

# Causal Inference in Instructional Workforce Research

**Annual Progress Report  
March 2005**

Mary M. Kennedy  
Betsy J. Becker

This project addresses two intertwined problems. One is the problem of drawing causal inference in cases where experimental conditions cannot be met, the other is the problem of ascertaining what existing research can tell us about the relative importance of different types of teacher qualifications. For both problems, we rely heavily on a large data base of studies that claim to tell us something about teacher qualifications. There are nearly 500 studies in this data base, of which only two are experimental. Thus if policy makers use strict experimental standards to identify useful research, they must reduce the available literature from 478 studies to 2. If, on the other hand, audiences want to include a larger pool of literature, they must find ways to evaluate and synthesize research that is highly various in its samples, designs, statistical approaches, and measurements. We prefer the latter approach, recognizing that numerous methodological issues arise in this approach, both with respect to how to identify strengths and weaknesses of the underlying studies and with respect to synthesizing highly various approaches to research.

Our approach to the problem has been to multi-pronged. We have developed a relatively large collection of projects, each of which slices into the issue in a different way. This report describes our work in three sections. The first section describes activities that support the entire project. The second lists the various analytic studies that are currently under way, and the third lists miscellaneous papers and presentations that we have completed (or nearly completed) during this year.

## Supporting Activities

Two activities support the overall project. One of these is a detailed digital data base, with a complicated coding scheme that allows us to characterize studies in quite a refined manner. The second is a web site that holds all of our paper products so far as well as a list of our library of studies.

### Data Base

Our data base currently contains 478 studies which rely on remarkably different samples, variables, and analytic strategies. Our coding scheme is designed to allow us to capture all of

these variations. It is now on line so that staff can access it from any location in order to code a study, review existing codes for one or more studies, or search for a set of studies that meets some specific criteria (e.g., regressions involving secondary math teachers). We have separate screens for characterizing the article, authors, sampling strategy, grade levels and subjects in the sample, qualifications examined, and outcomes examined. Once these general features are captured, we move on to more detailed examination of the study design, capturing all of the variables that are included as covariates or design controls. Below are some snapshots of some of the screens: The first screen shows how a study's sampling strategy is characterized. The second is used to list the particular qualifications that were included in the study. The third and fourth are from the coding section that characterizes the study design and control variables used in the study. For that section, we have a separate page for controls at each level of the system (student, teacher, school, etc)

Study ID	AndersDor81	Coder:	Yanxuan	Coder #:	2
<b>III. Sample Characteristics (Part 1)</b>					
<b>A. Sampling Strategies</b>					
<input type="checkbox"/> any national data base					
Existing database	A Local Data Base				
Target group	Elementary school teachers and students in St. Louis, Missouri				
Sampling method	<input type="checkbox"/> NI <input type="checkbox"/> Purposive <input type="checkbox"/> Convenience (inferred) <input checked="" type="checkbox"/> Census <input type="checkbox"/> Probability (random,stratified/cluster) <input type="checkbox"/> Other				
Other method					
<b>B. Assignment Method</b>					
Method of assignment	<input type="checkbox"/> NI <input type="checkbox"/> Matched <input type="checkbox"/> Naturally occurring groups <input checked="" type="checkbox"/> NA <input type="checkbox"/> Random <input type="checkbox"/> Other				
Notes on assignment					
<b>C. Setting</b>					
States:	<input type="checkbox"/> NI	#	1	(List)	MO
Regions/Districts	<input type="checkbox"/> NI	<input type="checkbox"/> NA	#	1	(List) St. Louis
Universities/COLLEGES	<input type="checkbox"/> NI	<input checked="" type="checkbox"/> NA	#		(List)
Schools (#)	<input type="checkbox"/> NI				
Location	<input checked="" type="checkbox"/> NI	<input type="checkbox"/> Urban	<input type="checkbox"/> Suburban	<input type="checkbox"/> Rural	
School type	<input type="checkbox"/> NI	<input checked="" type="checkbox"/> All Public	<input type="checkbox"/> All Private	<input type="checkbox"/> Mix	Public (#) _____ Private (#) _____
Classes (#)	<input checked="" type="checkbox"/> NI				
Special characteristics					
<a href="#">◀ Back to Author Data</a>			<a href="#">More Sample Data ... ▶</a>		

