NEERO CONFERENCE
COVER PAGE

Title:	Clinical Supervision: Making Meaning of Student Teaching Practices
Authors:	Linda Capalbo (contact person) Marie Lynch
Institutional Affiliation:	Rhode Island College
Contact Information:	lcapalbo@ric.edu 401-456-9571 mlynch@ric.edu 401-456-8763

Abstract

This study analyzed written records created by college clinical supervisors, of student teaching observations carried out during the Fall 2008 and Spring 2009 semester. Observations, conducted in public schools in Rhode Island, reflected the dual enrollment status of each student teacher; that is, each candidate was observed, multiple times, in both a general elementary or middle level classroom and in a setting focused on students with special educational needs. The purposes of the analysis were to 1) review the language and ratings used by clinical supervisors about teacher candidates' practice during student teaching; 2) demonstrate how supervisors, in institutional documents, account for, explain, or inform stakeholders of the practices observed; 3) explore implications about best practice feedback.

Keywords **–** Student teachers, supervision, assessment, observation protocols

CLINICAL SUPERVISION: MAKING MEANING OF STUDENT TEACHER PRACTICES

Objectives

Objectives of this roundtable presentation are to 1) review the findings regarding language and ratings used by clinical supervisors about teacher candidates' practice during student teaching; 2) demonstrate how supervisors, in institutional documents, account for, explain, or inform stakeholders of the practices observed; 3) explore implications about best practice feedback.

Theoretical Framework

Despite the perceived importance in teacher education of structured clinical observations of student teachers' classroom performance, "there is little systematic research on exactly what the most effective supervisors do". (Darling-Hammond, Hammerness, Grossman, Rust & Shulman, 2005, p. 412). While all teacher preparation programs have developed tools to aid in the evaluation of student teachers' performance (Blanton, Sindelar, Correa, 2006; Lang & Wilkerson, 2007), we have little insight as to how well those tools are used by supervisors to capture practice and to provide meaningful feedback for the candidate as well as the institution.

Although parameters for gauging effective practice delineated by the National Council for Accreditation in Teacher Education (NCATE, 2002) are helpful to institutions who prepare pre-service teachers, they don't speak to the processes by which judgments are made by the observer in the field. Given the range of backgrounds and experiences of supervisors, (Raths & Lyman, 2003) many adjunct faculty, the imperfections of any particular observational tool (Tucker & Stronge, 2005), and the varying circumstances in which teacher candidates practice (Hamel & Merz, 2005; Good & Brophy, 2000), how *do* professionals gauge levels of expertise in planning, acting, and reflecting demonstrated by novice practitioners and then inform stakeholders of the decisions made?

The growing number of students with disabilities in general education classrooms across the country has created a critical demand for general educators to have knowledge and skills beyond a particular content strand (Prater & Sileo, 2004; Rademacher, Wilhelm, Hildreth, Bridges, & Cowart, 1998; INTASC, 2001). Thus, many higher education institutions offer dual certification tracks (general and special education) to help candidates prepare for the "increasing diversity and inclusiveness of public school classrooms" (Van Laarhoven, Munk, Lynch, Wyland, Dorsch, Zurita, Bosma, & Rouse, 2006, p.209).

Some researchers suggest current models of supervision, may be more ecologically appropriate for candidates in general education classrooms than for candidates in special education settings (Albi, Clifford, & Macy, 2005; Blanton, et al, 2006; Coladorci & Breton, 1997). Many of the protocols/processes that define a clinical supervisor's role have been created and developed by general educators (Albi, et al, 2005; Prater & Sileo, 2004).

Paris & Gespass (2001) point out that, given the varied teaching sites in which candidates are observed, "each [supervisor] may construct somewhat different relationships and processes with each student teacher." They add, "It will mean expanding our thinking about what counts

as data when we examine the student teacher's work together to include not only what is seen but what is thought" (p. 411).

Researchers (Darling-Hammond et al, 2005; Wilson, Floden, & Ferrini-Mundy, 2002) suggest that the nature of the support given to student teachers during their clinical experiences is key to their development. Cochran-Smith (2005) suggests that future research in the field of teacher education needs to create a chain of evidence linking teacher preparation to student achievement. Perhaps, an examination of how supervisors describe, differentiate and evaluate the performances of student teachers can help us better define the meaning of clinical supervision, providing one important link in that chain.

