The Renaissance Partnership Teacher Work Sample

2.

Evidence Supporting Score Generalizability, Validity, and Quality of Student Learning Assessment

Peter R. Denner, Antony D. Norman,
Stephanie A. Salzman, Roger S. Pankratz,
and C. Samuel Evans

Peter Denner is Assistant Dean of the College of Education and professor of educational psychology at Idaho State University. His current research interests are focused on standards-based performance assessments of teacher quality. He is both the institutional and assessment coordinator for the Renaissance Partnership work at ISU.

Antony "Tony" Norman is associate professor of psychology in the College of Education and Behavioral Sciences at Western Kentucky University (WKU). He serves as the WKU assessment coordinator for the Renaissance Partnership for Improving Teacher Quality. His scholarly interests include moral development, gifted education, and education reform.

Stephanie A. Salzman is Dean of the Woodring College of Education at Western Washington University. She was formerly the Associate Dean of the College of Education at Idaho State University, where she pioneered the introduction of teacher work sample assessment. Her scholarly interests include standards-based teacher education and assessment.

Roger Pankratz is director of the Renaissance Partnership for Improving Teacher Quality and professor of teacher education in the College of Education and Behavioral Sciences at Western Kentucky University. He is the former executive director of both the governor's Council on School Performance Standards and the Kentucky Institute for Education Research

(KIER). Under his direction, KIER published three annual reviews of research on the Kentucky Education Reform Act of 1990 and more than twenty statewide studies evaluating the results of school reform initiatives.

C. Samuel "Sam" Evans is the Associate Dean for Administration and Graduate Studies in the College of Education and Behavioral Sciences at Western Kentucky University. He serves as the WKU institutional coordinator for the Renaissance Project. His scholarship focuses on affective characteristics of effective teachers and teacher impact on student learning.

Abstract

The Renaissance Partnership Teacher Work Sample (RTWS) was investigated as an accountability measure for demonstrating teacher candidates' abilities to meet targeted teaching standards. The findings support the generalizability of the RTWS ratings. The results revealed high dependability coefficients for panels of three or more trained and experienced raters. Validity evidence for the RTWS was obtained using criteria suggested by Crocker (1997), including the frequency, criticality, necessity, and representativeness of the targeted teaching behaviors to actual teaching practice. The results also affirmed direct correspondence between the targeted RTWS tasks and seven of the ten Interstate New Teacher Assessment and Support Consortium (INTASC) standards. Finally, positive correlations between RTWS performances and independent ratings of the quality of learning assessments indicate that teacher candidates who score well on the RTWS provided better evidence of their impact on student learning than those who scored less well. Collectively, the findings demonstrate teacher work sample performance provides a credible means for teacher education programs to verify teacher candidate performance levels.

Based on the belief that quality teaching results in student achievement, a national trend to improve teacher quality has emerged. Prompted by major works, such as *A Nation at Risk* (National Commission on Excellence in Education, 1983), *Tomorrow's Teachers* (The Holmes Group, 1986), and *A Nation Prepared: Teachers for the 21st Century* (The Carnegie Forum on Education and the Economy, 1986), federal and state policy makers have turned their focus on teachers' ability to positively impact the learning of students. Teaching organizations such as the National Commission for Teaching and America's Future (1996), the National Education Association, and the American Federation of Teachers (Bradley, 1998) have followed suit.

At the same time, a growing body of research confirms the relationship between knowledge of teaching and learning acquired in teacher preparation programs and student achievement. In a study of 900 Texas school districts, Ferguson & Ladd (1996) reported a strong correlation between teacher expertise, measured by licensing exam scores, master's degrees, and years of experience, and student achievement. Other studies (Darling-Hammond, 2000; McRobbie, 2001; Sanders & Rivers, 1996) have reached similar conclusions. Furthermore, this connection persists even when taking into account student poverty and limited English proficiency, as well as selected school resource measures. In every teaching field, stronger preparation resulted in greater success with students and the increased likelihood of continuing in the teaching profession (McRobbie, 2001).

This evidence of the impact of teaching performance on student achievement has prompted various accrediting bodies to create more rigorous standards by which to judge teacher preparation programs and their candidates. Accordingly, the National Council of Accreditation of Teacher Education (NCATE, 2000) requires affiliate institutions to develop assessment systems that document teacher candidates' preparation to meet national or state standards and their impact on P–12 student learning.

In response to the coming changes in accreditation standards, a five year initiative by ten (now eleven) institutions, titled, "*Improving Teacher Quality through Partnerships that Connect Teacher Performance to Student Learning*" (Pankratz, 1999), began with the expressed purpose of advancing "a paradigm shift from a focus on the teaching process to learning results and connecting teacher performance to student learning" (p. 1). These institutions pledged to "implement programs and practices that build their capacity to be accountable for the impact of their teacher candidates and graduates on student learning" (p. 1). As a first action of the initiative, institutional representatives met and jointly identified seven teaching processes as essential to facilitating the learning of all students: (1) using contextual factors to plan instruction, (2) selecting learning goals, (3) developing an assessment plan, (4) designing instruction, (5) making instructional decisions, (6) analyzing student learning, and (7) reflecting on the teaching and learning process.

To measure teacher candidates' abilities regarding these processes, the partnership adapted the Western Oregon University Teacher Work Sample Methodology (Schalock, Schalock, & Girod, 1997). The result has been the development of the Renaissance Teacher Work Sample (The Renaissance Partnership for Improving Teacher Quality, 2001), which consists of seven performance tasks related to each of the above teaching processes. The Renaissance Teacher Work Sample (RTWS) requires teacher candidates to

produce a 20-page narrative, plus charts and attachments, that becomes a culminating teaching performance exhibit developed during student teaching (the RTWS assessment may be viewed at: *http://fp.uni.edu/itq/*). Central to this culminating performance is the requirement that teacher candidates demonstrate the end result of their teaching in terms of its impact on student learning. In addition, the partnership institutions collectively have developed scoring guides and rubrics to judge teacher candidates' level of performance on each of the seven teaching process standards, as well as their overall performance.

Although, as a measure of teaching standards, teacher work samples hold great promise, Denner, Salzman, and Bangert (2001) assert that this methodology is not without its critics. Important issues include the validity of teacher work samples as a measure of teaching performance standards and whether the degree of generalizability of scores derived from teacher work samples is sufficient for making high-stakes decisions regarding teaching performance levels with respect to those standards.