Study ID MonkKing94

Coder: MJ

Coder #: 2

#### IV. Hypotheses and Indicators of Teacher Qualifications (Part 1)

##### General Ability or Achievement

- |  |  |
|--|--|
| <input type="checkbox"/> 1a IQ                           | <input type="checkbox"/> 1g GPA                  |
| <input type="checkbox"/> 1b Verbal ability               | <input type="checkbox"/> 1h Institutional status |
| <input type="checkbox"/> 1c SAT/ACT                      | <input type="checkbox"/> 1i Other test score     |
| <input type="checkbox"/> 1d AP placement tests           | <input type="checkbox"/> 1j Self assessment      |
| <input type="checkbox"/> 1e Basic Skills Praxis I, other | <input type="checkbox"/> 1k Other                |
| <input type="checkbox"/> 1f NTE                          |  |

Any indicators of general ability

##### Content Knowledge

Any indicator of content knowledge

- |   |  |
|---|--|
| <input type="checkbox"/> 2a Certificate in subject                                  | <input type="checkbox"/> 2h GPA in subject                             |
| <input type="checkbox"/> 2b Degree in subject (BA/BS in subject) vs. not            | <input type="checkbox"/> 2i Test in subject, TE, Praxis II, GRE, Other |
| <input type="checkbox"/> 2c Degree level (BA vs. MA vs. other in subject)           | <input type="checkbox"/> 2j Self assessment                            |
| <input checked="" type="checkbox"/> 2e Major vs minor vs neither in subject         | <input type="checkbox"/> 2k Other                                      |
| <input checked="" type="checkbox"/> 2g Number of courses or credit hours in subject |  |

##### Pedagogical Knowledge

Any indicator of pedagogy

- |   |  |
|---|--|
| <input type="checkbox"/> 3a Certificate in Ed.                            | <input type="checkbox"/> 3i Test of pedagogical knowledge like Praxis II |
| <input type="checkbox"/> 3b BA/BS in education (vs not)                   | <input type="checkbox"/> 3j Self assessment                              |
| <input type="checkbox"/> 3c Degree level in education (BA vs MA vs Other) | <input type="checkbox"/> 3k Other  |
| <input type="checkbox"/> 3e Major vs minor vs neither in education        |  |
| <input type="checkbox"/> 3g Number of courses or credits in education     |  |
| <input type="checkbox"/> 3h GPA in education                              |  |

##### Pedagogical Content Knowledge

Any indicator of PCK

- |  |  |
|--|--|
| <input checked="" type="checkbox"/> 4g # of courses or credits in methods of teaching specific content |  |
| <input type="checkbox"/> 4i Test of PCK  |  |
| <input type="checkbox"/> 4j Self assessment  |  |
| <input type="checkbox"/> 4k Other  |  |

##### Practical Knowledge

Any indicator of practical knowledge

- |   |  |
|---|--|
| <input checked="" type="checkbox"/> 5a Years experience | <input type="checkbox"/> 5g Career ladder status |
| <input type="checkbox"/> 5b Novice vs experienced       | <input type="checkbox"/> 5i NPPTS certification  |
| <input type="checkbox"/> 5c Test like Praxis III        | <input type="checkbox"/> 5j Self assessment      |
| <input type="checkbox"/> 5e INTASC assessment           | <input type="checkbox"/> 5k Other                |

More TQs ... ▶

◀ Back to Sample Data, Part 2

Study ID	AaronsonBS03	Link	Student	Teacher	School	District
Sample ID	1	Variables Controlled: Students				
LinkID	1	Control Variables				
TQ_ID	1					
QT_ID	1	Variable	N of instances	Type of control		
		Race/ethnicity	1	Statistical controls		
		Student SES	1	Statistical controls		
		Student age	1	Statistical controls		
		Student sex	1	Statistical controls		
		Study activities				
		Other activities				
		Affective variables				
		Pretest scores	1	Statistical controls		
		Other test scores				
		Family structure	1	Statistical controls		
		Parental education				
		Parental expectations				
		Parent school involvement				
		Other student variables	2	Statistical controls		
		N of Student Controls	8			