Methodology & Data Sources

This mixed-method study analyzed 144 observational reports of 12 student teachers created by elementary and special education clinical supervisors and cooperating teachers during the Fall 2008 and Spring 2009 semesters. Each candidate was observed at least six times, three times in either a traditional elementary or middle level classroom, and three times in a setting that supported special education students. The protocols, developed within the college's School of Education, were standard across placements and participants.

Text data in the form of open-ended commentary generated from the observational reports were analyzed using content analysis to explore the relationship between elementary and special education commentary for 12 teacher candidates (Krippendorf, 1980). Each observational report required the clinical supervisor to also rate and identify the teacher candidate's practice on a 4-point rating scale as either exemplary, acceptable, developing or unsatisfactory.

Data were analyzed quantitatively as well, using Cohen's (1960) Kappa Coefficient of Agreement (Conger, 1985; Fleiss, 1971) to assess the level of agreement between paired raters (e.g. elementary education supervisor/special education supervisor; elementary education supervisor/cooperating teacher; special education supervisor/cooperating teacher. It should be noted that we viewed agreement between/among raters strictly to mean that each rater gave the student teacher the same rating as indicated on the observation protocol with respect to a candidate's ability to plan, act and reflect on their lesson during student teaching. It was assumed that the ratings of a "competent teacher" vs. an "exemplary teacher" would be distinguishable and thus comparable in both elementary and special education settings. Typically, agreement of at least .70 is viewed as acceptable, but we considered *moderate* (.41-.60) or *substantial* (.61-.80) levels of agreement between raters when they were noted as statistically significant (Sim & Wright, 2005).

Results

To analyze data quantitatively, 144 observational reports that documented the candidate's teaching three times in each setting (e.g. elementary and special education), were considered. The observational protocol, standard across all observations, emphasizes three strands aligned with the School of Education's Conceptual Framework that focuses on creating teachers who are *Reflective Practitioners who Plan, Act, & Reflect*. Within each strand or section, subsequent statements are used to rate the student teacher on various areas that are briefly outlined in Table 1.

Table 1. Overview of Observation Report

Plan	Act	Reflect
Selection of Content	Instructional Opportunities	Works Collaboratively
Integrated Unit & Lessons	Positive Learning Environment	Accepts Constructive Criticism
Developmentally Appropriate Instruction	Classroom Management Techniques	Implements Suggestions
Diverse Learner Needs	Oral/Written Communication Skills	Analyzes Teaching
Uses Formal/Informal Assessment	Uses Formal/Informal Assessment	Follows Policy & Procedures
	Projects Professional Image	Professional Interactions

For example, the Plan section of the observation report requires the rater to assess how well the student teacher planned for his/her students regarding the content area chosen for the lesson, whether the lesson was part of an overall integrated unit, offered developmentally appropriate instruction, attended to the diverse learner needs of the students, and provided for adequate assessment opportunities that were both formal and informal in nature.

Thus, each section had 5-6 indicators that required the elementary or special education clinical supervisor and/or cooperating teacher to rate the student teacher on a 1-4 scale. The entire protocol had a total of 17 indicators, whereby on which candidates were rated as Exemplary, Competent, Developing, or Unsatisfactory. In order to calculate the level of agreement by percentage, the ratings were assigned a numerical score in the following way: Exemplary (4); Competent (3); Developing (2); Unsatisfactory (1). In some cases, the rating of C+ or C/E was given, which indicated a rating between a 3 and 4. This score was then assigned a rating of 3.5. In other instances a combined score of D/C was noted between Developing and Competent, which was then given a 2.5 numerical score.

Results indicate that there is 100% agreement between raters that no teacher candidate was rated as “unsatisfactory”. Raters also fully agreed that by the third observation in both the elementary and special education settings, the teacher candidates achieved at least a “competent” overall rating. However, agreement became less evident when rating student teachers as either “competent” or “exemplary”. The mean level of agreement was calculated across 6 observations/pair, in each of the three strands**.** It should be noted that *moderate* to *substantial* agreement did not necessarily indicate a positive correlation in terms of candidates’ practice but strictly meant that supervisors agreed on the rating given.