Investigation of these issues was the goal of three consecutive partnership meetings (June 2001, January 2002, and June 2002) that included multiple representatives from each of the eleven project institutions. The first purpose of our investigation was to determine score generalizability for the performance scores derived from each of the RTWS scoring rubrics when raters from across the partnership institutions evaluated RTWS performances. The second purpose was to investigate the validity of the RTWS as a measure of actual teaching behaviors and as a measure of national teaching standards. Our third purpose was to evaluate the degree to which performances on the RTWS provided quality assessment evidence for student learning.

Method

TEACHER WORK SAMPLE SETS

The teacher work samples (TWS) evaluated in this investigation were collected from across nine of the universities participating in the Renaissance Partnership to Improve Teacher Quality. The RTWS sets examined in this study were selected from two TWS collections: a collection of N = 110 TWS, gathered in June 2001, and a collection of N = 87 TWS, gathered in June 2002. Both collections contained TWS covering a broad range of subject areas and all grade levels from K to 12. Following a benchmarking process developed by Denner, Salzman and Bangert (2001), each TWS within each

collection was assigned to one of four categories along a developmental continuum from beginning to expert level performance. The benchmarking process is described later in the procedures section. After the benchmarking process, smaller sets (n = 10) of TWS were selected for scoring by groups of raters.

From the first RTWS collection, two exemplar sets of TWS, Set A and Set B, were developed. These sets consisted of 10 TWS each. The 10 TWS were chosen from the benchmarked exemplars of each performance category and were randomly assigned by category to the two sets (Set A and Set B). Each TWS exemplar set contained 2 Beginning, 3 Developing, 3 Proficient, and 2 Expert work samples. A third random set, Set C, was also complied containing the same breakdown of beginning to expert TWS. In a later phase of the investigation, a fourth 10 TWS set, Set D, was created from a random selection by category merger of the Set A and Set B work samples. From the second collection of TWS (N = 87) in June 2002, a fifth set, Set E, of 10 TWS was selected. The 10 Set E TWS were chosen at random by category after the entire collection of TWS had been organized into four categories from beginning to expert, following the same benchmarking process as had been used the previous year. Due to an incorrect identification of one of the TWS, the Set E TWS consisted of 1 Beginning, Developing, 4 Proficient, and 2 Expert TWS.

INSTRUMENTS

RTWS Scoring Rubrics. To rate each TWS, two rubrics were developed: the *RTWS Analytic Scoring Rubric* and the *RTWS Modified-Holistic Scoring Rubric.* The rubrics were based on the required components outlined in the RTWS Prompt and assessed the teaching process standards targeted by the RTWS assessment (to view the standards, RTWS Prompt, and analytic rubric go to: *http://fp.uni.edu/itq/*). Both the RTWS prompt and accompanying rubrics were collaboratively developed in an earlier three and a half day meeting of representatives from all partnership institutions. On the modified-holistic scoring rubric, each of the seven targeted standards for the TWS was rated on a 3-point scale: 1 = *Standard Not Met;* 2 = *Standard Partially Met;* and 3 = *Standard Met.* Summing across the seven standards, the total modified-holistic scores could vary from 0 to 21 points. On the analytic scoring rubric, the multiple targeted indicators for each standard were rated on a 3-point scale: 1 = *Indicator Not Met;* 2 = *Indicator Partially Met;* and 3 = *Indicator Met.* Across the seven standards, there were 32 total indicators; therefore, total analytic scores could vary from 0 to 96 points.

Validity Questionnaire. To gather validity evidence, a questionnaire asked a panel of raters (n = 42) about the alignment among the RTWS prompt, the targeted teaching processes (the RTWS standards), and the scoring rubrics on a four point scale: 1 = *Poor;* 2 = *Low;* 3 = *Moderate;* and 4 = *High.* In addition, we applied criteria suggested by Crocker (1997) for judging the content representativeness of performance assessments and scoring rubrics with regard to four criteria: (1) the frequency of the teaching behaviors in actual job performance, (2) the criticality (or importance) of those behaviors, (3) the authenticity (or realism) of the tasks to actual classroom practice, and (4) the degree to which the tasks were representative of the targeted standards. These criteria were rated using a four point scale from 1 = *Not at All* to 4 = *Very,* or in the case of the frequency criterion, a five point scale from 1 = *Never* to 5 = *Daily.* To assess evidence for validity of the RTWS requirements with regards to state and national teaching standards, we chose to focus on the INTASC standards (Interstate New Teacher Assessment and Support Consortium, 1992). The panel of raters were asked to indicate the extent to which the RTWS standards aligned with INTASC standards on a three point scale: 1 = *Not at All;* 2 = *Implicitly;* and 3 = *Directly.*

Quality of Learning Assessment Rating Scale. To independently assess whether RTWS performances reflected a robust representation of teacher impact on student learning that provided quality evidence for student learning, we developed a Quality of Learning Assessment (QLA) rating scale. The QLA scale focused on important criteria for sound student learning assessment, such as whether the learning goals reflected several types of learning and were significant and challenging (see Appendix). The criteria for judging the quality of assessments came from several contemporary textbooks on assessment (Chase, 1999; Gredler, 1999; Stiggins, 2001). Across the items, the criteria were rated as 0 = *Does Not Meet Criterion,* 1 = *Partially Meets Criterion,* or 2 = *Meets Criterion.* Summing the ratings across the items provided a total score. The original scale employed in June 2001 had only 10 items, so scores on the rating scale could vary from zero to 20. When used in June 2002, the scale was modified by the addition of two items. The added items were "assessments were congruent with the targeted learning goals in content and cognitive complexity" and "assessment directions and procedures are clear and would be understood by the students." Scores on the modified scale could vary from zero to 24. The Appendix A presents the full 12-item version of the QLA scale.

