Relating Teacher Candidate Performance to their Students' Subject Specific Academic Achievement using TWS methodology

By

Mark J. Fenster

University of Massachusetts Lowell

Diane L. Judd

Valdosta State University

Paper presented at the annual meeting of the American Educational Research Association. New York, NY, March 26, 2008. The Lorne H. Woollatt Distinguished Paper Award winner from the Northeastern Educational Research Association. Session 43.087.

Address correspondence to Mark Fenster, The Graduate School of Education, University of Massachusetts Lowell, 523 O'Leary Library, 61 Wilder Street, Lowell, MA 01854-3051, e-mail: mjfenster@gmmail.com or fenster_21stcentury@hotmail.com,

Or Diane Judd, Department of Early Childhood and Special Education, Valdosta State University, Valdosta GA 31602-0092, e-mail: djudd@valdosta.edu.

Abstract

Teacher education preparation programs are under pressure from the public to provide evidence that their programs and teacher candidates are improving student achievement. However, the connections between teacher education preparation programs, teacher candidates' evaluation, and student achievement are often hard to disentangle from other educationally relevant effects. Teacher candidates were formally assessed by their university supervisor during their student teaching semester. Using Western Oregon University's teacher work sample methodology, we were able to assess the achievement of the teacher candidate's students. Results showed that students of teacher candidates' showed 52% gains in knowledge of learning goals. Additionally, improvement in teacher candidate's teaching ability as measured on an observation instrument was associated with higher gains by their students on the learning goals. Lastly, teacher candidates assessing their students' learning in the disciplines of science and math showed lower gains on the learning goals when compared to English/language arts, health or social studies.

Relating Teacher Candidate Performance to their Students' Academic Achievement using TWS methodology

Currently, teacher education preparation programs are under pressure from the public at the state and national levels to provide evidence that their programs and teacher candidates are improving student achievement (Levine, 2006). However, the connections between teacher education preparation programs, teacher candidates' evaluation, and student achievement are often complex. Teacher expertise is an influential factor in determining student achievement (Darling-Hammond, 1997; Darling-Hammond, Holtzman, Galin & Heilig, 2005; Gay, 2003, Delpit, 1995; NBPTS, 1999). However, student achievement is not limited to teacher's abilities, but comes from a myriad of different places; students, home, school, peers, teachers and principals, to mention six pertinent examples (Hattie, 2003). Analyzing cumulative teacher effects in mathematics from grades 3 to 5, Sanders and Rivers (1996, p. 6) found that "groups of students with comparable abilities and initial achievement may have vastly different academic outcomes as a result of the sequence of teachers to which they are assigned". Wright, Horn and Sanders (1997) found the single most dominant factor affecting student academic *gain* is teacher effect.

After reviewing different value added methodologies, McCaffrey, Lockwood, Koretz, and Hamilton (2004) state:

> We cautiously conclude from our review of the literature that teachers differentially affect student achievement. Across diverse studies using different age cohorts, different models and statistical approaches, and different types of achievement measures, the studies all find nonzero teacher effects. (p. 113)

Hattie (2003) goes on to conclude that the answer to improving student achievement:

> Lies in the person who gently closes the classroom door and performs the teaching act-the person who puts into place the end effects of so many policies, who interprets these policies, and who is alone with students during their 15,000 hours of schooling (p. 3).

This paper is an analysis of those people who are in preparation to gently close that classroom door and teach. The purpose of the study was to determine; (1) did pre-k-5 students achieve the learning goals set by the teacher candidates, and (2) were improving scores on the teacher observation instrument related to higher rates of pre-k-5 student achievement?

Theoretical Framework

The shift to standards and performance based evaluations provided data and evidence for teacher education program research. The teacher candidates going through this program developed teacher work samples during their student teaching, conceptually similar to the Western Oregon University model and supported by Renaissance Group 2 institutions (Pankratz, 1999; Schalock, 2002; Schalock, & Myton, 2002; Wright, 2002). The teacher work samples contained information on the student achievement of the teacher candidate's students with respect to learning goals, achievement, and analysis of students' achievements presented by the teacher candidates.

The conceptual framework for the teacher candidate observation instrument was developed by a collaboration of a team from three universities and the assistance of a systemic teacher education program (STEP) grant. This conceptual framework is

aligned to state and national standards including Interstate New Teacher Assessment and Support Consortium (INTASC) (1992) and National Board for Professional Teaching Standards (NBPTS) (1998). On-line supporting resources for the framework can be found at: http://www.teachersbridge.org. Teacher candidates were evaluated on 13 indicators. The instrument can be found in Appendix A.

The importance of evaluating teacher candidates on their students' learning

Standards-based teaching and learning, as in instruction-based systems, requires teacher candidates to show high performance levels in the knowledge of the content they teach and pedagogy. However, a standards-based program also requires evidence that graduates can positively impact student learning (Pankratz, 1999). The Program Standards for Elementary Teacher Preparation document developed by the National Council for the Accreditation of Teacher Education (NCATE) describes four major attributes for performance-based teacher preparation (National Council for the Accreditation of Teacher Education, 2000). They are; (1) teacher candidate knowledge, (2) teaching performances (methodology), (3) teacher dispositions (values and commitments), and (4) positive effects on student learning.

As Schalock, Schalock, and Myton (1998) noted:

> We believe that any quality assurance system for teachers must include demonstrable teacher effectiveness, as measured by the learning gains of students. A demonstrably effective teacher (in contrast to a teacher who is merely knowledgeable or skillful) is able to integrate and apply whatever knowledge and skills are needed to advance the learning of a particular

> group of students toward a particular learning goal under a particular set of conditions (resources, time, and so on) (¶ 7).