Indicated in Table 2, five of 17 areas showed statistically significant *moderate* or *substantial* agreement between elementary and special education supervisors’ ratings with respect to competent and exemplary categories (Sim & Wright, 2005). For example, under the Plan section, *substantial* agreement between supervisors was noted under for the indicators: Appropriate Instruction and Uses Formal/Informal Assessments. The Act strand yielded *moderate* agreement for Projects Professional Image, but *substantial* agreement in Creates a

Positive Environment. The Reflect section indicated *substantial* agreement in one area: Analyzes Teaching.

Table 2. Statistically Significant Agreement between Supervisors

Strand	Topic	*M*
PLAN	Appropriate Instruction	.68
	Formal/Informal Assessment	.66
ACT	Projects Professional Image	.47
	Positive Environment	.69
REFLECT	Analyzes Teaching	.62

Comparisons between special education supervisors and special education cooperating teachers, and between elementary education supervisors and cooperating teachers were also conducted as shown in Table 3. Moderate agreement for both special education and elementary education rater pairs was indicated for almost all of the indicators, with the demonstrated statistically significant range from .43 to 60. Three common criteria were noted across program raters with *moderate* agreement for Formal & Informal Assessment (Plan), *moderate* agreement for Accepts Constructive Criticism (Reflect) and *substantial* agreement when rating Implemented Suggestions (Reflect) for the both special education and the elementary education pairs.

Raters also seemed to highlight specific criteria in their respective domains. For example, special education supervisors and cooperating teacher raters reached *moderate* agreement within the following indicators: Diverse Learner Needs (Plan); Instructional Opportunities (Act), Formal/Informal Assessment (Act), and Follows Policy/Procedures, Works Collaboratively; and Interaction with Colleagues (all Reflect). Conversely, elementary education supervisors and cooperating teachers indicated *moderate* agreement in the Plan strand: Selection of Content, Appropriate Instruction, and Integrated Unit, and Positive Environment in the Act section.

Table 3. Statistically Significant Agreement between Supervisors/Cooperating Teachers

Strand	Special Educators' Agreement	*M*	Elementary Educators' Agreement	*M*
PLAN	Diverse Learner Needs	.58	Selection of Content	.54
	Formal/Informal Assessment	.57	Appropriate Instruction	.53
			Formal/Informal Assessment	.51
			Integrated Unit	.51
ACT	Instructional Opportunities Formal/Informal	.60	Positive Environment	.58

	Assessment	.43		
REFLECT	Follows Policy/Procedures	.43	Accepts Constructive Criticism	.53
	Accepts Constructive Criticism	.50		
	Implemented Suggestions		Implemented Suggestions	.61
	Works Collaboratively	.66		
	Interaction with Colleagues	.44		
		.57		

Supervisors' Commentary

Open-ended commentary, generated from 66 supervisors' observational reports, was also compared using content analysis. Coding of open-ended comments written by supervisors to support ratings assigned to specific performance descriptors revealed three primary categories of language use: descriptive; functional, and evaluative. Descriptive language was used to indicate the presence of a particular strategy or technique, or to convey the type of setting or event in which a strategy or behavior occurred:

"K acted as primary teacher with the Cooperating Teacher as co-teacher." (#1SP09).

Descriptive language tended to be neutral in tone and content, merely informing the reader of the conditions extant in the observation. Language describing teaching/learning conditions was most often delivered in a third-person mode.

Functional language, often, but not always, delivered in a first-person mode, was used to provide direction to the teacher candidate by providing specific suggestions, advice, or tips to improve teaching performance or to support student learning:

"You could have circulated to check each team's answer." (#2SP09).

Functional language gave the candidate "something to take away with them" from the observation. Functional comments, generally, were neutral in tone, but at least implied a negative judgment of performance since the comment highlighted a behavior deemed missing from the candidate's teaching repertoire.

Evaluative language was used to make (relatively) explicit the quality of the candidate's performance in the view of the observer and, ostensibly, to provide qualitative evidence to support overall ratings for each performance descriptor:

"This was an excellent lesson…extremely well thought out and developed." (#9SP09).

Language used to evaluate candidate performance reflected both first and third person modes.

A second analysis of supervisors' open-ended comments more closely examined the qualifying or evaluative language used to characterize candidates' teaching expertise by both special

education and elementary education supervisors. Several descriptive words and phrases appeared multiple times across observation reports, accounting for 107 of the 298 evaluative comments made by raters. Most commonly used descriptors included: well planned; well organized; well implemented; appropriate; consistent; professional; clear; engaging, and works well with ... .

