facial emotions. This social sensitivity impacts Black students' behaviors in the classroom (Shade). Ewing and Yong (1993) also found a preference for a visual learning mode among gifted African American students.

Mexican Americans

A sensitivity in the social interaction of Mexican Americans, together with an orientation to collective or collaborative efforts, reflects the traditional Mexican cultural values of close affiliation with family and community (Shade, 1997). Slonin (1991) suggested that Hispanic culture is based on cooperation, interpersonal relationships, a "relaxed" time perception , a preference for physical proximity, and traditional sex roles. Vasquez (1990) suggested that Hispanic American students' orientation of loyalty to family and groups may predispose them toward cooperative learning. Dunn and Dunn (1978) reported that Mexican American students were peer oriented and were more likely to perform well in cooperative group situations. In a large study of Mexican American immigrant and first generation elementary students and Anglo American elementary students (n=687), Dunn, Griggs, and Price (1993) found that the Mexican American students were more peer-oriented than were the Anglo students, with the Mexican American girls more peer oriented than the boys. They also found that the Mexican American boys had the strongest preferences for tactile learning and that the Mexican American girls in general showed less tactile learning preferences and a more varied approach to learning than the boys. Similarly, Ewing and Yong (1993) reported that gifted Mexican American students preferred a kinesthetic learning style over an auditory or visual one. Mori (1991) reported that Mexican students with higher English proficiency continued to show a stronger orientation for active learning in compari-

son with high English proficiency Japanese students who did not prefer this modality.

Saracho (1991) cautioned against making assumptions about cognitive style in Mexican American children, however. She asserted that, although a generally field dependent, prosocial orientation has been assumed in Mexican American children, field independence/dependence "is a relative rather than an absolute tern (and that) extensive data must be collected and analyzed before accepting any generalizations" (p.23). In a study of Mexican American kindergarten children from an agricultural community, Saracho found a range of field independence/dependence on The Children's Embedded Figures Test (CEFT). She also found significant differences in the children's play behavior and social competence. In other words, Saracho found both a diversity of cognitive styles and a range of social competency related to that stylistic diversity.

Saracho (1997) further suggested that both the amount of traditional procedures in child rearing and the degree of generational distance from migration to the United States both affect cognitive style. Several findings on differences in cognitive style among Mexican Americans in relation to Anglo contact support Saracho's view. Some of these findings are similar to the findings for Hmong students with regard to United States residency (Timm et al., 1998). For example, Buriel (1975) reported that first and second generation Mexican immigrants had cognitive styles similar to traditional communities, but the third generation did not. Ramirez and Castaneda (1974) reported that Mexican American students were inclined toward a field sensitive learning style but that style varied in relation to assimilation, distance from Mexico, length of residence in the United States, impact of urbanization, and amount of prejudice encountered. Ramirez, Castaneda, and Herold (1974) studied three different types of communities: (a) Mexican American members with a primarily traditional Mexican culture; (b) dualistic with Mexican American members and a mixture of Mexican and Mexican American cultures; and (c) Mexican American members with manifest values from Anglo-American culture. They reported that the students from the dualistic community were in between the more field dependent members of the traditional community and the less field dependent members of the Anglo-oriented community. Other studies have reported similar results from traditional and dualistic communities (Laosa and DeAvila, 1979).

This section has focused on Mexican American students but other students may experience shifts in their cognitive style in relation to type of community, demographic considerations, and length of residency in the United States. In light of Creason's report (1992) that 40% of Hispanic students drop out of school, there is clearly a need for more research in this area.

Native Americans

Smith and Shade (1997) cited some Native American cultural factors that are conducive to a field sensitive cognitive style and socially sensitive interactive style. Among these are a conviction of the inherent good of all people, a belief that all people are interconnected with each other and with nature, and a view that cooperation is important for solving problems. According to Pepper and Henry (1997), socialization among Native Americans tends to be permissive and children are encouraged to experiment and to explore. Discipline does not mean obedience, but development of self control whereby children come to regard non-interference as normal. "Respect for individual dignity and personal autonomy are valued and youngsters are taught not to interfere in the affairs of others" (p. 170). Socialization further emphasizes observational and contextually relevant learning. Thus a cognitive style emerges that includes a preference for visual processing, an informal and exploratory learning preference, and a sensitivity to social cues.