The current paper focuses on three of the four attributes for performance-based teacher preparation with particular attention to analyzing impacts of student learning following the Western Oregon University model of teacher work samples (TWS). Teacher dispositions were not dealt with in this paper.

Pankratz (1999) points out the Western Oregon University's program requires teacher candidates to complete ten teaching tasks, each having a required product that has a performance measure: (1) describing a unit of study; (2) mapping the classroom context; (3) identifying learning outcomes; (4) developing assessments for outcomes; (5) administering pre-instruction assessments; (6) developing a design for instruction and assessment for all pupils; (7) implementing the instructional plan; (8) administering the post-instructional assessment; (9) summarizing, interpreting, and reporting the growth of each pupil and selected groups of pupils in the class; (10) and reflecting and evaluating the teaching and learning process for the instructional unit. Pankratz notes that,

> At Western Oregon, a foundation for the knowledge, skills, and processes required for the ten tasks is provided in the early years of the program. The set of ten tasks is completed for two separate units of two to five weeks of instruction during their student teaching experience. The work sample methodology provides direct evidence of a teacher candidate's effect on student learning in a relatively short time period and clearly connects the elements of standards-based teaching and learning. (¶ 29-30).

These ten tasks were components incorporated into the early childhood candidate preparation program evaluated in this paper.

Examples of learning goals taken from the teacher work sample

The Early Childhood Education Program TWS includes prompts to assist the teacher candidates in the development of their teacher work samples. The prompts for the learning goals explain to the teacher candidates that the learning goals will guide the planning, teaching, and assessment of their unit. The learning goals should describe what the teacher candidate expects their students to know and do at the end of the unit. In addition, the learning goals should be significant, challenging, varied, and appropriate. Table 1 shows examples of learning goals taken from five TWS.

Table 1

Examples of learning goals taken from five teacher work samples

Teacher Candidate	Subject and Grade	Learning Goals
1	Science 1st grade	Topic: Animals Learning Goal 1: Student will be able to identify and describe the characteristics of the following animal groups: mammals, reptiles, amphibians, birds, fish, insects, and spiders. Learning Goal 2: Student will be able to classify animals into specific animal groups.
2	Science 1st grade	Topic: The Human Body Learning Goal 1: Student will identify the parts and major function of the skeletal system.

		Learning Goal 2: Student will identify and describe types of germs, the sicknesses they cause, and how sicknesses are passed from one person to another.
3	Mathematics 4th grade	Topic: Fractions Learning Goal 1: Student will relate models to fractions, read and write fractions, and use fractions to represent part of a group. Learning Goal 2: Student will be able to show and use equivalent fractions.
4	Social Studies 4th grade	Topic: The Revolutionary War Learning Goal 1: Student will identify the source of dissatisfaction that led to the American Revolution. Learning Goal 2: Student will describe the contributions of key individuals in the American Revolution including Colonial and British men, women, and minorities.
5	Social Studies K grade	Topic: Transportation Learning Goal 1: Student will recognize different modes of transportation. Learning Goal 2: Students recognize transportation used for safety purposes.

Evidence that teacher work samples impact pre-k through 12 student learning

There has been a growing body of research showing a positive relationship between teacher expertise and student achievement (Darling-Hammond, Holtzman, Gatlin, & Heilig, 2005; Darling-Hammond, 2000; McCaffrey, Lockwood, Koretz, & Hamilton

2004; Sanders & Rivers, 1996). In every teacher field studied, McRobbie (2001) found stronger preparation resulted in greater success with students. However, the relationship between TWS and student achievement was found to be more ambiguous. A major goal of the teacher work sample was to "connect teacher performance to its impact on student learning" (Denner, Norman, Salzman, & Pankratz, 2003, p. 24). Studies have found that teacher candidates who scored well on the TWS provided better evidence of their impact on student learning than those who did not score well (Denner, et al.), but the relationship between performance on the TWS and evidence that learning goals were achieved was less clear and statistically insignificant (Denner et al.).

Hypotheses

Given the ambiguous relationship found in prior studies between TWS and learning goals we hypothesized:

H_1: Scores on a teacher observation instrument were related to gains in learning goals.

Methods

Teacher candidates were formally assessed by their university supervisor three times during their student teaching semester; early during the period (usually in the second or third week), during the middle (six to eight weeks) and at the end (weeks twelve to thirteen). These supervisor ratings provided the basis to look at initial ratings (the baseline) and any change from those rating during the student teaching experience.

Participants. All teacher candidates in the student teaching component of an early childhood and reading program during the 2005-2006 academic year participated in the study (n=103). Four students were not included in the study due to extensive missing data in teacher

observation or the learning goals, giving us 99 participants in the study. The 99 teacher candidates were student teaching in grades pre-k to 5.

A pre-test and post-test was given to pre-k to grade 5 students on all learning goals for each unit taught by the teacher candidate. After collecting data from the teacher work sample on student achievement (the pre-test and post-test) we then looked at the relationships between the supervisors' evaluations of teacher candidates' work and the student achievement of those teacher candidates' students.

The teacher candidates have primary responsibility for teaching at least three weeks during their student teaching experience. They gradually pick up subjects one at a time from the mentor teacher, building to the three week period when they have primary responsibility for teaching. At the end of the three week period, the teacher candidate gradually gives back subjects to the mentor teacher. The student achievement analysis that follows came from the three week period where the teacher candidates had primary responsibility for teaching the class.