TEACHER WORK SAMPLE RATERS

In June 2001, a group of 36 raters from across the Renaissance Partnership institutions assembled in St. Louis, Missouri. The raters included administrators, teacher education faculty members, arts & sciences faculty members, and public school teachers from the regions served by the universities in the partnership. The raters were randomly assigned to groups of six raters to score the Set A, Set B, and Set C TWS using either the modified-holistic rubric or the analytic rubric. In January 2002, two additional groups of raters were selected from the 55 trained raters assembled in St. Louis. The raters for the Set D TWS consisted of 2 administrators, 6 faculty members, and 2 teachers. The ten Set D raters were selected on the basis of their approximation to a scoring criterion after a practice scoring session. The ten raters were randomly assigned by rater type (administrator, faculty member, or teacher) to two groups of 5 raters each. The two groups were then randomly assigned to scoring method (modified holistic versus analytic). In June 2002, six additional raters were asked to score the Set E TWS. The six Set E raters were all teacher education faculty members who had been nominated as experienced raters by their respective institutions.

PROCEDURES FOR SCORING THE TEACHER WORK SAMPLES

RTWS Rater Training. For all TWS raters, the training consisted of two hours of a review of the teaching processes and standards targeted by the RTWS assessment, examination of the relationship between the standards and the RTWS components, instruction on how to use the scoring rubrics to rate TWS performances, and anti-bias training (based on procedures described in Denner, Salzman & Bangert, 2001) during which raters completed a series of activities to uncover and create a reference list of potential sources of scoring bias.

RTWS Benchmarking. After training, groups of raters sorted the TWS gathered in each collection (June 2001 and June 2002) according to a set of holistic category descriptions. The categories described TWS performances along a continuum: 1 = *Beginning,* 2 = *Developing,* 3 = *Proficient,* and 4 = *Expert.* To accomplish this task, the raters were divided into cross-institutional groups of 4 raters each. Each group first performed a quick read of 15 to 20 percent of the work samples. When a group reached consensus on the holistic category, they placed the TWS in that pile. In the afternoon, a different mix of raters were grouped to examine the TWS within each category and to pick category exemplars. Following group discussion, four to six exemplars

of performance in each category were identified. As described previously, TWS sets were then created by either randomly assigning the exemplar TWS by category to the TWS sets (Set A and Set B) or selecting TWS at random from within each of the four-benchmark categories (Set C and Set E).

RTWS Scoring. At this stage, all raters scored their assigned set of TWS independently using their assigned scoring rubric (analytic or modified-holistic). As they scored, the raters continued to use their personal lists of biases to remind them to ignore these factors when scoring. They were exhorted to score the TWS on the basis of the standards and the scoring rubrics only. Across all TWS sets, the average grading time per TWS for raters using the analytic rubric was about 28 minutes. The average grading time per TWS for raters using the modified-holistic rubric was about 27.5 minutes.

Validity Ratings. The validity data were gathered in June 2002. The validity assessment panel consisted of 42 representatives from across the 10 partner-ship institutions. None of the validity assessment panel members had been involved in the TWS development process. Most of the panel members were faculty members from the partnership institutions who were being intro-duced to the RTWS assessment for the first time. The panel included a mix of administrators, faculty members, and public school teachers. The panel members had received training as RTWS raters (in the same manner as described previously) and had rated at least two work samples prior to com-pleting the validity questionnaire. All panel members independently com-pleted the sections of the content validity questionnaire.

PROCEDURES FOR QUALITY OF LEARNING ASSESSMENT

Expert Raters. Independent panels of experts consisting of 2 to 3 expert raters were asked to evaluate three sets of RTWS (Set A, Set B, and Set E) using the Quality of Learning Assessment (QLA) rating scale. All of the QLA raters had extensive backgrounds in testing and measurement. All were experienced in the development and use of scoring rubrics

Scoring Procedures. Following acquaintance with the RTWS assessment and full rater training, the QLA raters for this study received intensive training that focused on the QLA items and the possible locations and sources of evi-dence for each item within the various RTWS components. The raters reached consensus regarding key definitions and concepts embedded in the QLA items and practiced locating the evidence using an example TWS. The

QLA raters then independently scored their assigned set of n = 10 TWS. The raters averaged about 20 minutes per work sample to complete their QLA ratings.

Design. To evaluate the reliability of the scores from the RTWS rubrics, we employed a research design from Generalizability Theory (Shavelson & Webb, 1991). A single facet design was used to assess the effect of rater for both the modified-holistic and the analytic scoring methods. This design was analyzed separately for each of the RTWS sets using repeated measures ANOVA. The rater facet served as the repeated-measures factor in each case. Using variance component estimates generated from the ANOVA results, Generalizability Theory permits the calculation of two types of coef-ficients, depending upon whether the measure is to be used to make deci-sions about the "relative standing or ranking of individuals" or about "the absolute level of their scores" (Shavelson & Webb, 1991, p. 84). Because the RTWS was designed to measure teacher education candidates' abilities to meet the seven targeted teaching process standards (an absolute decision about performance levels with respect to the standards), the formula pre-sented by Shavelson and Webb (1991) for computing an index of depend-ability for absolute decisions was employed in this study. An index of dependability indicates the proportion of the score that can be generalized across the raters and reflects the performance level of the candidate. The same formula can be adjusted to provide information regarding the number of raters necessary for making high-stakes decisions about the absolute level of teaching performance of teacher candidates using the RTWS assessment.

Pearson product-moment correlation was used to correlate the RTWS scores with the QLA rating scores. All total scores on all measures were averaged across raters. Percentages were calculated for reporting the responses of the validity assessment panel to the content validity question-naire. For all statistical analyses, the level of statistical significance was set at $\square = .05$.