An association between culture, ecology, and cognitive style has been reported among Native Americans by Kleinfeld (1970). In a testing situation for visual memory which required the ability to recall complex visual patterns, rural Inuit native children of all ages outperformed urban White children. These results were attributed to the ecology of a sparse Arctic landscape and to socialization that included a hunting tradition, both of which require visual acuity and an ability to perceive slight variations in the environment. Berry (1971) also reported visual acuity among urban Inuit subjects, in spite of less hunting experience. This finding suggests that Inuit child rearing practices emphasize visual learning, imitation, and non-verbal instruction.

Phillips (1978) reported that Native American students show a preference for learning by observation before they attempt to perform a task themselves. According to More (1987), Native American students prefer a visual to verbal learning mode and use images to learn concepts. These characteristics suggest a field sensitive cognitive style. Caldwell (1989) and Kasten (1992) reported a preference for cooperation in learning situations among Native American students.

Gender and Social Class

Within diverse groups, cognitive style may be mediated by gender and socioeconomic status. In a large study (636 boys and 638 girls), Park (1997) reported gender differences in style preferences across auditory, visual, and tactile modes and a significant gender difference in kinesthetic preference, with the girls reporting a higher preference. Some findings of gender differences within groups are reported above for Hmong and Mexican American students. Social class differences are also sometimes overlooked in the reporting of cognitive styles. Blackwell (1975) reported that African Americans in the professional/middle class and skilled blue collar class are more oriented to achievement, social striving, and consumerism in comparison with the economically disadvantaged. In an early study of Chinese, Jewish, Black, and Puerto Rican children from middle class and low income homes, however, Stodolsky and Lesser (1967) reported different patterns in cognitive processes for each ethnic group

regardless of social class. In other words, ethnicity appeared to influence cognitive style more than social class. Banks (1988) also reported similar findings of the effect of ethnicity over social class and further reported that ethnic differences remained even when social class had changed for the better. These findings suggest that the interrelationship between ethnicity, gender, and social class is a complicated one in which cognitive styles may not necessarily be assumed by one dimension alone.

Educational Implications

The research findings considered in this review raise some important issues for classroom application. First, not all students in any cultural group necessarily approach learning in the same way. As Irvine and York (1995) assert, stereotyping occurs when inaccurate or general characteristics of a group are ascribed to, or assumed, for individuals. Second, educators must remember that learning is a fluid process and that students' cognitive styles are not static but may change across time. Findings of acculturation effects among Mexican American, Hmong, and other Asian Americans suggest that individual differences and acculturating experiences must be considered. Third, individuals may use different approaches to learning and problem solving, depending on the nature of the problem. Timm (1996) reported the following anecdote:

> . . . a teacher was required to take a widely marketed learning style test by her school administrator. During the test she considered how she approached the task of writing a report and answered the test items accordingly. Being suspicious of the test's validity, she asked to retake the test immediately. Because her hobby was sewing, this time she considered how she approached the task of creating a dress of her own design. The results of her two tests indicated two totally different learning styles (p. 190).

In other words, the creators of learning style tests have not generally taken into consideration the fact that people may use a variety of approaches that best suit the task at hand.

In spite of these caveats, the above review does reveal some general patterns for diverse groups. Shade (1997) suggested that cognitive processes are the result of socialization and cultural experiences and that the environment is interpreted through cultural filters and responded to accordingly. Thus, people who share common experiences develop similar processes of "conceiving, judging, and reasoning" (p. 134). Shade, Kelly, and Oberg (1997) offer a variety of teaching strategies for working in culturally responsive classrooms. As educators, we need to remember that our own interpretations, problem solving strategies, and communication styles are the result of our cultural experiences, but we sometimes forget our own ethnocentrism in these matters. And worse, we make judgments about the abilities of students that are filtered through our own cultural lenses.

I will close by sharing an incident, told to me by a Wisconsin teacher, that dramatically illustrates how a school task may be culturally biased and fail to take diverse cultural styles into account. Hmong students in a Wisconsin school were given a sorting test and asked to draw a circles around objects that did **not** belong. One test item included a picture of a hammer, a saw, a hatchet, and a fire. The "correct" answer was the fire because it was not a tool, but the Hmong students choose the hammer. Rather than assuming that the students were wrong, the teacher asked them why they had chosen the hammer. They told her that "you would use a saw or a hatchet to cut the wood for the fire but not the hammer." This context oriented and procedurally based answer is not surprising in Hmong culture. There is a lesson here for all of us.