Overall Info	Link	Student	Teacher	School	District
Study ID AaronsonBS03					Variables Controlled: Teachers
Sample ID 1					Control Variables
Link ID 1					
TQ_ID 1		Variable	N of instances	Type of control	
QT_ID 1		Subject taught			
		Teacher age			
		Teacher certification status	6	Statistical controls	
		Teacher's college	6	Statistical controls	
		College major	3		
		Courses taken			
		Teacher GPA			
		Teacher class size			
		Grade taught			
		Highest degree earned	2	Statistical controls	
		Teacher job satisfaction			
		Teacher race / ethnicity	4	Statistical controls	
		Teacher SES			
		Teacher sex			
		Teacher years of experience	6	Statistical controls	
		Teacher perceptions of support			
		Teacher perceptions of school problems			
		Teacher perceptions of stu-tchr relations			
		Teacher activities			
		Teachers' TE programs			
		Other teacher variables			
		Number of teacher controls	27		

## **Web Site**

The second important supporting activity in the project is our TQQT web site. We began this as a vehicle for storing project related documents so that everyone could have easy access to them. The site currently holds a list of the studies in our data base, most of the papers we have written to date, as well as coding forms and most of our proposals and progress reports as well.

The site has expanded substantially since we started it and we have recently had to ask the university for more space to contain it. During the next year of our grant, we expect to engage in a major overhaul of the site and make it more of a public site rather than a private one. The new site will remove some of our internal working documents and focus more on public documents and summaries of research that might be of interest to a wider audience.

The site, in its current incarnation, can be found at <http://www.msu.edu/~mkennedy/TQQT>

## **Analytic Studies Currently Under Way**

Our analytic studies are of two kinds. Some aim to synthesize particular subsets of literature while others aim to develop new strategies for synthesizing research. These two strands of work obviously interact. We find that efforts to synthesize provoke new questions about method, and that efforts to devise new methods give us new ideas about syntheses already under way. A central concern we have is that extant methods of synthesis do not allow for inclusion of the full range of analyses our focal studies have used. In fact, often those studies that have the highest quality are least amenable to meta analysis, either because they have engaged in taken unique steps in their design or analysis to control for confounding variables or because they are using relatively new methodologies, such as hierarchical linear modeling or structural equation modeling. So while our synthesis projects focus on relatively simpler study designs (simple group comparisons or correlational studies), our methodological work aims to stretch these by finding ways to accommodate more complex designs. We list these projects separately here, but with the caveat that the distinctions are not always neat and our final products may not reflect the labels we currently are using.

## **History of Certification Policies**

One problem we discovered relatively early in this project was that many of the variables commonly used in this literature do not have stable meanings. For instance, “certification” is a frequently-measured teacher qualification, but the meaning of certification can change dramatically from one state to another, and can change over time as states alter their requirements. This problem applies both to “traditional” certification and to “alternative routes” to certification. To help us better interpret study findings that purport to measure the effects of teachers’ certification, we have commissioned a history of certification policies over the past 40 years. This project examines trends in state policies from 1960 to the present and provides us with an important context for interpreting research findings during this period.

## **Effects of Certification on Teacher Quality**

Teachers begin their teaching career with various kinds of teacher certification requirements. The big concern is whether these teachers with different kind of certificates teach equally well. Thirty three studies comparing traditionally certified teachers, alternatively certified teachers or emergency certified teachers were synthesized to determine whether teachers with a certain kind of teacher certificate perform better than teachers with another kind of certificate. Their effectiveness is evaluated in terms of both classroom teaching performance and student academic achievement. Although we include only studies comparing teachers with three different kinds of certification (traditional, alternative, and emergency) the variation among the included studies is still tremendous. This is embodied in two ways. First, the definition for each kind of teacher certification, especially the definition for alternative certificate, varies across different studies. Second, different studies have different controlling variables by different ways, either design control (matching) or statistical control (ANCOVA or Multiple Regressions). These variations in study design may shield the detection of some effects. There are also 5 regression studies about the effectiveness of teacher certification. Instead of synthesizing the slopes from the regression studies with the effect sizes from comparison studies, the regression results will be described and presented qualitatively.

## **Examination of Commercial Interviews as Hiring Criteria**

Commercial teacher interviews, like SRI/Gallup's Teacher Perceiver Instrument (TPI), claim to identify teacher candidates who are most likely to succeed on the job based on their attitudinal and personality characteristics. Instruments like the TPI use a battery of multiple-choice questions to measure the candidate's beliefs, attitudes, and values along a series of affective orientations (generally called "themes"). Scholars like Martin Haberman at the University of Wisconsin-Milwaukee have argued that a teacher's beliefs, attitudes, and values are the most important indicator of his or her quality as a teacher; Haberman has even designed and marketed his own commercial teacher interview aimed at identifying teachers most likely to be successful in urban schools. In 2005 the TPI was re-design from a personally scored, face-to-face interview into an impersonally scored, online screener. However, the Perceiver's affective themes remain the underlying theoretical and evidentiary premise. What is unclear is whether the affective values and traits measured by commercial interviews like the TPI are valid representations of teaching ability, or whether they instead reflect the teacher's likeability or skill at strategizing responses to questions about teacher attitudes. If the instrument identifies personality traits that increase the likelihood that principals will think highly of them, it does not necessarily follow this makes them more effective teachers. To examine these questions, we are engaging in a meta-analysis study of the predictive value of the TPI. The synthesis examines data from more than two dozen studies conducted over the past three decades.