Change Scores and the problem of the measurement of change

The hypothesis relates the performance of the teacher candidate on the observation instrument to *gains* in their students' learning. There has been considerable controversy in the literature regarding the best way to measure change. As Kissane (n.d.) noted, researchers have been able to identify and articulate the problems with change scores better than they have been able to overcome many of those problems satisfactorily. Change scores were once seen as unreliable from a measurement perspective (Cronbach & Furby, 1970). More recent research found the reliability of change scores was not a serious problem (Llabre, Spitzer, Saab, Ironson & Schneiderman, 1991; Williams &

Zimmerman, 1996). Typically, in change scores, the size of the base is negatively related to the change. If scores are high with a pre-test, the change score based on the pre-test is likely to be low. Conversely, if scores are low with a pre-test, the change score based on the pre-test is likely to be high.

In order to analyze the changes in learning goals to the teacher observation instrument, we require a formula to percentage change in our dependent variable, while at the same time showing some sensitivity to the size of the baseline score. Moran and Fenster (1982, p. 462) point out that

> The standard change score formula [Percentage change = a (t) – a (t-1) / a (t-1)] is unsatisfactory for two reasons. The magnitude of a measure thus derived is sensitive to the directionality of change…Second, this formula results in percentage changes that are overly sensitive to the size of the initial base; that is generally, the greater the size of the initial base, the smaller the change and vice versa. An alternative formula,
> Percentage change = [a (t) – a (t-1) / (a (t-1) + a (t))]/2
> is superior in two respects. First it obviates the magnitude/directionality problem: It produces the same magnitude percentage change figure for both positive and negative changes… The denominator is divided by 2 to return the percentage change figure to scale, since the appropriate base is deemed to be a point midway between the initial and subsequent base. Second, it minimizes (although it does not eliminate) the problem of initial base size by using both bases in the denominator (p. 426)

In the example taken from the present research, using the standard percentage change formula, an increase in the percentage of students knowing learning goals from a value of 50% to a value of 75% yields a figure of +50%, while a decrease from a value of 75% to a value of 50% yields a figure of -33%. Using the second percentage change formula [percentage change = a (t) – a (t-1) / (a (t-1) + a (t))/2], an increase in the percentage of students knowing learning goals from a value of 50% to a value of 75% yields a figure of +40%, while a decrease from a value of 75% to a value of 50% yields a figure of -40%.

Both the standard and the second percentage change scores were taken from two alternative elasticity formulas commonly used in micro-economics (Peterson, 1976).

Teacher observation instrument. The university supervisor evaluated the teacher candidate using a teacher observation instrument (see appendix A). Teacher candidates were evaluated on 13 dimensions using a four point scale where a score of "1" was low and "4" was high. The ordinal level scale was treated as interval level data (Labovitz, 1970). Thus, a teacher candidate averaging a 4.0 on the instrument received the maximum possible score on all 13 indicators.

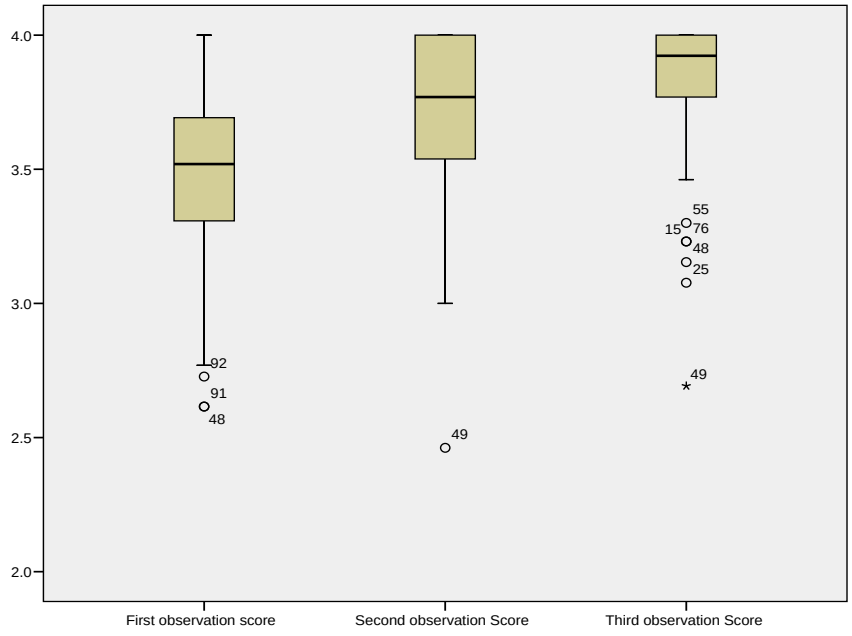
Results

Teacher candidate performance

Teacher candidate performance improved from the first to second observation, and improved again from the second to third observation. At the first observation (the baseline), the mean overall rating on the teacher observation instrument was 3.48 (out of a possible 4.00) and 3 (3%) of teachers were evaluated with the top score on all indicators of the instrument. At the second observation, the mean overall rating on the teacher

observation instrument increased to 3.73 and 26 (26%) of teachers were evaluated with the top score on all indicators of the instrument. At the third observation the mean overall rating on the teacher observation instrument was 3.82 and 41 (41%) of teachers candidates were evaluated with the top score on all indicators of the instrument. Figure 1 summarizes this information in a box plot.

Figure 1: Teacher candidates' scores on the observation instrument



However, ratings for the teacher candidates did not always increase. There were seven teachers whose overall ratings dropped between the 1st and 2nd administrations of the instrument and ten teachers whose overall rating dropped between the 2nd and 3rd administrations.