References

Anderson, J. A. (1988). Cognitive styles and multicultural populations. *Teacher Education*, *39*(1), 2-9.

Banks, J. A. (1988). Ethnicity, class, cognitive and motivational styles: Research and teaching implications. *Journal of Negro Education*, *57*(4), 452-466.

Bennett, C. I. (1990). *Comprehensive multicultural education*, (2nd ed.). Boston: Allyn and Bacon.

Berry, J. (1971). Ecological and cultural factors in spatial perceptual development. *Canadian Journal of Behavioral Science*, *3*(4), 324-336.

Berry, J. W. (1976). *Human ecology and cognitive style: Comparative studies in cultural and psychological adaptation*. New York: Wiley.

Blackwell, J. E. (1975). *The Black community: Diversity and unity*. New York: Dodd, Mead and Co.

Boykin, A. W. (1986). The triple quandary and the schooling of Afro-American children. In U. Neisser (Ed.), *The school achievement of minority children*, (pp. 57-92). Hillsdale, NJ: Lawrence Erlbaum Associates.

Buriel, R. (1975). Cognitive styles among three generations of Mexican-American children. *Journal of Cross-Cultural Psychology*, *6*(4), 417-429.

Cabezas, A. (1981). *Early childhood development in Asian and Pacific American families: Families in transition*. San Francisco: Asian, Inc.

Caldwell, A. J. (1989). *Cultural learning styles: American Indian students in the classroom*. Madison, WI: Wisconsin Department of Public Instruction.

Chickering, A. W. (1976). Commentary: The double bind of field dependence/independence in program alternatives for educational development. In S. Messick (Ed.), *Individuality in learning*, (pp. 79-89). San Francisco: Jossey-Bass.

Cole, M. (1996). *Cultural psychology: A once and future discipline*. Cambridge, MA: Harvard University Press.

Creason, P. (1992). Changing demographics and the importance of culture in student learning styles. (ERIC Document Reproduction Service No. Ed 361 270).

Curry, L. (1990). A critique of the research on learning styles. *Educational Leadership*, *48*(2), 50-56.

Duffy, J. (1997, October). Ideologies of literacy: The historical development of reading and writing in Laos: 1921-1975. Paper presented at the meeting of The Mid-Western Educational Research Association, Chicago, IL.

Dunn, R. S. and Dunn, K. J. (1978). *Teaching students through their individual learning styles: A practical approach*. Reston, VA: Reston Publishing.

Dunn, R. S. and Dunn, K. J. (1979, January). Learning styles/teaching styles: Should they . . . can they. . . be matched? *Educational Leadership*, *36*(4), 238-244.

Dunn, R., Griggs, S., and Price, G. E. (October, 1993). Learning styles of Mexican-American and Anglo-American elementary students. *Journal of Multicultural Counseling and Development*, *21*, 237-247.

Entwistle, N. (1981). *Styles of learning and teaching*. Chichester: Wiley.

Entwistle, N. (1988). Motivational factors in students' approaches to learning. In R. Schmeck (Ed.), *Learning strategies and learning styles*. New York: Plenum.

Ewing, N. J. and Yong, F. L. (1993). Learning style preferences of gifted minority students. *Gifted Education International*, *9*(1), 40-44.

Fischer, B. B. and Fischer, L. (1979). Styles in teaching and learning. *Educational Leadership*, *36*(4), 245-254.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gitter, A. G., Black, H., and Mostofsky, D. (1972). Race and sex in the perception of emotion. *Journal of Social Issues*, *28*(4), 63-78.

Gregorc, A. F. (1982). *Transaction ability inventory*. Department of Secondary Education, University of Connecticut, Storrs.

Guild, P. (1994, May). The culture/learning style connection. *Educational Leadership*, *51*, 16-21.

Halverson, C. B. (1993). Cultural context inventory: The effects of culture on behavior and work style. *The 1993 Annual: Developing Human Resources*. (131-139). San Diego, CA: Pfeiffer and Company.

Heppner, P. and Krauskopf, C. (1987). An information processing approach to personal problem solving. *Counseling Psychologist*, *15*(3), 371-447.

Hvitfeldt, C. (1985). Picture perception and interpretation among preliterate adults. *Passage: A Journal of Refugee Education*, *1*(1), 27-30.

