DOCUMENT RESUME

ED 474 079 SP 041 444

AUTHOR Denner, Peter R.; Norman, Antony D.; Salzman, Stephanie A.; Pankratz, Roger S.

TITLE Connecting Teaching Performance to Student Achievement: A Generalizability and Validity Study of the Renaissance Teacher Work Sample Assessment.

PUB DATE 2003-02-17

NOTE 43p.; Paper presented at the Annual Meeting of the Association of Teacher Educators (Jacksonville, FL, February 17, 2003).

PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)

EDRS PRICE EDRS Price MF01/PC02 Plus Postage.

DESCRIPTORS *Accountability; Elementary Secondary Education; Evaluation Methods; Higher Education; Preservice Teacher Education; *Student Teacher Evaluation; Student Teachers; *Teacher Competencies; Teaching Skills

ABSTRACT

This study investigated use of the Renaissance Partnership Teacher Work Sample (RTWS) as an accountability measure for demonstrating teacher candidates' ability to meet targeted teaching standards. Teacher work samples were collected from nine universities participating in the Renaissance Partnership to Improve Teacher Quality. Trained raters assessed the teacher work samples. Results supported the generalizability of RTWS ratings made using an analytic scoring rubric. Results revealed high dependability coefficients for panels of three or more trained and experienced raters. Support for the content representativeness of 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. Results also indicated direct correspondence between the targeted RTWS tasks and seven of the ten INTASC standards. Positive correlations between the RTWS performances and independent ratings of the quality of learning assessments indicated that teacher candidates who scored well on the RTWS provided better evidence of their impact on student learning than those who scored less well. The Quality of Learning Assessment Scale is appended. (Contains 11 tables and 26 references.) (SM)

Connecting Teaching Performance to Student Achievement: A Generalizability and Validity Study of the Renaissance Teacher Work Sample Assessment

Peter R. Denner

Idaho State University

Antony D. Norman

Western Kentucky University

Stephanie A. Salzman

Western Washington University

Roger S. Pankratz

Western Kentucky University

Research session presented at the annual meeting of the Association of Teacher Educators, Jacksonville, FL, February 17, 2003. Correspondence may be addressed to Peter Denner, Assistant Dean, College of Education, Campus Box 8059, Idaho State University, Pocatello, ID 932099 or dennpete@isu.edu.

Abstract

The use of the Renaissance Partnership Teacher Work Sample (RTWS) as an accountability measure for demonstrating teacher candidates' abilities to meet targeted teaching standards was investigated. The findings support the generalizability of the RTWS ratings made using an analytic scoring rubric. The results revealed high dependability coefficients for panels of three or more trained and experienced raters. Support for the content representativeness of 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 indicated direct correspondence between the targeted RTWS tasks and seven of the ten INTASC standards. Finally, positive correlations between the 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.

Connecting Teaching Performance to Student Achievement: A Generalizability and Validity Study of the Renaissance Teacher Work Sample Assessment

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, one such body, 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, who are part of the Renaissance Group, a consortium of colleges and universities throughout the United States with a major commitment to educating teachers, 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." (Pankratz, 1999, 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. 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 decision regarding teaching performance levels with respect to those standards.

At three consecutive partnership meetings (January 2002, June 2002, and January 2003), representatives from the eleven project institutions met to investigate whether the Renaissance Teacher Work Sample (RTWS) provided sufficient credible evidence of teacher candidates' abilities with respect to the targeted teaching standards to warrant its use for the purpose of high-stakes assessment and program accountability. 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 content representativeness of the RTWS and to examine its validity 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 three TWS collections: a collection of N = 110 TWS gathered in June 2001, a collection of N = 87 TWS gathered in June 2002 and a collection of N = 115 TWS gathered in January, 2003. All three 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), all TWS within each collection were 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 a set of 10 TWS (Set 1) was created from a random selection of exemplar TWS by holistic category . The Set 1 TWS consisted of 2 Beginning, 3