It should be noted that the absence of a particular evaluative phrase did not signify a negative evaluation, merely that the rater(s) chose other phrases to characterize the candidates' performance. Raters had complete discretion in choosing language to qualify their narrative ratings of candidates.

Although all teaching candidates were being evaluated against the same set of criteria, it is to be expected that not all candidates will perform at the same levels of expertise. An overall evaluation of "competent" is the measure by which candidates pass or fail student teaching. Many supervisors (approximately 44%) chose to qualify candidates' scores on the rating statements in each strand (Plan, Act, Reflect); that is candidates might receive a C/E or C+ on a particular statement. Consequently, the overall rating on a given observational report placed it in a category of: developing (1), competent (29), competent plus (20) or exemplary (16). Further analysis of the language supervisors used to differentiate performance revealed that just as the same types of language (descriptive, functional, and evaluative) were used across all categories, so were many of the same evaluative phrases used to describe differing levels of performance. Thus, phrases such as "well-planned" and "appropriate" appeared as descriptors for developing, competent, competent plus and exemplary performances.

To try to distinguish what characterized an exemplary from a competent performance, a key focus of this study, the 298 evaluative statements were examined for differences in degree, since differences in kind appeared negligible. When the language was analyzed in this way, key words such as "very", "highly", "strong", "most" and "excellent" emerged as more frequent qualifiers. Table 4 shows the distribution of these qualifiers of degree by level and by supervisor group.

Table 4. Qualifiers of Degree

Qualifying Ratings	SPED Supervisors	ELED Supervisors
Competent	15.8%	35%
Competent plus	40%	52%
Exemplary	29%	59%

Note: Percentages correspond with the total number of evaluative statements made divided by the qualifier ratings. The total number varied ranging from 24-63 (SPED) and 40-63 (ELED).

A final note on qualifiers of degree: although it would be expected that evaluative statements would be as likely to be negative as positive, certainly when describing behaviors or practices that qualify as developing or even competent, the data revealed that overtly negative evaluative statements were rare. Of the 7 "developing" statements, 2 were negative, but in a functional form rather than evaluative ("I would suggest"; "in my opinion"). Of 103 "competent" statements, 4 were negative (1 sped; 3 eled). In the competent plus category, 1 (sped) of 120 statements was negative, and in the "exemplary" category there were no negative statements among the 68 overall.

Cooperating Teacher' Commentary
Subsequent to the analysis of supervisors' data, a follow up content analysis of the open-ended commentary from 72 cooperating teachers' observations of the same candidates was conducted. The three categories of language use identified in supervisors' reports – descriptive, functional, and evaluative – also were present in cooperating teachers' comments. It appeared that the cooperating teachers commentary in general provided a wider range of types of language; that is, they often described one event or teaching behavior using all three language types. In this regard, cooperating teachers' written comments could be characterized, at least in some cases, as providing "thicker" description than that of supervisors, potentially providing more useful feedback to teacher candidates.

The use of qualifiers (e.g. well-planned; well implemented, etc.) to distinguish competent-plus and exemplary performances was also consistent with that of the supervisors. As noted above, this type of evaluative language was often embedded in the descriptive and directive components of Cooperating Teachers' commentary, thus providing multiple layers of feedback.

Discussion
The original purposes of the analysis were to 1) examine the language used by the observer that both describes and evaluates the student teacher's performance, particularly as it differentiates levels of expertise; and 2) explore the level of agreement offered within and across (multiple) reports about the same teacher candidate. Our hope was to better define the similarities and differences that exist across the diverse fields of elementary and special education and make meaning of the clinical supervision/cooperating teacher feedback provided to candidates.

Types of Language: Supervisors
The types of language used by the supervisors in their written comments, descriptive, functional, and evaluative, are not surprising. The role of the supervisor is typically thought to comprise these functions. One brings to the surface for examination the professional behaviors of the candidate, both productive and unproductive, acts as a mentor to help the candidate increase the former and eliminate or reduce the latter, and serves as the arbiter (along with the cooperating teacher) of the progress toward mastery that the candidate has made in her/his student teaching placement. The protocol designed by the institution to capture these professional behaviors and interactions is designed to promote all of these through observation and written and oral communication among all parties involved.