Table 2 shows the performance of the teacher candidates at baseline, the first observation, taken in the second or third week of student teaching on the 13 indicators assessed by the observation instrument. Teacher candidates were performing best on the subject specific content classroom indicator (3.70) and worst on the communication indicator (3.07) on a scale where a code of "1" was the lowest possible score and a code of "4" was the highest possible score.

Between the first and third observations (baseline and final) teacher candidate performance improved on all 13 indicators measured on the instrument. The biggest overall improvement came from the indicator assessing communication, the indicator they were weakest in at baseline. The smallest gain came from the indicator assessing subject specific content/concepts, the indicator they were strongest on at baseline (Table 3 shows the gains on all 13 indicators).

Table 2

Baseline scores on Teacher Observation Instrument

Indicators	Mean score at Baseline
Subject Specific Content/Concepts	3.70
Reflection	3.69
Students' Development	3.67
Classroom Environment	3.60
Assessment	3.60
Monitoring and Adjustments	3.57
Resources and Technology	3.52
Instructional Strategies	3.52
Classroom Management	3.51
Pedagogical Content	3.47
Lesson Plan and Instruction	3.45
Content Connection	3.28
Communication	3.07

Table 3

Improvement on Teacher Observation Instrument, first to last administration

Indicators	Mean Gain time 1 to time 3
Communication	0.55
Content Connection	0.45
Lesson Plan and Instruction	0.43
Instructional Strategies	0.40
Resources and Technology	0.39
Monitoring and Adjustments	0.31
Classroom Environment	0.29
Reflection	0.27
Assessment	0.26
Classroom Management	0.25
Students' Development	0.24
Pedagogical Content	0.18
Subject Specific Content/Concepts	0.15

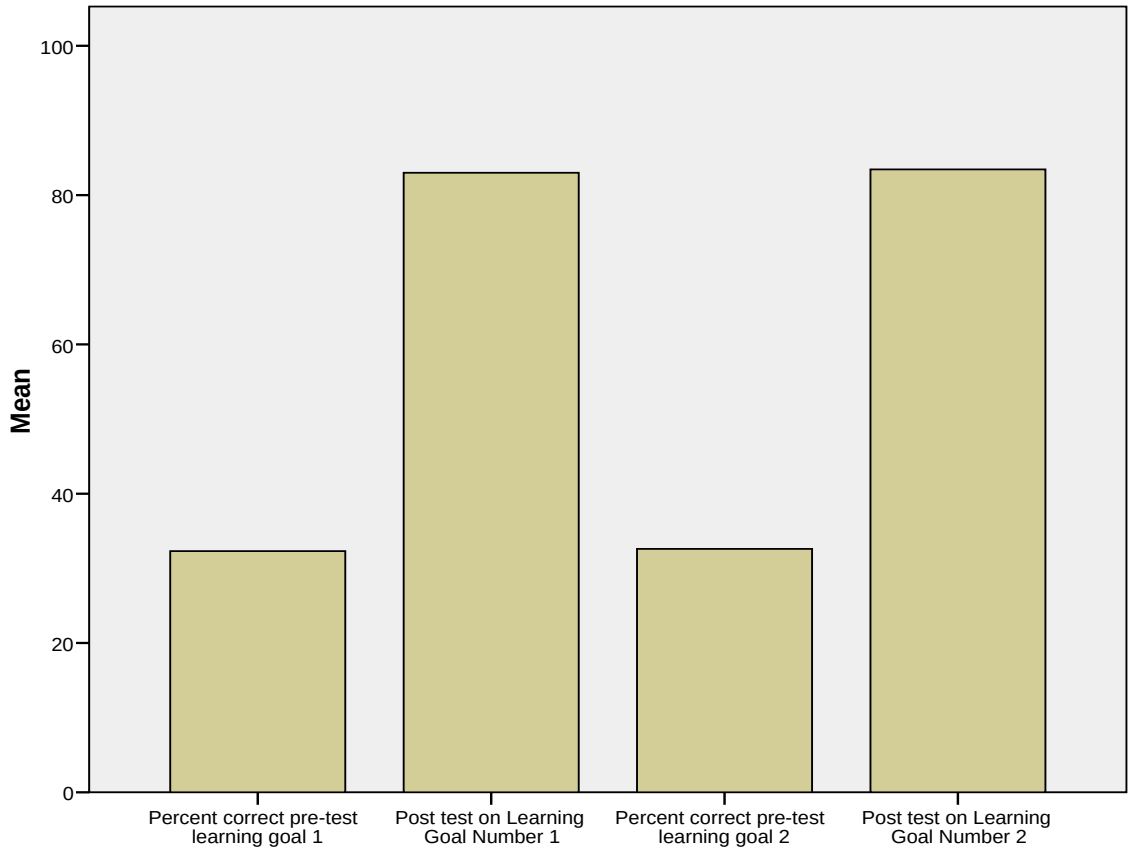
Analysis on student achievement

Teacher candidates identified learning goals and gave a pre-test to their students on the learning goals they were to cover in their class, as shown in figure 2. On learning goal number 1, the average percentage of correctly answered questions was 32.0%. On learning goal number 2, the average percentage of correctly answered questions was 32.6%.

At the end of the unit, teacher candidates gave an assessment on the material they covered, collecting data on the percentage of students' in their classes exhibiting knowledge with respect to their chosen learning goals. On learning goals number one and number two, the average percentage of correctly answered questions on the post-test was 82%.