## **Comparison of Alternative Indicators of Math Knowledge**

Evidence that knowledge of mathematics is relevant to the quality of teaching has been surprisingly elusive. Ever since Begle's famous 1972 study, researchers have sought clear evidence that knowledge of mathematics improves mathematics teachers' effectiveness. This project synthesizes this literature. We have located 46 studies that investigate the relationship between some indicator of teachers' mathematics knowledge, on one side, and their students' mathematics achievement on the other. Of interest is whether some indicators (e.g., courses in mathematics) might be better predictors than others (e.g., test scores in mathematics). However, the project, like all of our projects, confronts certain methodological challenges. One challenge is to combine studies using a variety of data analytic techniques such as correlation, analysis of covariance (ANCOVA), regressions, and hierarchical linear model (HLM). No techniques now exist for obtaining effect sizes from analyses such as HLM and regression. In addition, we don't know how to ensure the comparability of effect sizes when studies use different control variables. For instance, each regression equation presents a unique set of control variables. Our first step in this project is to focus on studies that provide correlation coefficients along with their more elaborated analyses. Here we found very small average correlation coefficients, presumably because these data do not control for other important variables such as students' prior achievement and students' SES. We are now in the process of examining other analytic techniques in search of a way to represent the effects of teacher's subject matter knowledge on student mathematics achievement when more complex and various statistical models are used.

## **Predictors of Science Achievement**

This project focuses specifically on studies that use students' science achievement scores as their outcome and that use primarily correlation coefficients as their analytic strategy. 11 reports are included in this study, all of conducted in the period from 1967 to 1995. In this study, teacher preparation refers to the number of hour of credits in science education or/and in specific science subjects, which teachers took in college. This study examines the average size of effect across these studies, the degree of homogeneity of their findings, and compares their predictor variables. So far we find heterogeneity among the correlations between teacher preparation and students' science achievement. The global effect size in correlation form is 0.13, which is significant. Similarly, teaching experience was correlated positively with students' science achievement. The corresponding value of the correlation coefficient was 0.15. Heterogeneity was also found among the correlations between teaching experience and students' science achievement. Four other mediators, sampling method, student achievement score type, teachers' science course type and teaching experience type, are also being used to explain the variability among the correlations in the these studies.

## **Three Hypotheses Regarding Teacher Education Curricula for Math Teachers**

This project recognizes three distinct hypotheses regarding the type of educational background that is most likely to lead to effective teaching practice. First is the “Bright, Well-Educated Person” hypothesis, which argues that the best teachers are those who attend relatively more selective schools and who receive a strong liberal education while there. The second hypothesis argues for strong content knowledge but takes a more vocational stance than the first, arguing that teachers cannot teach what they do not know. The third hypothesis is that teachers need professional knowledge—knowledge about schools, students and families, about how to manage classrooms, motivate students and foster learning. These hypotheses suggest different curricula for college students intending to teach. Focusing only on the teaching of mathematics, this project synthesizes research that examines teachers’ college course taking including their majors and minors, and the status of the institutions they attended. The goal is to see whether the extant evidence offers more or less support for any of these three hypotheses.

## **Synthesis of Regression Studies**

Two methodological projects investigate the use of regression results in meta-analysis.

(A) *Methods for combining slopes.* This project focuses on regressions in which the outcomes of interest are slopes for continuous variables (such as years of teaching experience, teacher test scores, and the like). One paper listed below has been sent to a journal, and several presentations on the topic have been made by Becker (August 2004) or Wu and Becker together. The issue here is that complications arise when regressions are to be synthesized, primarily because different research studies typically do not estimate identical regression models. In such cases the slopes presented in different studies will not represent exactly the same relationships, so combinations across studies may not be advisable. It is not clear, however, whether this will cause major problems when study results are synthesized. It may be that summaries can proceed in some cases, particularly if the other predictors in the model are fairly independent of the focal predictor. So in addition to examining the methods that can be used to synthesize slopes, the team is considering what the limits are on combining slopes from different regression models across studies. As a part of this, Wu is about to propose a dissertation to investigate the use of missing-data analysis methods in the context of summarizing regression models.