Results

SCORE GENERALIZABILITY

Effect for Raters across TWS Sets. As might be expected, for the initial groups of raters, who had received only minimal training, the effect for rater was found to be statistically significant across all three RTWS sets (Set A, Set B, and Set C). For the groups of six novice raters assigned to the modified-holistic scoring rubric, the effect for rater was statistically significant for the Set A TWS, $F(5, 45) = 6.11$, $MSE = 6.67$, $p < .001$, the Set B TWS, $F(5, 45) =$

3.85, $MSE = 8.18$, $p = .005$, and the Set C TWS, $F(5, 45) = 3.50$, $MSE = 6.20$, $p = .009$. Likewise, for the groups of six novice raters assigned to use the analytic scoring rubric the effect for rater was also statistically significant for the Set A TWS, $F(5, 45) = 4.17$, $MSE = 93.06$, $p = .003$, the Set B TWS, $F(5, 45) = 6.14$, $MSE = 39.78$, $p < .001$, and the Set C TWS $F(5, 45) = 6.00$, $MSE = 78.86$, $p < .001$. Seven months later, following better training, independent groups of five raters, who had been selected on the basis of their ability to meet a scoring criterion, also displayed a statistically significant effect for rater when scoring the Set D TWS for both the modified-holistic scoring rubric, $F(4, 36) = 3.89$, $MSE = 21.71$, $p = .01$, and the analytic scoring rubric, $F(4, 36) = 6.28$, $MSE = 59.21$, $p = .001$. Importantly, after one year, when the partnership institutions nominated six experienced raters to score the Set E TWS using the analytic scoring rubric, the effect for rater was not found to be statistically significant, $F(5, 45) = 1.07$, $MSE = 100.94$, $p = .39$. Together, these findings suggest rater experience may be an important factor influencing score consistency when cross-institutional raters are asked to assess complex teacher work sample performances.

Dependability Coefficients. Table 1 presents the variance components estimates derived from the ANOVA results used in the formulas for computing the dependability coefficients for each of the TWS sets and scoring methods. When groups of novice raters used the modified-holistic rubric to score the Set A, Set B, and Set C TWS, the results yielded six rater coefficients of dependability of .59, .77, and .71 respectively. For the analytic scoring rubric, the six-rater coefficients were computed to be .62 for Set A, .91 for Set B, and .64 for Set C. For raters who were given better training and who were selected on the basis of the degree of match of their practice scores with a scoring criterion, the five-rater coefficients of dependability for the Set D TWS were .74 for the modified-holistic scoring rubric and .88 for the analytic scoring rubric. For the experienced raters, who scored the Set E TWS using the analytic scoring rubric, the six-rater coefficient of dependability was computed to be .87. Together, these coefficients suggest a high proportion of the TWS score differences among teacher education candidates can be generalized across raters.

Adjusting the number of raters included in the formulas revealed that an acceptable level of dependability of .77 to .82 could be achieved with as few as three raters when using the analytic scoring rubric based on the results from the Set D and Set E TWS. Table 2 displays the dependability coefficient estimates for different numbers of raters by scoring method using the results obtained across TWS sets. Overall, the results indicate that scores on the Renaissance TWS performance assessment can be used to make decisions regarding the quality of teaching performance that can be generalized across raters when panels of three or more trained and experienced raters are used.

TABLE 1

Estimates of Variance Components for the Total Modified-Holistic Scores and the Total Analytic Scores across Teacher Work Sample Sets

Estimated Variance Components					
	Modified-Holistic Scores				
Source	Set A (6 raters)	Set B (6 raters)	Set C (6 raters)	Set D (5 raters)	
Person	2.37	5.81	3.13	4.82	
Rater	3.41	2.33	1.55	1.88	
Residual	6.66	8.19	6.20	6.39	
	Analytic Scores				
Source	Set A (6 raters)	Set B (6 raters)	Set C (6 raters)	Set D (5 raters)	Set E (6 raters)
Person	33.55	113.26	35.08	138.38	111.64
Residual	93.06	39.78	78.87	59.21	100.94

VALIDITY

Alignment. For the alignment between the TWS elements presented in the guidelines and the targeted standards, 78.6 percent ($f = 33$) of validity assessment panel members indicated a high degree of alignment, and 21.4 percent ($f = 9$) said moderate alignment. For the alignment between the TWS task elements and the analytic scoring rubric, 69 percent ($f = 29$) of the panel members said there was a high degree alignment, 28.6 percent ($f = 12$) said moderate alignments, and 2.4 percent ($f = 1$) said low alignment. For the alignment of the analytic scoring rubric with the targeted standards, 73.8 percent ($f = 31$) said there was high alignment, 23.8 percent ($f = 10$) said moderate alignment and 2.4 percent ($f = 1$) said low alignment.

TABLE 2

Dependability Coefficient Estimates by Number of Raters for Each Scoring Rubric

Dependability Coefficient Estimates

Number of Raters	Modified-Holistic Rubric				
	Set A	Set B	Set C	Set D	
6 Raters	.59	.77	.71	.78	
3 Raters	.41	.62	.55	.64	
1 Rater	.19	.36	.29	.37	
Number of Raters	Analytic Rubric				
	Set A	Set B	Set C	Set D	Set E
6 Raters	.62	.91	.64	.90	.87
3 Raters	.45	.85	.47	.82	.77
1 Rater	.22	.65	.23	.60	.53

Frequency. Table 3 presents the judgments made by the validity assessment panel with regard to how frequently they would expect a teacher to engage in the teaching behaviors targeted by the RTWS. All the teaching behaviors were considered to be high frequency activities for teachers with 83.3 to 100 percent of the raters indicating "weekly" or "daily" for all but one of the behaviors. The targeted teaching behavior that required teacher candidates to "use assessment data to profile student learning and communicate information about student progress and achievement" was rated "weekly" ($f = 20$) or "daily" ($f = 7$) by only 64.3 percent of the raters.

Criticality. To assess the *criticality* of the tasks performed while completing the RTWS, the validity assessment panel rated the importance of the teaching behaviors required. Table 4 presents the number and percent of the validity panel members indicating the importance to effective teaching (or criticality) of the teaching behaviors targeted by the Renaissance TWS. All

TABLE 3

Number and Percent of Panel Members Indicating How Frequently They Would Expect a Teacher to Engage in the Teaching Behaviors Targeted by the Teacher Work Sample (N = 42)

Teaching Behaviors Targeted By Teacher Work Sample	Never	Yearly	Monthly	Weekly	Daily
Use information about the learning-teaching context and student individual differences to set learning goals and plan instruction and assessments.		2 4.8%	5 11.9%	10 23.8%	25 59.5%
Set significant, challenging, varied, and appropriate learning goals.			5 11.9%	26 61.9%	11 26.2%
Use multiple assessment modes and approaches aligned with learning goals to assess student learning before, during, and after instruction.			2 4.8%	14 33.3%	26 61.9%
Design instruction for specific learning goals, student characteristics and needs, and learning contexts.			1 2.4%	19 45.2%	22 52.4%

(continues on next page)

TABLE 3 (continued)					
Teaching Behaviors Targeted By Teacher Work Sample	Never	Yearly	Monthly	Weekly	Daily
Use ongoing analysis of student learning to make instructional decisions.				7 16.7%	35 83.3%
Use assessment data to profile student learning and communicate information about student progress and achievement.		1 2.4%	14 33.3%	20 47.6%	7 16.7%
Reflect on instruction and student learning in order to improve teaching practice.		1 2.4%	5 11.9%	5 11.9%	31 73.8%

of the teaching behaviors were considered to be "important" or "very important."