Figure 2

Percentage of students giving correct answers on the learning goals, pre-test and post-test



Bivaraite analysis, the relationship between teacher candidate performance and pre-k to 5 student achievement

We see teacher candidate observation scores increasing and student achievement improving. We are now in a position to test our core research hypotheses, the relationship between teacher candidate observation scores and gains in student achievement.

Test of hypothesis scores on a teacher observation instrument were related to gains in learning goals. Table 4 presents the test of our hypothesis. The correlation between performance on the teacher observation instrument and the learning goal number one and two change scores was positive in all six cases, and statistically significant in five out of six instances. Effect sizes ranged from moderate to small (Cohen, 1988). Thus, we find some support for H_1.

Table 4

Correlations between scores on the teacher observation instrument and gain scores on the learning goals

Observation Instrument	Learning Goal Number One Gain Score
First observation Score	.24*
Second observation Score	.30*
Third observation Score	.24*
Observation Instrument	Learning Goal Number Two Gain Score
First observation Score	.22*
Second observation Score	.21*
Third observation Score	.12

Note: *=prob. <.05

Lastly, we extend the hypothesis test to look at whether gains in the learning goals were similar across disciplines. Teacher candidates chose the learning goals. They were strongly urged to choose one discipline for those learning goals (mathematics, science, social studies, English/language arts, or health), and all the teacher candidates limited their learning goals to one discipline. Table 5 presents the results of a linear multiple regression equation predicting to gains on learning goal number one. The independent variables hypothesized to predict to learning goal gains were; scores on the second observation score (the observation most closely aligned with the teacher candidate having responsibility for teaching all the classes' subjects), and two disciplines that researchers have identified as being more difficult for students to master; science and mathematics (Vygotsky, 1978).

Table 5 shows that each increase of one point on the observation instrument predicts a .26 increase in the gain score for learning goal one. We find that the mean gain in learning goal number one decreased by .33 when the teacher candidates' used science and decreased by .19 when teacher candidates used mathematics as the discipline for their learning goals. The three predictor variables were each statistically significant at $p<.05$. R^2, a measure of the overall fit of the equation was .22.

Table 5

Linear multiple regression equation predicting to gains on learning goal one

Independent Variable	Unstandardized Coefficients		t	Significance
	B	Std. Error		
(Constant)	-.367	.394	-.930	.355
Second observation Score	.259	.104	2.485	.015
Learning goal science	-.332	.091	-3.664	.000
Learning goal math	-.194	.089	-2.172	.033

(N=86)

Discussion

Teacher candidates can be part of the 30% of the variance in student achievement attributed to teachers by Hattie (2003). We found that teacher candidates had large and positive learning effects on their elementary school students in a variety of content areas. The pre-tests showed that students came into the content units with little knowledge of the material (32%) but showed considerable gains at the end of the unit (gain scores of 52%). Additionally, we found that those teacher candidates that exhibited higher quality teaching as measured by the observation instrument were associated with higher levels in their student's achievement on the learning goals. Thus, we find, the relationship between TWS and student learning gains were positive and generally statistically significant, a finding contrary to Denner et al. (2003). One possible explanation for the difference in the studies is the number of case and the resulting statistical power to reject the null hypothesis. Denner et al. analyzed ten TWS, we analyzed 99. If we had analyzed 10 TWS, the correlations reported in Tables 4 would have been statistically insignificant. Lastly, this achievement effect took place during the three week window where the teacher candidate taught the class, showing the Western Oregon University TWS model can work logistically in early childhood teacher preparation programs with tight windows for assessment of student learning.

The very high percentage (nearly 84%) of k-5 students achieving the learning goals may mean that teacher candidates can conceivably set higher standards. The instructions on learning goals to the teacher candidate were the teacher sets significant, challenging, varied and appropriate learning goals. With 100% of students achieving these learning goals in some classes, perhaps it would be beneficial for the institution's early childhood and reading program to push teacher candidates to set an even higher bar with their students with more challenging

learning goals than they do now, and determine the percentage of students that can meet these higher bars with the TWS methodology. If the percentage of students showing learning gains drops precipitously, then additional adjustments can be made to the difficulty of the learning goals. Additionally, perhaps the program should consider proving additional discrimination with respect to quality teaching on the high end of the teacher observation instrument. With 41% of the teacher candidates topping out the observation instrument with scores of "4" on every indicator, there is a need for a more refined scoring system at the high end of the continuum. A score of "5" is built into the instrument, but a score of "5" is reserved for induction level and experienced teachers who are consistently exemplary over time; therefore, level 5 should not be used to evaluate teacher candidates during practica or student teaching (see Appendix A.)

We also found the results vary by subject matter. The gains in the learning goals for science and mathematics were significantly less than the other disciplines. Science and mathematics might be harder to learn in the elementary school than other disciplines as elementary students function mostly out of episodic long term memory and not semantic (conceptual) long term memory (Vygotsky, 1978). The variability in children making the switch from episodic long term memory to (conceptual) long term memory can be considerable; timing is not the same for all children, making it harder for teacher candidates to know how to best pitch the difficulty of the learning goals in disciplines like science and mathematics, when compared to English/language arts, social studies or health.

Limitations

The learning goals were defined by teacher candidates. The learning goals had to be aligned with the state's standards and be part of the curriculum in the student teacher's

classroom, but the student teacher was able to select learning goals from topics from any of the content areas; English/language arts, math, science, social studies, or health.