Developing, 3 Proficient, and 2 Expert TWS. From the second collection of TWS (N = 87) in June 2002 as second set of 10 TWS was selected (Set 2). The 10 Set 2 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 2 TWS consisted of 1 Beginning, 3 Developing, 4 Proficient, and 2 Expert TWS. From the third collection, following the same type of benchmarking procedure, TWS were randomly selected (except for those TWS categorized at the beginning level as explained below) by holistic category as follows: 4 Beginning, 10 Developing, 10 Proficient, and 5 Expert. The set 3 TWS had only four TWS at the beginning level because they were all of the TWS categorized at that level in the January, 2003 collection.

Instruments

RTWS Scoring Rubrics. The RTWS Scoring Rubric was 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 RTWS 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 teaching process standards, there were 32 total indicators; therefore, total analytic scores could vary from 0 to 96 points.

Validity Questionnaire. To establish content-related evidence for validity, a questionnaire was developed to ask 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 the content-related 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; Grendler, 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 twelve items provided a total score from zero to 24 (see Appendix).

Teacher Work Sample Raters

In January 2002, five raters were selected from the 55 raters assembled and trained in St. Louis. The raters included an administrator, 3 faculty member and 1 public school teacher. In June 2002, six additional raters were asked to score the Set 2 TWS. The six Set 2 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 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 were assigned the task of sorting the TWS gathered in each collection 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% -20% of the work samples. When a group reached consensus on the holistic category, they placed the TWS in that pile. In the afternoon, the TWS within each category were examined by a different mix of raters assigned to pick exemplars of the assigned category. Following group discussion, four to six exemplars of performance in *each* category were identified. As described previously, TWS Set 1 was created by randomly selecting exemplar TWS by category. Set 2 and Set 3were created by random selection from within each of the four benchmark categories (except for the Set 3 beginning level category where all four TWS at that level were selected for inclusion in the set).

RTWS Scoring. All raters scored their assigned set of TWS (Set 1 or Set 2) independently using the RTWS scoring rubric. 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. The average grading time per TWS for the raters of Set 1 and Set 2 was about 28 minutes.

Content Validity Ratings. Content validity data was gathered in June 2002. The validity assessment panel consisted of 42 representatives from across the 10 partnership 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 introduced 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 practiced rating at

least two work samples prior to completing the validity questionnaire. All panel members independently completed the sections of the content validity questionnaire.

Procedures for the Quality of Learning Assessment

Expert Raters. An independent panel of measurement experts consisting of 3 expert raters was asked to evaluate the Set 2 TWS using the Quality of Learning Assessment (QLA) rating scale. The QLA raters had extensive background in testing and measurement. All were experienced in the development and use of scoring rubrics. Using repeated measures ANOVA, the effect of rater on the QLA scores was not found to be statistically significant, $F(2, 18) = .44$, $MSE = 8.40$, $p = .65$. The three rater coefficients of dependability for the QLA scores was calculated to be .84. The meaning of a dependability coefficient is explained later in the design section. Together, these findings suggest sufficient inter-rater agreement for the purpose of this investigation.

QLA 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 evidence 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.

Procedures for Analysis of Evidence for Learning

Two of the researches examined the Set 3 TWS and reached consensus as to whether or

not each TWS contained evidence for learning gains by achievement goal and by student. They also reached consensus as to whether or not each TWS contained evidence for student achievement of the stated criteria for each targeted learning goal. This process took about 5 hours.

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 on scores derived from the RTWS scoring rubric. 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 coefficients depending upon whether the measure is to be used to make decisions about the "relative standing or ranking of individuals" or about "the absolute level of their scores" (Shavelson & Webb, 1991, p. 84). We used the formulas for computing an index of dependability for absolute decisions 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). 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 coefficients of dependability were calculated using formulas supplied by Shavelson and Webb (1991). The same formulas were 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. A chi-square test for linear trend (discussed in Steel & Torrie, 1960) was used to determine whether the evidence for learning gains and accomplishment of learning goals increased with TWS category level. 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 questionnaire. For all statistical analyses, the level of statistical significance was set at $\alpha = .05$.