Authenticity. The validity assessment panel judged how authentic the tasks required by the RTWS are to success as a classroom teacher. Table 5 presents the number and percent of the panel members rating each of the nine major TWS tasks as authentic. All tasks required by the RTWS were considered to be authentic or very authentic to success as a classroom teacher by a majority of the panel members. The percentages varied from 61.9 percent for (item # 8) "Teacher uses graphs or charts to profile whole class performance on pre-assessment and post-assessment, and to analyze trends or differences in student learning for selected subgroups" to 97.6 percent for (item #6) "Teacher uses on-going analysis of student learning and responses to rethink and modify original instructional design and lesson plans to improve student progress toward the learning goals(s)."

TABLE 4

Number and Percent of Panel Members Indicating the Importance to Effective Teaching of the Teaching Behaviors Targeted by the Teacher Work Sample (N = 42)

Teaching Behaviors Targeted By Teacher Work Sample	Degree of Importance			
	Not at all Important 1	Some-what Important 2	Important 3	Very Important 4
Use information about the learning-teaching context and student individual differences to set learning goals and plan instruction and assessments.			10 23.8%	32 76.2%
Set significant, challenging, varied, and appropriate learning goals.			4 9.5%	38 90.5%
Use multiple assessment modes and approaches aligned with learning goals to assess student learning before, during, and after instruction.			6 14.3%	36 85.7%
Design instruction for specific learning goals, student characteristics and needs, and learning contexts.			6 14.3%	36 85.7%

(continues on next page)

TABLE 4 (continued)				
	Degree of Importance			
Teaching Behaviors Targeted By Teacher Work Sample	Not at all Important 1	Some-what Important 2	Important 3	Very Important 4
Use ongoing analysis of student learning to make instructional decisions.			5 11.8%	37 88.1%
Use assessment data to profile student learning and communicate information about student progress and achievement.			12 28.6%	30 71.4%
Reflect on instruction and student learning in order to improve teaching practice.			4 9.5%	38 90.5%

Representativeness. The validity assessment panel also considered the degree to which the tasks required by the RTWS reflect and represent the targeted standards (See Table 7). Once again, the majority (88.1 to 97.6 percent) of the panel members thought the tasks were "representative" or "very representative" of the targeted standards, with most panel members indicating very representative (59.5 to 73.8 percent).

Match to INTASC Standards. Finally, the panel of experts indicated the extent to which the tasks required for the RTWS reflected the Interstate New Teacher Assessment and Support Consortium (INTASC) standards (Interstate New Teacher Assessment and Support Consortium, 1992). Although not directly designed to assess the INTASC standards, the teaching processes targeted by the RTWS are very similar to those addressed by

TABLE 5

Number and Percent of Panel Members Indicating How Authentic the Tasks Required by the Teacher Work Sample Are to Success as a Classroom Teacher (N = 42)

Tasks Required By the Teacher Work Sample	Degree of Authenticity			
	Not at all Authentic 1	Some-what Authentic 2	Authentic 3	Very Authentic 4
Teacher uses understanding of student individual differences and community, school, and classroom characteristics to draw specific implications for instruction and assessment.		3 7.1%	15 35.7%	24 57.1%
Teacher sets significant, challenging, varied and appropriate learning goals for student achievement that are aligned with local, state, or national standards.		4 9.5%	13 31.0%	25 59.5%
Teacher designs an assessment plan to monitor student progress toward learning goals, using multiple assessment modes and approaches to assess student learning before, during, and after instruction.		6 14.3%	13 31.0%	23 54.8%

TABLE 5 (continued)				
	Degree of Authenticity			
Tasks Required By the Teacher Work Sample	Not at all Authentic 1	Some-what Authentic 2	Authentic 3	Very Authentic 4
Teacher designs instruction aligned to learning goals and with reference to contextual factors and pre-assessment data, specifying instructional topics, learning activities, assignments and resources.		2 4.8%	17 40.5%	21 50.0%
Teacher designs instruction with content that it accurate, logically organized, and congruent with the big ideas or structure of the discipline.		2 4.8%	15 35.7%	25 59.5%
Teacher uses on-going analysis of student learning and responses to rethink and modify original instructional design and lesson plans to improve student progress toward the learning goal(s).		1 2.4%	17 40.5%	24 57.1%

TABLE 5 (continued)				
	Degree of Authenticity			
Tasks Required By the Teacher Work Sample	Not at all Authentic 1	Some-what Authentic 2	Authentic 3	Very Authentic 4
Teacher analyzes assessment data, including pre/post assessments and formative assessments, to determine students' progress related to the unit learning goals.		4 9.5%	13 31.0%	25 59.5%
Teacher uses graphs or charts to profile whole class performance on pre-assessments and post-assessments, and to analyze trends or differences in student learning for selected subgroups.	4 9.5%	12 28.6%	15 35.7%	11 26.2%
Teacher evaluates the effectiveness of instruction and reflects upon teaching practices and their effects on student learning, identifying future actions for improved practice and professional growth.		2 4.8%	15 35.7%	25 59.5%

TABLE 6

Number and Percent of Panel Members Indicating the Degree to Which the Tasks Required by the Teacher Work Sample Reflect and Represent the Targeted Standards (N = 42)

Tasks Required By the Teacher Work Sample	Degree of Representativeness			
	Not at all Repre-sentative 1	Some-what Repre-sentative 2	Repre-sentative 3	Very Repre-sentative 4
Teacher uses understanding of student individual differences and community, school, and classroom characteristics to draw specific implications for instruction and assessment.		2 4.8%	15 35.7%	25 59.5%
Teacher sets significant, challenging, varied and appropriate learning goals for student achievement that are aligned with local, state, or national standards.		1 2.4%	11 26.2%	30 71.4%
Teacher designs an assessment plan to monitor student progress toward learning goals, using multiple assessment modes and approaches to assess student learning before, during, and after instruction.		1 2.4%	10 23.8%	30 71.4%