There has been some concern that teacher candidates might set easy to meet criteria on the post-test, thereby showing learning gains (Airasian, 1997). There were protections in the TWS methodology of this program that made it harder for teacher candidates to set easy to meet criteria on the post-test. First, if students scored high on the pre-test, the learning goals were modified under the direction of the university supervisor to make sure that the curriculum didn't cover material that the class already knew. While this requirement wasn't always met as evidenced by a few classes with over 70% of the students already exhibiting knowledge of the material before it was taught, most classes few students exhibited knowledge of the material at pre-test. Second, the teacher candidates had many hours of discussion with their mentor teacher and their university supervisor about the preparation and teaching of their unit during the developing stages of the unit. This may account for some of the high percentages of gains achieved by the elementary students on the learning goals.

If the TWS were a high stakes assessment for teacher candidates and such candidates had to demonstrate an impact on student learning, there may have been greater incentive for student teachers to choose easy to meet learning goals. Teacher candidates in this study were required to produce a TWS, but the TWS was not a stakes component in their certification process. The teacher candidates did not have to show gains in knowledge on the learning goals in order to be certified. Thus, there were no incentives in the TWS methodology to reward teacher candidates for setting easy to achieve learning goals.

We recognize some additional major limitations of the present study. We had ceiling effects with both the teacher observation instrument and the learning gain scores. At the third

(and final) rating, 41% of the teacher candidates topped out scores on the observation assessment. Additionally, about one-fifth of the classes topped out the learning goal gain measure, many classes improving from 0% to 100%. The ceiling effects on both variables could have adversely affected the strength of some of the reported correlations. We note here that non-parametric correlation (the rank order correlation) showed virtually the same correlations as the Pearson Product Moment Correlations, so the assumptions made to use a measure of association based on an interval level scale on this data did not impact the substantive conclusions of this paper (Labovitz, 1970).

Looking at the distributions showing gains on the learning goals, we can conclude that the TWS is a promising path for colleges of education to use in order to collect evidence that their programs and, especially importantly, teacher candidates, are improving student achievement in the schools. As Levine (2006, p. 105) notes, "The job of a teacher education program is to prepare teachers who can promote student achievement". The TWS gives Colleges of Education *some* evidence on this important question. At a minimum, the TWS emphasizes the importance of analyzing student learning to future teachers, university supervisors and program developers.

Lastly, the increases in leaning gain scores reported here was all short term. Pankratz (1999, ¶ 46) cautions us, "what can be expected in a given time period? Are the conditions different for short-term progress than for long-term gains? Can short-term gains be a valid measure of teacher performance"?

References

Airasian, P. W. (1997). Oregon teacher work sample methodology: Potential and problems. In J. Millman (Ed.). *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 46-52). Thousand Oakes CA: Corwin Press.

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

Cronbach, J. & Furby, L. (1970). How should we measure change - Or should we? *Psychological Bulletin, 105,* 68-80.

Darling-Hammond, L. (1997). *The right to learn: A blueprint for creating schools that work.* San Francisco: Jossey-Bass, Publishers.

Darling-Hammond, L. (2000). *Solving the dilemmas of teacher supply, demand, and standards: How we can insure a competent, caring and qualified teacher for every child*. New York, NY: National Commission on Teaching and America's Future.

Darling-Hammond, L., Holtzman, D. J., Gatlin, S. J., & Heilig, J. V. (2005). Does teacher preparation matter? Evidence about teacher certification, Teach for America, and teacher effectiveness. *Education Policy Analysis Archives, 13*(42). Retrieved February 21, 2006 from http://epaa.asu.edu/epaa/v13n42/.

Delpit, L. (1995). *Other people's children: Cultural conflict in the classroom.* New York: The New Press.

Denner, P. R., Norman, A. D., Salzman, S. A., & Pankratz, R. S. (2003). Connecting Teaching Performance to Student Achievement: A Generalizability and Validity Study of the Renaissance Teacher Work Sample Assessment. Paper presented at the annual meeting of the Association of Teacher Educators, Jacksonville, FL, February 17, 2003.

Gay, G. (Ed.). (2003). *Becoming multicultural educators: Personal journey toward professional agency.* San Francisco: Jossey-Bass.

Hattie, J. (2003). *Teachers make a difference. What is the research evidence?* Retrieved January 27, 2006 from http://www.acer.edu.au/workshops/documents/Teachers_Make_a_Difference_Hattie.pdf.

Interstate New Teacher Assessment and Support Consortium (1992). *Model standards for beginning teacher licensing and development: A resource for state dialogue.* Washington, DC: Author.

Kissane, B. V. (n.d.). *The measurement of change as the study of the rate of change*. Retrieved January 27, 2006 from http://www.rasch.org/erp5.htm.

Labovitz, S. (1970). The assignment of numbers to rank order categories. *American Sociological Review, 35,* 515-524.

Levine, A. (2006). *Educating School Teachers*. Retrieved September 20, 2006 from http://www.edschools.org/pdf/Educating_Teachers_Report.pdf.

Llabre, M. M., Spitzer, S. S., Saab, P. G., Ironson, G. H., & Schneiderman, N. (1991). The reliability and specificity of delta versus residualized change as measures of cardiovascular reactivity to behavioral challenges. *Psychophysiology, 28,* 701-711.

McCaffrey, D. C., Lockwood, J. R., Koretz, D. M., & Hamilton, L. S. (2004). *Evaluating Value-Added Models for Teacher Accountability*. RAND Corporation: Santa Monica, CA.

McRobbie, J. (2001). *Career-long teacher development. Policies that make sense*. San Francisco, CA: West Education.