Results

Score Generalizability

Effect for Raters across TWS Sets

The effect of rater was statistically significant for the Set 1 TWS, $F(4, 36) = 6.28$, $MSE = 59.21$, $p = .001$, but not for the Set 2 TWS, $F(5, 45) = 1.07$, $MSE = 100.94$, $p = .39$, for the TWS total scores when experienced raters were nominated by their institutions. 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 that were used in the formulas for computing the dependability coefficients for both TWS sets. For Set 1 TWS, for raters who were selected on the basis of the degree of match to a scoring criterion, the five rater coefficient of dependability was computed to be .88. For the experienced

raters, who scored the Set 2 TWS, the six-rater coefficient of dependability was computed to be .87. Because the second set had less variability among the TWS, the coefficient is somewhat lower. However, taken 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 an acceptable level of dependability of .77 to .82 could be achieved with as few as three raters. Table 2 displays the dependability coefficient estimates for different numbers of raters using the results obtained from both TWS sets. These findings suggest TWS can be feasibly administered and scored by raters from across teacher education institutions with sufficient inter-rater agreement to make absolute decisions about the overall performance levels of teacher education candidates with respect to the targeted performance standards.

Content Validity

To evaluate the content validity of scores derived from the RTWS, we applied criteria suggested by Crocker (1997) for judging the validity of performance assessments. These criteria included alignment of the standards and the tasks with the scoring rubric, the frequency of the targeted behaviors in actual practice, the importance or criticality of the targeted behaviors to real performance, the authenticity of the tasks to actual performance situations, and the representativeness of the tasks with respect to the targeted performance standards. Each of these criteria will be addressed separately in the sections that follow.

Alignment

Table 3 presents the judgments made by our validity assessment panel regarding alignment among the RTWS Guidelines, the targeted teaching processes (e.g., the TWS standards), and the analytic scoring rubrics. For the alignment between the TWS elements presented in the guidelines and the targeted standards, 78.6% (f =33) of panel members indicated a high degree of alignment. For the alignment between the TWS task elements and the analytic scoring rubric, 69% (f = 29) of the panel members said there was a high degree alignment. For the alignment of the analytic scoring rubric with the targeted standards, 73.8% (f = 31) said there was high alignment. Overall, the evidence supports this criterion for quality performance assessments.

Frequency

Table 4 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% 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% of the raters. These results support the frequency criterion of content representativeness.

Criticality

To support the *criticality* of the tasks performed while completing the RTWS, we asked the validity assessment panel to rate the importance of the teaching behaviors required. Table 5 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 of the teaching behaviors were considered to be "important" or "very important." Thus, the Renaissance TWS assessment satisfies this criterion.

Authenticity

Next, we asked our validity assessment panel to judge how authentic the tasks required by the RTWS are to success as a classroom teacher. Table 6 presents the number and percent of the panel member ratings 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% 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% 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)." Across all nine tasks, the results support the authenticity criterion for valid performance assessment.

Representativeness

We also asked the validity assessment panel to consider the degree to which the tasks required by the RTWS reflect and represent the targeted standards. The ratings of the panel members are presented in Table 7. Once again, the majority (88.1% to 97.6%) 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%). Therefore, this criterion of valid performance assessment was also met.

Match to INTASC Standards

Finally, we asked our panel of experts to indicate the extent to which the tasks required for the RTWS reflected the 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 many of the INTASC standards. Table 8 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 8, 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 TWS, 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 to be measured by the TWS and these standards were not targeted by the RTWS. Overall, the results

support the RTWS as a measure of many of the INTASC standards.