TABLE 6 (continued)

Tasks Required By the Teacher Work Sample	Degree of Representativeness			
	Not at all Repre-sentative 1	Some-what Repre-sentative 2	Repre-sentative 3	Very Repre-sentative 4
Teacher designs instruction aligned to learning goals and with reference to contextual factors and pre-assessment data, specifying instructional topics, learning activities, assignments and resources.		2 4.8%	13 31.0%	27 64.3%
Teacher designs instruction with content that it accurate, logically organized, and congruent with the big ideas or structure of the discipline.		1 2.4%	14 33.3%	27 64.3%
Teacher uses on-going analysis of student learning and responses to rethink and modify original instructional design and lesson plans to improve student progress toward the learning goal(s).		1 2.4%	10 23.8%	31 73.8%

TABLE 6 (continued)				
	Degree of Representativeness			
Tasks Required By the Teacher Work Sample	Not at all Repre-sentative 1	Some-what Repre-sentative 2	Repre-sentative 3	Very Repre-sentative 4
Teacher analyzes assessment data, including pre/post assessments and formative assessments, to determine students' progress related to the unit learning goals.		2 4.8%	9 21.4%	30 71.4%
Teacher uses graphs or charts to profile whole class performance on pre-assessments and post-assessments, and to analyze trends or differences in student learning for selected subgroups.	2 4.8%	3 7.1%	12 28.6%	25 59.5%
Teacher evaluates the effectiveness of instruction and reflects upon teaching practices and their effects on student learning, identifying future actions for improved practice and professional growth.		1 2.4%	12 28.6%	29 69.0%

TABLE 7

Number and Percent of Panel Members Indicating the Teacher Work Sample Tasks Matched the Interstate New Teacher Assessment and Support Consortium Standards (N = 42)

INTASC Standards	Not at all	Implicitly	Directly
Knowledge of Subject Matter: The teacher understands the central concepts, tools of inquiry, and structures of the content area(s) taught and creates learning experiences that make these aspects of subject matter meaningful for learners.		13 31.0%	26 61.9%
Knowledge of Human Development and Learning: The teacher understands how students learn and develop, and provides opportunities that support their intellectual, social, and personal development.		16 38.1%	24 57.1%
Adapting Instruction for Individual Needs: The teacher understands how students differ in their approaches to learning and creates instructional opportunities that area adapted to learners with diverse needs.	1 2.4%	7 16.7%	32 76.2%
Multiple Instructional Strategies: The teacher understands and uses a variety of instructional strategies to develop students' critical thinking, problem solving, and performance skills.	1 2.4%	11 26.2%	28 66.7%

TABLE 7 (continued)			
INTASC Standards	Not at all	Implicitly	Directly
Classroom Motivation and Management Skills: The teacher understands individual and group motivation and behavior and creates a learning environment that encourages positive social interaction, active engagement in learning, and self-motivation.	10 23.8%	22 52.4%	8 19.0%
Communication Skills: The teacher uses a variety of communication techniques including verbal, nonverbal, and media to foster inquiry, collaboration, and supportive interaction in and beyond the classroom.	4 9.5%	26 61.9%	10 23.8%
Instructional Planning Skills: The teacher plans and prepares instruction based upon knowledge of subject matter, students, the community, and curriculum goals.		5 11.9%	35 83.3%
Assessment of Student Learning: The teacher understands, uses, and interprets formal and informal assessment strategies to evaluate and advance student performance and to determine program effectiveness.		4 9.5%	36 85.7%
Professional Commitment and Responsibility: The teacher is a reflective practitioner who demonstrates a commitment to professional standards and is continuously engaged in purposeful mastery of the art and science of teaching.		16 38.1%	24 57.1%

TABLE 7 (continued)			
INTASC Standards	Not at all	Implicitly	Directly
Partnerships: The teacher interacts in a professional, effective manner with colleagues, parents, and other members of the community to support students' learning and well being.	13 31.0%	19 45.2%	8 19.0%

many of the INTASC standards. Table 7 presents the number and percent of responses made by our panel of experts for each of the INTASC standards. The RTWS was seen by a majority of the experts to directly measure seven of the ten INTASC standards. As can be seen from Table 7, the highest rated were those INTASC standards most closely aligned with the seven teaching process standards targeted by the RTWS. Other INTASC standards were judged to be implicitly measured because knowledge and skills related to them might be used in completing a RTWS, even though indicators of these standards are not directly included in the Renaissance scoring rubrics. Of significance is the fact that three of the INTASC standards were not seen as measured by the RTWS and these standards were not targeted by the RTWS.

QUALITY OF LEARNING ASSESSMENT

Effect of Rater. Using repeated measures ANOVA, the effect of rater on the Quality of Learning Assessment (QLA) scores was not statistically significant for the Set A TWS, $F(1, 19) = .85$, $MSE = 5.89$, $p = .38$ or the Set E TWS, $F(2, 18) = .440$, $MSE = 8.40$, $p = .65$, but it was statistically significant for the Set B TWS, $F(2, 16) = 4.07$, $MSE = 5.80$, $p = .04$. The two-rater coefficient of dependability for the QLA scores for the Set A TWS was calculated to be .69. The three-rater coefficients of dependability for the QLA scores for the Set B and Set E TWS were calculated to be .71, and .84 respectively. Together, these findings suggest sufficient inter-rater agreement for the purpose of this investigation.

Correlation with Renaissance TWS Total Scores. Table 8 presents the correlations among the analytic total scores, modified-holistic total scores, and Quality of Learning Assessment (QLA) total scores for the Set A, Set B, and

TABLE 8

Correlations Among the Total Analytic Scores, Total Modified-Holistic Scores, and the Total Quality of Learning Assessment Scores

	Analytic	Modified-Holistic
Set A Teacher Work Samples (n = 10) Quality of Learning Assessment Analytic	.24	.48 .77*
Set B Teacher Work Samples (n = 10) Quality of Learning Assessment Analytic	.89*	.86* .89*
Set E TWS (n = 10) Quality of Learning Assessment	.70*	

*p < .05

Set E teacher work samples. All total scores were averaged across the raters of each set. As can be seen from Table 8, the correlations for the Set B TWS and Set E TWS were positive and high. These correlations indicate a strong positive relationship between total work sample performance as measured by the analytic and the modified-holistic rubrics and the total scores on the Quality of Learning Assessment measure. Together, these data support the idea that teacher education candidates who scored well on the Set B TWS and the Set E TWS used quality assessments methods to demonstrate their impact on student learning.