Moran, J., and Fenster, M. 1982. Voter Turnout in Presidential Primaries: A Diachronic Analysis. *American Politics Quarterly 10*:453-76.

National Board for Professional Teaching Standards (1998). Southfield, MI: Author.

National Board for Professional Teaching Standards (NBPTS). (1999). *What teachers should know and be able to do.* Washington, DC: NBPTS.

National Council for the Accreditation of Teacher Education. (2000). *Program Standards for Elementary Teacher Preparation*. Washington, DC. (Author).

Pankratz, R. (1999). Becoming accountable for the impact of graduates on students and schools: Making operational the shift from teaching to learning. Paper presented at the annual meeting of The American Association of Colleges for Teacher Education, Washington, DC, February 24-27, 1999. Retrieved January 23, 2006, from http://fp.uni.edu/itq/Paper_Publication/Becoming_Accountable.htm

Peterson, W. (1976). *Principles of Economics: Micro.* Homewood, IL.: Richard D. Irwin.

Sanders, W.L. & Rivers, J.C. (1996). Cumulative and residual effects of teachers on future student academic achievement. Knoxville, TN: University of Tennessee Value-Added Research Center.

Schalock, D., Schalock, M., & Myton, D. (1998). Effectiveness—along with quality—should be the focus. *Phi Delta Kappan, 79* (6): 468-470.

Schalock, H. D. (2002). Teacher work sample methodology with a standards orientation. Pp. 33-64. In, *Connecting teaching and learning: A handbook for teacher educators on teacher work sample methodology*. G. R. Girod (Ed.). Washington, DC: American Association of Colleges for Teacher Education. (ERIC Document Reproduction Service No. ED 463282).

Schalock, H. D., & Myton, D. (2002). Connecting Teaching and Learning: An Introduction to Teacher Work Sampling. Pp. 5-32. In, *Connecting teaching and learning: A handbook for teacher educators on teacher work sample methodology*. G. R. Girod (Ed.). Washington, DC: American Association of Colleges for Teacher Education. (ERIC Document Reproduction Service No. ED 463282).

Vygotsky, L.S. (1978). *Mind and society: The development of higher mental processes*. Cambridge, MA: Harvard University Press.

Williams, R. H. & Zimmerman, D. W. (1996). Are simple gain scores obsolete? *Applied Psychological Measurement, 20,* 59-69.

Wright, D. M. A. (2002). Teacher work sample methodology in early childhood and elementary preparation: A case study. Pp. 361-374. In, *Connecting teaching and learning: A handbook for teacher educators on teacher work sample methodology*. G. R. Girod (Ed.). Washington, DC: American Association of Colleges for Teacher Education. (ERIC Document Reproduction Service No. ED 463282).

Wright, S., Horn, S., & Sanders, W. (1997). Teacher and Classroom Context Effects on Student Achievement: Implications for Teacher Evaluation. *Journal of Personnel Evaluation in Education 11*, 57-67.

Appendix A

Teacher Candidate Observation Instrument
XYZ College of Education

Candidate:_____Observer:_____
School:_____ Subject/Grade Level:_____Date:_____
Directions**:** (a) Under each indicator, circle the professional practices that are observed, (b) list specific evidence observed for the indicator, (c) Under the # column, circle the number that reflects the practices observed for each indicator. Stronger performances should be rated the higher number within that level.

Note: Level 1 = Indicator Not Demonstrated
Level 2 = Indicator Partially Demonstrated
Level 3 = Indicator Adequately Demonstrated
Level 4 = Indicator Effectively Demonstrated

Level 5 = Indicator Exceptionally Demonstrated (Reserved for induction level and experienced teachers who are consistently exemplary over time; therefore, Level 5 should not be used to evaluate teacher candidates during practica or student teaching.)

I. CONTENT AND CURRICULUM: Teachers demonstrate a strong knowledge of content area(s) appropriate for their certification levels.		
Indicators/Professional Practices	**#**	**Evidence/Comments**
I-A. Subject–Specific Content/Concepts		
1-2. Uncorrected teacher content/concept errors; uncorrected student errors	1 2	
3-4. Shows knowledge of content/concepts; corrects teacher and student errors	3 4	
5. Accurate content/concept knowledge; consistently helps students recognize and correct errors	5	
Not Able To Rate	NATR	
I-B. Pedagogical Content (Instructional Methods)		
1-2. Uses inappropriate instructional method; little evidence of making content appropriate for diverse learners; lacks connections to students' prior knowledge	1 2	
3-4. Uses appropriate instructional methods; makes content appropriate for diverse learners; connects learning to students' prior knowledge	3 4	
5. Consistently uses a variety of appropriate instructional methods; consistently makes content appropriate for diverse learners; consistently connects learning to students' prior knowledge	5	
Not Able To Rate	NATR	
I-D. Content Connections		
1-2. Little or no evidence of making connections to other subject areas; little or no evidence of making content relevant to students' everyday lives	1 2	
3-4. Connects content to other subject areas; makes content relevant to students' everyday lives;	3 4	
5. Consistently connects content to other subject areas; consistently makes content relevant to students' everyday lives; affords students opportunities to make their own connections	5	
Not Able To Rate	NATR	