Correlation of QLA Total Scores with RTWS Total Scores and Sub-scale Scores

Table 9 presents the correlations among the RTWS scores and the Quality of Learning Assessment (QLA) total scores for the Set 2 TWS. All scores were averaged across raters. As can be seen from Table 11, the correlation was positive, $r = .70$, $n = 10$, $p = .025$, for the total score relationship between the QLA and RTWS scores. For a variety of reasons, related to the fact that the RTWS measures multiple teaching process standards, only some of which are focused on the candidates' documentation of their impact on student learning, it is not surprising this correlation is only at a moderate to high level. Examination of the correlations of the RTWS sub-scale scores with the QLA scores revealed high and statistically significant correlations between the QLA scores the RTWS Learning Goals sub-scale scores, $r = .80$, $p = .005$ and the RTWS Analysis of Student Learning sub-scale scores, $r = .91$, $p < .001$. A statistically significant positive correlation was also found for the relationship between the QLA scores and the RTWS Instructional Decision-Making sub-scale scores, $r = .65$, $p = .042$. These data support the idea that teacher education candidates who scored well on Set 2 TWS used quality assessments methods to demonstrate their impact on student learning. It should noted, however, that due to the constraints of this study, these correlations were based on a rather small number of work samples.

Evidence for P-12 Student Learning

Table 10 presents the number and percent of TWS showing evidence for learning gains by achievement goal and by student for the Set 3 TWS. Table 11 shows the number and percent

of the Set 3 TWS containing clear evidence for whether or not each student achieved the targeted learning goals. As can be seen in both tables, the percentage of TWS containing clear evidence increases across the four TWS performance categories. The chi-square test for linear trend was statistically significant for the evidence for learning gains, χ^2 (df = 1) = 7.30, $p < .05$, but it did not reach statistical significance for the evidence for accomplishment of the targeted learning goals, χ^2 (df = 1) = 2.91, $p > .05$.

Discussion

The Renaissance Teacher Work Sample (RTWS) is an authentic, multifaceted performance assessment intended to be completed by preservice 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 support for the content validity of the RTWS for the purpose of making high-stakes decisions about teacher candidates' overall 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 Interstate New Teacher Assessment and Support Consortium (INTASC) standards (Interstate New Teacher Assessment and Support Consortium, 1992). In addition, we investigated the generalizability of the RTWS scores when the RTWS performances were evaluated by raters from across teacher preparation institutions. Finally, using groups of

measurement experts, we examined whether RTWS performances provided credible instruction – embedded evidence for teacher candidates' impact on learning gains, accomplishment of targeted learning goals, and for the use of sound assessment practices to demonstrate their impacts on student learning. Our findings support the RTWS as a method for providing credible evidence of teacher candidate performance with respect to state and institutional teaching standards and for instruction embedded evidence of their impacts on student learning.

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. To examine this, we applied a research design from Generalizability Theory (Shavelson & Webb, 1991) to assess the consistency of the RTWS scores assigned by cross-institutional panels of raters, which included faculty members, administrators, and public school teachers affiliated with institutions, when using the RTWS scoring rubrics. Although we found significant effects for less experienced raters, we did not find a significant effect for experienced raters when using the RTWS scoring rubric. Our 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.

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 judgement when scoring, it is

natural to expect a certain degree of scoring variability. Generalizability Theory (Shavelson & Webb, 1991) also provides two kinds of summary coefficients (for absolute and relative decisions) that reflect scoring consistency. We chose to compute dependability coefficients indicating the degree of consistency in scores for making absolute (criterion-referenced) decisions about candidate performance levels. The formulas for computing dependability also permit determination of the required number of raters necessary for making dependable decisions. Based on five rater and six rater panels, we found 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 (absolute levels) that can be generalized across raters. Adjusting the number of raters in the formulas, we found sufficient dependability could be obtained when panels of three or more experienced raters are used. Hence, our findings indicate the RTWS can be administered and scored with sufficient inter-rater dependability to be used to make high-stakes decisions about overall teaching performance across the targeted teaching performance standards.

Support for Content 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 the seven targeted 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). The panel of raters indicated a direct correspondence between RTWS tasks and INTASC standards for those standards that matched the seven teaching processes targeted by the RTWS, and a lesser alignment where there was a lesser potential for match. 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.