Discussion

The Renaissance Teacher Work Sample (RTWS) is an authentic, multifaceted performance assessment completed by pre-service teacher candidates during student teaching to demonstrate their level of teaching proficiency relative to seven targeted teaching standards (The Renaissance Partnership for Improving Teacher Quality, 2001). The seven teaching process standards all address teaching actions influential to student learning. The RTWS was developed to assess teaching performance levels when teacher candidates are asked to show evidence of their impacts on student learning. In this investigation, we examined the generalizability of RTWS scores for two scoring methods (modified-holistic and analytic) when RTWS performances were evaluated by raters from across teacher preparation institutions. In addition, we examined support for the validity of the RTWS for the purpose of making high-stakes decisions about teacher candidates' abilities to meet the targeted teaching process standards. We also examined the link between the targeted standards and national teaching standards as represented by the INTASC standards (Interstate New Teacher Assessment and Support Consortium, 1992). Finally, using groups of measurement experts, we examined whether RTWS performances provided credible evidence for the use of sound assessment practices when teacher candidates' are required to demonstrate their impacts on student learning. Overall, our findings support the RTWS as a method for providing credible evidence of teacher candidate performance.

EVIDENCE FOR SCORE GENERALIZABILITY

A major issue for all performance assessments is the extent to which different raters provide similar judgments with respect to the quality of the observed performances. Applying Generalizability Theory (Shavelson & Webb, 1991), the results revealed significant effects for novice raters using both scoring methods (the analytic rubric and the modified holistic rubric), but not for experienced raters when using the analytic scoring rubric. These findings suggest the training and experience of the raters are important considerations when using the RTWS to make decisions about the quality of teaching performance levels. This finding is consistent with the general findings for other types of performance assessments (Dunbar, Koretz, & Hoover, 1991).

Nevertheless, the important issue for complex performance assessments, like the RTWS, is not whether or not there are scoring differences among the raters, but rather the extent of those differences and the dependability of the score decisions made by the panel of raters. Because performance assessments require the application of professional judgment when scoring, it is natural to expect a certain degree of scoring variability. To determine the degree of consistency in the RTWS scores for making absolute (criterion-referenced) decisions about candidate performance levels, Generalizability Theory (Shavelson & Webb, 1991) was applied to compute dependability coefficients. The formula for computing these coefficients also permitted

determination of the required number of raters necessary for making dependable decisions. Based on five-rater and six-rater panels, we found moderate to very high dependability coefficients for scores derived from the RTWS scoring rubrics. This means a large proportion of RTWS scores reflect differences in teacher candidate performances levels (criterion-referenced) that can be generalized across raters.

Dependability coefficients were found to be higher in general for the analytic scoring method than for the modified-holistic scoring method. Coupled with the fact that scoring times were nearly identical for the two scoring methods, the data from this investigation support the use of the analytic scoring method when high-stakes decisions are planned. Adjusting the number of raters in the formulas, we found sufficient dependability could be obtained using the analytic scoring method when panels of three or more experienced raters are used. Collectively, these findings suggest teacher work samples can be administered and scored by raters from across teacher education institutions with sufficient inter-rater agreement to make high-stakes decisions about the performance levels of teacher education candidates with respect to the targeted performance standards. However, multiple scorers remain essential to produce credible measures of performance for high-stakes decisions.

SUPPORT FOR VALIDITY

Contemporary thinking (Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education, 1999) about validity considers it to be a unitary concept—that is, there are not different types of validity, but rather different types of evidence. Validity does not inhere in the instrument but rather is related to uses of the results for certain purposes. Furthermore, validity is an ongoing argument, combining both logical and empirical elements. This study provides initial support for important aspects of the content validity of the RTWS when used for the purpose of assessing teacher candidates' abilities with respect to seven teaching process standards.

Our empirical findings support the alignment of the RTWS Prompt, the targeted standards, and the RTWS scoring rubrics. We also found support for Crocker's (1997) criteria for judging the content representativeness of performance assessments and scoring rubrics—namely, the frequency, criticality, authenticity, and representativeness of the required RTWS tasks to actual teaching performance. Our findings also yielded evidence of the alignment of the RTWS tasks with national teaching standards in the form of

the INTASC standards (Interstate New Teacher Assessment and Support Consortium, 1992). Together, the results support the content validity of the RTWS for the purpose of assessing teacher education candidates' abilities to meet the targeted teaching standards.

Because this study has validated a direct link between the teaching process standards and teaching behaviors measured by the RTWS and the INTASC (1992), the findings of this study are likely to generalize to other teacher education programs whose state and program standards are based on or similar to the INTASC standards. Hence, the RTWS could be considered by other teacher education programs for inclusion as one of their methods for providing evidence of their candidates' abilities to meet such standards as required for unit accreditation (National Council for Accreditation of Teacher Education, 2000).

EVIDENCE FOR QUALITY STUDENT LEARNING ASSESSMENT

Airasian (1999) has expressed concern about the quality of the pre- and post assessments used in teacher work samples. Faced with the demand to demonstrate impact on student learning, there is the possibility teacher candidates' might select only low-level, easy-to-meet learning goals or set easy to meet criteria for their students' responses on the post assessment. Hence, absent explicit evidence for the quality of the assessments, can work samples provide valid and credible evidence of teacher impact on student learning?

The RTWS scoring criteria take into consideration the significance of the learning goals, quality of the assessments, and student performance relative to the chosen learning goals. Hence, teacher impact on student learning is addressed by building explicit criteria relative to these factors into the RTWS scoring rubrics. Thus, the RTWS scores reflect the abilities of teacher candidates to develop quality pre- and post-assessments of student learning aligned with learning goals; to disaggregate assessment data on the pre- and post-assessments to profile student learning; to assess the impacts of their instruction on the learning of their students; and to communicate information about student progress clearly and accurately. The quality and strength of the evidence determines the rating the RTWS receives from the panel of expert raters.