(B) *Combining slopes and standardized-mean-difference effect sizes.* A second issue that has arisen as we have examined the literature on teacher qualifications is the question of how to combine results from comparison studies (often represented in terms of standardized mean differences) with results from regressions. This problem was first noted in our project in the literature on alternate routes to teacher certification. In this domain, many studies present analyses such as t-test comparisons or analyses of variance to examine differences in outcomes between teachers who have been certified via traditional programs and others from alternate-

routes programs (or with emergency certificates). In many studies the design is simple – two groups are compared, so an appropriate study index would be the standardized mean difference. But in others more complex multiple regressions are estimated where the type of certificate held by the teachers is represented by a series of dummy (yes/no) variables. In such cases the slopes for the dummy variables represent functions of the mean differences, but with other variables held constant. When only two groups (and one dummy variable) are involved, there is an algebraic relation between the slope for the dummy variable and the standardized mean difference. However when other variables are included the mean difference is a partial relation or partial mean difference, with the other regression predictors held constant. When several dummy variables are in the model then the slopes may be combined to represent simpler mean differences, but again if other predictors appear in the model these slopes will represent partial relations, holding other included variables constant. Currently our research involves both an algebraic investigation of the slope estimates and their relation to the simple zero-order mean differences represented in the common effect-size indices, and a simulation study that will address the question of when regression models become too complex to warrant combining effects estimated from partial slopes with simpler zero-order effect sizes.

## **Role of Study Quality Scores in Meta-Analysis: A Simulation Study**

When combining studies, the quality of the primary research being combined is critical, due to its potential influence on the overall outcome. Researchers have identified two general challenges in dealing with the quality of primary studies in meta-analysis. One is that no universally accepted tools exist to assess the quality of primary studies included in a meta-analysis . The other challenge, which is the main focus of this study, is the lack of statistical evidence about incorporating quality of primary studies in a meta-analysis. This project examines the impact of incorporating quality scores for primary studies in meta-analysis, based on a simulation in which the magnitude of the mean population effect-size and its variance, the number of studies, and within-study sample sizes are manipulated. The quality of the overall effect-size estimates is first evaluated. Then, the impact on the quality of the overall effect-size estimates of the number of studies, within-study sample sizes, the magnitude of population effect-size and its variance, and relationships between effect-size and quality score is examined. This study made clear the influence of quality weighting on effect-size estimates – in short, quality weighting adds uncertainty to average effects.

## **Papers and Presentations This Year**

Ahn, Soyeon (2005, February) Incorporating Quality Scores of Studies in Meta-analysis: A Simulation Study. Presented at the 5<sup>th</sup> annual meeting of the Campbell Collaboration, Lisbon, Portugal.

Becker, B.J., & Wu, M.-J. (under review). Multivariate meta-analysis: Contributions of Ingram

Olkin and new results for the synthesis of slopes. *Statistical Science*.

Becker, B.J., (2004, August). Multivariate meta-analysis: Contributions of Ingram Olkin (and various others). Paper presented at the annual meeting of the American Statistical Association, Toronto, Canada.

Becker, B.J. (2004, October). Personality (?) versus training: Meta-analyses of teaching quality. Paper presented at the Society for Multivariate Experimental Psychology, Naples, FL.

Becker, B.J. (2004, November). Teacher qualifications and the quality of teaching: Results from an ongoing research synthesis. Invited paper presented at the meeting of the Florida Educational Research Association, Tampa, FL.

Becker, B. J. And M. M. Kennedy (2005, February) Synthesizing across widely varying designs. Paper presented at the annual meeting of the Campbell Collaboration, Lisbon Portugal.

Colitti, M. (2005) Examining the Relationship between Teachers' Professional Examination Scores and Indicators of Teacher Quality. Presented at the Annual meeting of the American Educational Research Association, Montreal.

Kennedy, M. M. (Under review) The Value of Teacher Qualifications as seen through Qualitative Research. *Review of Educational Research*.

Metzger, S. (2005) Conceptualizations of teacher qualification: An overview of the past six decades. Submitted for presentation to the History of Education Association.