II. KNOWLEDGE OF STUDENTS AND THEIR LEARNING: Teachers support the intellectual, social, physical, and personal development of all students.		
Indicators/Professional Practices		**Evidence/Comments**
II-C. Students' Development		
1-2. Not responsive to the intellectual, social, physical, and personal developmental needs of all students	1 2	
3-4. Responsive to the intellectual, social, physical, and personal developmental needs of all students	3 4	
5. Consistently sensitive, alert, and responsive to the specific intellectual, social, physical, and personal developmental needs of all students	5	
Not Able To Rate	NATR	
III. LEARNING ENVIRONMENTS: Teachers create learning environments that encourage positive social interaction, active engagement in learning, and self-motivation.		
Indicators/Professional Practices		**Evidence/Comments**
III-B. Classroom Environment		
.1-2. Inefficient management of time, space, and learning resources for diverse students' learning; students not productively engaged	1 2	
3-4. Appropriate management of time, space, and learning resources for diverse students' learning; students actively engaged	3 4	
5. Consistent and appropriate management of time, space, and learning resources for diverse students' learning; active/equitable engagement of students	5	
Not Able To Rate	NATR	
III-C. Classroom Management		
1-2. Little or no evidence of a management plan; reactive classroom management style; behavior not monitored ; inconsistent/inappropriate responses to student behavior	1 2	
3-4. Follows classroom management plan; aware of student behavior; appropriate responses to student behavior; corrects misbehavior with minimal loss of instructional time	3 4	
5. Consistently follows classroom management plan; proactive classroom management style; subtle/preventative monitoring; fair, respectful responses to student behavior; students monitor/adjust own behavior when appropriate	5	
Not Able To Rate	NATR	
III-G. Communication		
1-2. Errors in spoken/written language; ineffective nonverbal communication; unclear directions; does not use effective questioning skills	1 2	
3-4. Error free spoken/written language; effective nonverbal communication; directions are clear or quickly clarified after initial student confusion; effective questioning and discussion strategies	3 4	
5. Consistently uses enriched/appropriate spoken/written language; effective nonverbal communication; effective questioning stimulates discussion in various ways	5	
Not Able To Rate	NATR	

IV. ASSESSMENT: Teachers understand and use a range of formal and informal assessment strategies to evaluate and ensure the continuous development of all learners.		
Indicators/Professional Practices		**Evidence/Comments**
IV-C Assessment		
1-2. Inappropriate or no assessment of instructional objectives/outcomes/essential questions; assessments do not align with the instructional objectives/outcomes/essential questions and lesson procedures.	1 2	
3-4. Uses appropriate assessments that align with the instructional objectives/outcomes/essential questions and lesson procedures	3 4	
5. Consistently uses a variety of authentic and traditional assessments that align with instructional objectives/outcomes/essential questions and lesson procedures; assessments are used to modify learning goals for individuals and groups	5	
Not Able To Rate	NATR	

V. PLANNING AND INSTRUCTION: Teachers design and create instructional experiences based on their knowledge of content and curriculum, students, learning environments, and assessments.		
Indicators/Professional Practices		**Evidence/Comments**
V-B. Lesson Plan and Instruction		
1-2. Lesson plan and instruction lack clear organization and sequence; inefficient pacing of lesson; instruction does not extend most students' understanding of concepts and/or content; components of the lesson plan are not aligned	1 2	
3-4. Lesson plan and instruction are logically organized and sequenced; pacing appropriate; instruction extends students' understanding of concepts and/or content; all components of the lesson plan are aligned	3 4	
5. Lesson plan and instruction consistently reflect findings from scientifically based research; appropriate organization and sequencing; appropriate pacing	5	
Not Able To Rate	NATR	
V-C. Instructional Strategies		
1-2. Inappropriate or no instructional strategies are used to engage and support learning; strategies inappropriately matched to subject matter	1 2	
3-4. Plans for and uses appropriate strategies that engage and support student learning; strategies appropriately matched to subject matter	3 4	
5. Consistently plans for and uses various strategies that engage and support diverse learners; provides multiple perspectives on key concepts, problems, and areas of knowledge	5	
Not Able To Rate	NATR	
V-D. Monitoring and Adjustments		
1-2. Does not monitor lesson or monitors lesson but adheres firmly to instructional plan; no adjustment for students who do not understand or who have already mastered the content	1 2	
3-4. Monitors lesson; makes appropriate modifications to instructional plans during the lesson to address students' needs; probes for understanding; uses students' questions to direct instruction	3 4	
5. Consistently monitors lesson and provides constructive and ongoing feedback; consistently and successfully makes modifications before and during the lesson to address student needs	5	
Not Able To Rate	NATR	

V-F. Resources and Technology		
1-2. Little evidence of using resources and materials other than assigned textbook and/or worksheets; technology is used superficially and does not enhance instruction	1 2	
3-4. Uses a variety of appropriate materials and resources; resources enhance instruction for diverse learners; uses technology effectively	3 4	
5. Consistently uses and monitors the effectiveness of a variety of appropriate materials and resources; resources consistently enhance instruction for diverse learners; students utilize resources, materials, and technology in their learning	5	
Not Able To Rate	NATR	
POST CONFERENCE		
VI. PROFESSIONALISM: Teachers recognize, participate in, and contribute to teaching as a profession.		
Indicators/Professional Practices		**Evidence/Comments**
VI-D. Reflection		
1-2. Does not examine his/her teaching; does not suggest modifications to improve teaching practices and student achievement	1 2	
3-4. Examines own teaching; suggests modifications that would lead to improved teaching practices and student achievement	3 4	
5. Consistently examines own performance in the classroom**;** provides evidence of modifying teaching practices to increase student achievement	5	
Not Able To Rate	NATR	
Comments:		

_____

Observer's Signature/Date Candidate's Signature/Date