Hvitfeldt, C. (1986). Traditional culture, perceptual style, and learning: The classroom behavior of Hmong adults. *Adult Education Quarterly*, *36*(2), 65-77.

Irvine, J. J. and York, D. E. (1995). Learning styles and culturally diverse students: A literature review. In J. A. Banks and C. A. M. Banks (Eds.), *Handbook of research on multicultural education*, (pp. 484-497). New York: Macmillan Publishing USA.

Jahoda, G. (1980). Theoretical and systematic approaches in cross-cultural psychology. In H. C. Triandis and W. W. Lambert (Eds.), *Handbook of cross-cultural psychology: Perspectives, Vol. 1*, (pp. 69-142). Boston: Allyn and Bacon.

Jones, J. M. (1979). Conceptual and strategic issues in relationship of Black psychology to American social science. In A. W. Boykin, A. J. Franklin, and J. F. Yates (Eds.), *Research directions of Black psychologists*. New York: Russell Sage Foundation.

Kagan, J. (1964). American longitudinal research on psychological development. *Child Development*, *35*(1), 1-32.

Karp, S. A. and Konstadt, N. L. (1971). Children's embedded figures test. In H. A. Witkin, P. K. Oldman, E. Raskin, and S. A. Karp (Eds.), *A manual for the embedded figures tests*, (pp. 21-26). Palo Alto, CA: Consulting Psychologists.

Kasten, W. C. (1992). Bridging the horizon: American Indian beliefs and whole language learning. *Anthropology and Education Quarterly*, *23*(2), 108-119.

Keefe, J. W. and Languis, M. (1983). Untitled article. *Learning Stages Network Newsletter*, *4*(2), 1.

Kleinfeld, J. (1970). C*ognitive strength of Eskimos and implications for education*. University of Alaska. Institute of Social, Economic and Government Research.

Kraemer, A. J. (April, 1973). A cultural self-awareness approach to improving intercultural communication skills. (ERIC Document Reproduction Service No. ED 079 213).

Kuchinskas, G. (1979). Whose cognitive style makes the difference? *Educational Leadership*, *36*(4), 269-271.

Laosa, L. M. and DeAvila, E. A. (1979). Development of cognitive styles among Chicanos in traditional and dualistic communities. *International Journal of Psychology*, *14*(2), 91-98.

Lee, G. Y. (1986). Culture and adaptation: Hmong refugees in Australia. In G. L. Hendricks, B. T. Downing, and A. S. Deinard (Eds.), *The Hmong in transition*, (pp. 55-72). Staten Island, NY: Center for Migration Studies.

More, A. J. (1987). Native -American learning styles: A review for researchers and teachers. *Journal of American Indian Education*, *27*(1), 17-29.

Mori, S. (1991, Spring-Fall). ESL classroom personality. *Journal of Intensive English Studies (JIES)*, *5*, 37-54.

Nedd, A. N. and Gruenfeld, L. W. (1976). Field dependence-independence and social traditionalism: A comparison of ethnic subcultures of Trinidad. *International Journal of Psychology*, *11*(1), 23-41.

Nieto, S. (1992). *Affirming diversity: The sociopolitical context of multicultural education*. New York: Longman.

Ogbu, J. U. (1983). Minority status and schooling in plural societies. *Comparative Education Review*, *27*, 168-190.

Park, C. C. (1997). A comparative study of learning style preferences: Asian-American and Anglo students in secondary schools. Paper presented at the American Educational Research Association, Chicago, IL.

Pepper, F. C. and Henry, S. (1997). Social and cultural effects on Indian learning style: Classroom implications. In B. J. R. Shade (Ed.), *Culture, style, and the educative process*, (2nd ed.), (pp. 168-177). Springfield, IL: Charles C. Thomas.

Phillips, J. C. (January 16, 1978). College of, by, and for Navajo Indians. *Chronicle of Higher Education*, 10-12.

Ramirez, M. and Castaneda, A. (1974). *Cultural democracy, bicognitive development and education*. New York: Academic Press.

Ramirez, M., Castaneda, A., and Herold, P. L. (1974). The relationship of acculturation to cognitive style among Mexican Americans. *Journal of Cross Cultural Psychology*, *5*, 424-433.

Reid, J. (1987). The learning style preferences of ESL students. *TESOL Quarterly*, *21*(1), 87-111.