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. Airasian (1999) asked whether teacher work samples can provide valid and credible evidence of

teacher impact on student learning absent explicit evidence for the quality of the learning assessments.

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 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 RTWS performance scores. 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. Although our investigations in this area are still preliminary, this finding indicates that our approach may provide a way to incorporate impacts on student learning into teaching performance assessments that embody national, state, and institutional standards.

Evidence for Impact on Student Learning

A major goal of the Renaissance partnership project has been to connect teacher performance to its impact on student learning . The RTWS is a teaching performance assessment that requires teacher candidates to demonstrate their impact on student learning using instruction embedded assessments. As part of the tasks required by the RTWS, teacher candidates' must profile the learning of their students with respect to the unit's targeted learning goals through the use of graphs that show pre-assessment to post-assessment learning gains. In addition to analyzing the assessment data for the whole class, the candidates' must also disaggregate the assessment data to explain progress and achievement toward the learning goals by subgroups of students and by selected individual students. To validate that TWS performance is a reflection of teacher candidates' abilities to be accountable for and to show evidence of their impacts on student learning, we had assessment experts examine a set of RTWS for evidence of learning gains and for evidence of meeting the criteria set for achievement of the unit's targeted leaning goals. The findings affirmed RTWS performance levels were linearly associated with evidence for learning gains across achievement goals and students. This is an important finding because it means RTWS performance is an indication of teacher candidates' abilities to show positive impacts on student learning. The evidence was less clear for accomplishment of the targeted learning goals according to the criteria set by the teacher candidates but there was a similar linear trend across RTWS performance levels. This latter finding was largely do to the fact that the teacher candidates did not always explicitly state their assessment criteria, so it was hard to determine whether or not the learning goals were met without inferring an acceptable

performance level from the candidates' general reflections on their students' progress and success. This points to the need for teacher education programs to do a better job mentoring teacher candidates to set explicit criteria for student learning success.

Suggestions for Future Research

Future research should examine the predictive validity of RTWS performances as teacher education candidates enter the profession and become teachers. The importance of examining the predictive validity of work sample assessments has also been noted by McConney et al. (1998). 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. Finally, future research should also examine the relationship between RTWS performances and student learning when measured by independent, but curriculum linked, achievement assessments, such as high-stakes state mandated achievement tests designed to assess state achievement standards.

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.

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

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.

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.

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, DC: 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.

Isaac, S., & Michael, W. B. (1971). *Handbook in research and evaluation*. San Diego, CA: Knapp.

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, DC: 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 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, DC: 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.

Smith, M. L., & Glass, G. V. (1987). *Research and evaluation in education and the social sciences*. Englewood Cliffs, NJ: Prentice-Hall.

Steel, R. G. D., & Torrie, J. H. (1960*). Principles and procedures of statistics*. New York: McGraw-Hill.

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

Table 1

Estimates of Variance Components for Total RTWS Scores for Each TWS Set.

	Estimated Variance Components	
	Set 1 (5 raters)	Set 2 (6 raters)
Person	138.38	111.64
Residual	59.21	100.94

Table 2

Total Score Dependability Coefficient Estimates by Number of Raters for each TWS set

Number of Raters	Dependability Coefficient Estimates	
	Set 1	Set 2
6 Raters	.90	.87
3 Raters	.82	.77
1 Rater	.60	.53

Table 3

Number and Percent of Panel Members Indicating Alignment Between the Renaissance TWS Guidelines, TWS Standards and TWS Scoring Rubric (N = 42)

Overall Alignment	Degree of Alignment			
	Poor	Low	Moderate	High
	1	2	3	4
Alignment of the Renaissance TWS Guidelines & Prompts with the targeted teaching process standards and indicators			9 21.4%	33 78.6%
Alignment of the Renaissance TWS Guidelines & Prompts with the analytic scoring rubric		1 2.4%	12 28.6%	29 69.0%
Alignment of the analytic scoring rubric with the targeted teaching process standards and indicators		1 2.4%	10 23.8%	31 73.8%