To validate the judgments of the RTWS raters and to address Airasian's (1999) concerns, we had independent measurement experts evaluate the quality of the assessments employed by the teacher candidates in their work samples. Our findings revealed significant high positive correlations between these independent evaluations of the quality of the learning assessments used by the teachers to demonstrate their impact on student learning

and the total RTWS performance on the analytic scoring rubric. Although lessened by the lower and nonsignificant correlations for the Set A teacher work samples, these initial findings do provide support for the idea that successful performance on a teacher work sample can be an indication of overall higher quality assessment of student learning. This finding indicates that the approach may provide a way to incorporate impacts on student learning into teaching performance assessments that embody national, state, and institutional standards.

SUGGESTIONS FOR FUTURE RESEARCH

Future research should examine the predictive validity of RTWS performances as teacher education candidates enter the profession and become teachers. McConney et al. (1998) also noted the importance of examining the predictive validity of work sample assessments. Future investigations should also focus on other aspects of score generalizability. One important aspect to consider is the generalizability of performance ratings across different occasions of work sample development by the same teachers or teacher candidates. Future research should also examine the relationship between RTWS performances and student learning when measured by independent achievement assessments, such as high-stakes state mandated achievement tests. In addition, more work needs to be done to find ways to streamline the process and make it more efficient while maintaining high standards of measurement.

Conclusion

The work of the Renaissance Partnership to Improve Teaching Quality presented in this study contributes to a growing body of research (Danielson, 1996; Denner, Salzman, & Bangert, 2001; Denner, Miller, Newsome, & Birdsong, 2002; National Board for Professional Teaching Standards, 2001) that supports the use of complex performance assessments as credible means for documenting candidate performance with respect to national, state, and institutional teaching standards and for linking teacher candidate performance to P-12 student learning. This is important in light of the general concern in the educational community about the use of standardized paper-and-pencil tests for this purpose (see, Darling-Hammond & Snyder, 2000). Specifically, this study has shown that teacher preparation institutions can use an authentic teacher performance assessment in the form of the

Renaissance Teacher Work Sample as they strive to align their programs with performance-based accreditation standards and to meet federal and state mandates for accountability.

Teacher education programs can also learn from the approach described here. The methods followed to establish credibility evidence for the RTWS can be used for other teacher performance assessments that are focused on standards (see, Denner, Miller, Newsome, & Birdsong, 2002 for an application to a case analysis assessment). The process of benchmarking, scorer training, and the procedures for collecting validity and generalizability data can all be applied by teacher education programs to their other performance assessments.

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 Oaks, CA: Corwin Press.

Bradley, A. (1998). NEA, AFT take up the thorny issues of teacher quality. *Education Week, 18,* p. 6.

Chase, C. I. (1999). *Contemporary assessment for educators.* New York: Longman.

Crocker, L. (1997). Assessing content representativeness of performance assessment exercises. *Applied Measurement in Education, 10,* 83–95.

Danielson, C. (1996). *Enhancing professional practice: A framework for teaching.* Alexandria, VA: Association for Supervision and Curriculum Development.

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

Darling-Hammond, L., & Snyder, J. (2000). Authentic assessment of teaching in context. *Teaching and Teacher Education, 16,* 523–545.

Denner, P. , Miller, T., Newsome, J., & Birdsong, J. (2002). Using complex case analysis to make visible the quality of teacher candidates. *Journal of Personnel Evaluation in Education, 16*(3), 153–174.

Denner, P., Salzman, S., & Bangert, A. (2001). Linking teacher assessment to student performance: A benchmarking, generalizability, and validity study of the use of teacher work samples. *Journal of Personnel Evaluation in Teacher Education, 15,* 287–307.

Dunbar, S. B., Koretz, D., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. *Applied Measurement in Education, 4,* 289–302.

Ferguson, R. F.& Ladd, H. F. (1996). How and why money matters: An analysis of Alabama schools. In H. Ladd (Ed.), *Holding schools accountable* (pp. 265–298). Washington, D.C.: Brookings Institute.

Gredler, M. E. (1999). *Classroom assessment and learning.* New York: Longman.

Interstate New Teacher Assessment and Support Consortium. (1992). *Model standards for beginning teacher licensing and development: A resource for state dialogue.* Washington, DC: Council of Chief State School Officers.

Joint Committee on Standards for Educational and Psychological Testing of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, D.C.: American Educational Research Association.

McConney, A. A., Schalock, M. D., & Schalock, H. D. (1998). Focusing improvement and quality assurance: Work samples as authentic performance measures of prospective teachers' effectiveness. *Journal of Personnel Evaluation in Education, 11,* 343–363.

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

National Board for Professional Teaching Standards. (2001). *The effect of National Board Certification on teachers.* Washington, D.C.: National Board for Professional Teaching Standards.

National Commission on Excellence in Education. (1983). *A nation at risk.* Washington, DC: Government Printing.

National Commission on Teaching and America's Future. (1996). *What matters most: Teaching for America's Future.* New York, NY: Author.

National Council for Accreditation of Teacher Education. (2000). *NCATE 2000 unit standards.* Washington, D.C.: Author.

Pankratz, R. (1999). *Improving teacher quality through partnerships that connect teacher performance to student learning.* Unpublished manuscript, Western Kentucky University.

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 and Assessment Center.

Schalock, H. D., Schalock, M., & Girod, G. (1997). Teacher work sample methodology as used at Western Oregon State College. In J. McMillan (Ed.) *Grading teachers, grading schools: Is student achievement a valid evaluation measure?* (pp. 15–45). Thousand Oaks, CA: Corwin Press.

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

Stiggins, R. J. (2001). *Student-involved classroom assessment* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

The Carnegie Forum on Education and the Economy. (1986). *A nation prepared: Teachers for the 21st century.* New York, NY: The Carnegie Forum on Education and the Economy.

The Holmes Group. (1986). *Tomorrow's teachers.* Lansing, MI: Author.

The Renaissance Partnership for Improving Teacher Quality. (2001). *Teacher work sample: Performance prompt, teaching process standards, scoring rubrics.* Retrieved from *http://fp.uni.edu/itq/ProjectActivities/index.htm.*