Rumbaut, R. and Ima, K. (1988). *The adaptation of Southeast Asian refugee youth: A comparative study*. Washington, DC: U. S. Department of Health and Human Services, Office of Refugee Resettlement.

Saracho, O. N. (1991). Cognitive style and social behavior in young Mexican American children. *International Journal of Early Childhood*, *23*(2), 21-38.

Saracho, O. N. (1997). Cultural differences in the cognitive style of Mexican-American students. In B. J. R. Shade (Ed.), *Culture, style, and the educative process*, (2nd Ed.), (pp. 118-125). Springfield, IL: Charles C. Thomas.

Shade, B. J. R. (1986). Is there an Afro-American cognitive style? *Journal of Black Psychology*, *13*, 13-16.

Shade, B. J. R. (1989). *Culture, style, and the educative process*. Springfield, IL: Charles C. Thomas.

Shade, B. J. R. (1997). *Culture, style, and the educative process*, (2nd Ed.). Springfield, IL: Charles C. Thomas.

Shade, B. J. R., Kelly, C., and Oberg, M. (1997). *Creating culturally responsive classrooms*. Washington, DC: American Psychological Association.

Slavin, R. E. (1997). *Educational psychology: Theory and practice*, (5th ed.). Boston, MA: Allyn and Bacon.

Slonin, M. B. (1991). *Children, culture, and ethnicity*. New York: Garland.

Smalley, W. A., Vang, C. K., and Yang, G. Y. (1990). *Mother of writing: The origin and development of a Hmong messianic script*. Chicago: The University of Chicago Press.

Smith, M. and Shade, B. J. R. (1997). Culturally responsive teaching strategies for American Indian students. In B. J. R. Shade (Ed.), *Culture, style, and the educative process*, (2nd Ed.), (pp. 178-186). Springfield, IL: Charles C. Thomas.

Snowman, J. (1989). Learning tactics and strategies. In G. D. Phy and T. Andre (Eds.), *Cognitive instructional psychology: Components of classroom teaching*. New York: Academic Press.

Staples, R. (1976). Black culture and personality. In R. Staples (Ed.), *Introduction to Black Sociology* (pp. 55-88). New York: McGraw-Hill.

Stodolsky, S. S. and Lesser, G. (1976). Learning patterns of the disadvantaged. *Harvard Educational Review*, *37*, 546-553.

Takaki, R. (1989). *Strangers from a different shore*. Boston, MA: Little, Brown and Company.

Timm, J. T. (1994). Hmong values and American education. *Equity and Excellence in Education*, *27*(2), 36-44.

Timm, J. T. (1996). *Four perspectives in multicultural education*. Belmont, CA: Wadsworth.

Timm, J. T. and Chiang, B. (1997). Hmong culture and cognitive style. In B. J. R. Shade (Ed.), *Culture, style, and the educative process*, (2nd Ed.), (pp. 105-117). Springfield, IL: Charles C. Thomas.

Timm, J. T., Chiang, B., and Finn, B. D. (1998). Acculturation in the cognitive style of Laotian Hmong students in the United States. *Equity and Excellence in Education*, *31*(1), 29-35.

Vasquez, J. A. (March, 1990). Teaching to the distinctive traits of minority students. *The Clearing House*, *63*, 229-304.

Witkin, H. A. (1979). Socialization, culture and ecology in the development of group and sex differences in cognitive style. *Human Development*, *22*(5), 358-372.

Witkin, H. A. and Goodenough, D. R. (1981). *Cognitive styles: Essence and origins, field dependence and field independence*. New York: International Universities Press.

Witkin, H. A., Moore, C. A., Goodenough, D. R., and Cox, P. W. (Winter, 1977). Field dependent and field independent cognitive styles and their educational implications. *Review of Educational Research*, *47*(1), 1-64.

Witkin, H. A., Oldman, P. K., Cox, P. W., Ehrlichman, E., Hamm, R. M., and Ringler, R. W. (1973). *Field-dependence-independence and psychological differentiation: A bibliography*. Princeton, NJ: Educational Testing Service

Witkin, H. A., Oldman, P. K., Raskin, E., and Karp, S. A. (1971). *A manual for the embedded figures tests*. Palo Alto, CA: Consulting Psychologists Press.

Worthley, K. M. E. (1987). Learning style factor of field dependence/ independence and problem solving strategies of Hmong refugee students. Unpublished master's thesis, University of Wisconsin Stout, Menomonie, WI.