Table 4

Number and Percent of Panel Members Indicating How Frequently They Would Expect a Teacher to Engage in the Teaching Behaviors Targeted by the TWS (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%
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%

Table 5

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

Teaching Behaviors Targeted By Teacher Work Sample	Degree of Importance			
	Not at all Important 1	Somewhat 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%
Use ongoing analysis of student learning to make instructional decisions.			5 11.9%	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%

Table 6

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	Somewhat 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%
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%
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 7

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 Represen-tative 1	Somewhat Represen-tative 2	Represen-tative 3	Very Represen-tative 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%
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		1 2.4%	10 23.8%	31 73.8%

and modify original instructional design and lesson plans to improve student progress toward the learning goal(s).				
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 8

Number and Percent of Panel Members Indicating the Extent to Which the Tasks Required by the Teacher Work Sample Reflect the INTASC 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%
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%
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%

Table 9

Correlations of RTWS Total Score and Sub-Scale Scores with the Total Quality of Learning Assessment Score for the Set 2 TWS (n = 10).

	Quality of Learning Assessment
A. RTWS Total Score	.70*
1. Contextual Factors	.02
2. Learning Goals	.80*
3. Assessment Plan	.58
4. Design for Instruction	.59
5. Instructional Decision-Making	.65*
6. Analysis of Student Learning	.91*
7. Reflection and Self-Evaluation	.63

*$p < .05$

Table 10

Percent of RTWS by Holistic Category Showing Evidence for Learning Gain for Each Student by Targeted Learning Goal for the Set 3 TWS (n = 29).

Holistic Category	*n*	Evidence for Learning Gain	
		No	Yes
4 = Expert	5	0%	100%
3 = Proficient	10	20%	80%
2 = Developing	10	50%	50%
1= Beginning	4	75%	25%

Table 11

Percent of RTWS by Holistic Category Showing Evidence for Achievement of the Learning Goals for Each Student by Targeted Goals for the Set 3 TWS (n = 29).

Holistic Category	*n*	Evidence for Achieving Learning Goals	
		No	Yes
4 = Expert	5	20%	80%
3 = Proficient	10	50%	50%
2 = Developing	10	60%	40%
1= Beginning	4	75%	25%

Appendix

Quality of Learning Assessment Rating Scale

1. Learning goals reflect several types of learning and are significant and challenging.
2. Learning goals are clearly stated as learning outcomes.
3. Learning goals are appropriate for the development and prerequisite knowledge, skills, and experiences of the students and other student needs.
4. Learning goals are explicitly aligned with national, state, or local standards.
5. Assessments are congruent with the learning goals in content and cognitive complexity.
6. Assessment criteria are clear and explicitly linked to the learning goals.
7. The assessment plan includes multiple assessment modes and assesses student performance throughout the instructional sequence
8. The assessments appear to be valid measures of the learning goals.
9. Scoring procedures are explained.
10. Assessment items or prompts are clearly written.
11. Assessment directions and procedures are clear and would likely be understood by the students.
12. Evidence of student learning includes data from assessments before and after instruction.

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: Connecting Teaching Performance to Student Achievemment: A Generalizability and Validity Study of the Renaissance Teacher Work Sample Assessment.	
Author(s): Peter Denner, Anthony Norman, Stephanie Salzman, Roger Pankratz	
Corporate Source:	Publication Date: 02-17-03

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

Level 1

X

Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.

The sample sticker shown below will be affixed to all Level 2A documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2A

Level 2A

☐

Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only

The sample sticker shown below will be affixed to all Level 2B documents

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY

Sample

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

2B

Level 2B

☐

Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Signature: Peter R. Denner	Printed Name/Position/Title: Peter Denner/Assistant Dean	
Organization/Address: Idaho State University 1550 E Terry Street Pocatello ID 83209-8059	Telephone: 208-282-4230	FAX: 208-282-4697
	E-Mail Address: dennpete@isu.edu	Date: 02-25-03

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfacility.org