Several points of interest arise when examining the types of language used in these protocols. First, a high proportion of the statements made in writing across observations tended to be descriptive, even in those sections clearly designated for evaluation of performance. Descriptions of performance, while needed, don't necessarily contribute to the mentoring or evaluative roles of the supervisor. In a number of cases, what could have become functional – specifically helpful to the candidate to know how to improve instruction – was left at the descriptive level. The candidate was made aware of what happened, not necessarily what was good or problematic about the situation.

"Directions for group work were given as well as roles for each student in group."

(#2SP09)

When functional language was used, it wasn't always as explicit as it might have been:

> "Your presentation of problem and use of overhead to depict concepts is well-planned but students needed to be supported more in understanding concepts." (#8SP09)

Thus candidates could understand what they needed to do, but not necessarily how. It is surmised that much of the functional language that occurred between the candidate and the supervisor took place during post teaching debriefings, but since that conversation is largely undocumented in the protocol, we can only speculate as to its presence or potential usefulness to the candidate. There were not many instances of specific recommendations being offered in the final section of the protocol, even though that was an explicit feature of the document.

An interesting phenomenon that occurred in relation to the use of evaluative language, was the relative absence of negative comments, and, for a very small number of supervisors, lack of any overtly positive qualifiers (very, highly, extremely, etc). Most of the written comments by supervisors designated as evaluative in nature, were positive, usually explicit but some implicit.

> "Excellent descriptive entries for each student." (#4SP09)

> "She is open to suggestions to improve on her skills." (#9SP09)

Of the total (298) evaluative written comments, only 7 were categorized as negative in tone or content.

> "Too ambitious…". (#3F08)
> " A good job, but…" . (#8SP09)

The question of why supervisors tend to record vastly more positive or even neutral comments than negative, especially in initial observations, given both the functional and evaluative roles they carry, needs to be examined further.

Types of Language: Cooperating Teachers

As noted earlier, many cooperating teachers combined descriptive, functional and evaluative statements when critiquing student teacher performances, thus offering candidates more explicit feedback on both the quality and content of their performances.

> " The lesson was well prepared. The pacing was appropriate for second grade. Students made connections before reading, which was especially important for diverse learners." [#6SP09]

> " Graphic organizer was a useful strategy to build background information. Questioning before, during, and after was developmentally appropriate – good informal assessment.

> Reader Response – good formal assessment – be sure to review writing rubric. Wait a little longer for students' attention." [#10SP09]

To some extent, cooperating teachers were a bit more effusive in their commentaries than supervisors, which may simply reflect the nature of the cooperating teacher/student teacher relationship, typically more intense and of longer duration than the relationship between supervisor and teacher candidate.

Levels of Expertise

Candidates were rated by both sets of clinical supervisors and cooperating teachers as either competent or exemplary by the time they completed the student teaching experience. Each candidate in this study completed their clinical experiences during the respective placements noted. However, the "competent" and "exemplary" ratings varied markedly and less agreement between/among paired raters was noted. Although we did not expect 100% agreement between/among raters, we anticipated that the ratings of a "competent teacher" vs. an "exemplary teacher" would be distinguishable and thus comparable in both elementary and special education settings.

Several issues became apparent during the analysis of the language used to evaluate levels of candidates' performance. First, because supervisors had latitude in generating open-ended comments, not all observers responded to the task in the same way. A small number opted to score the candidate's performance using the rating scale only, omitting any commentary at all. A greater number of observers wrote comments that were not related directly to the content of the prompts or statements that standardized the protocol. For example, comments about clarity of directions and skill in implementation might show up in the Reflect section. For a few observers, comments about the quality of a student teacher's performance appeared to contradict the rating of the same competence. For example, a comment might say" assessment is very well done", yet receive a D/C (developing/competent) as a rating on the assessment criterion. In some instances, the overall rating might be E (exemplary), while the description of the teaching performance was characterized as routine.

A second issue involved what might be referred to as the observer's style of interaction or perhaps philosophical perspective regarding their role, as that was reflected in the nature of the commentary. In some instances, candidates' overall ratings improved from the first to the third, certainly a marker to be hoped for in the professional growth process. However, a close examination of the types or frequency of evaluative comments made in each supervisor report did not always support the idea of growth. That is, the comments remained generally similar in content and tone over time. This may have reflected a personal tendency on the part of the observer to maintain a "low-key" style, or it may have reflected a fairly common pattern in evaluation situations of deliberately starting low to account for novice behaviors, regardless of the actual performance level shown in the first observed teaching. Interestingly, there was an overall lack of specific recommendations for improvement (short and long-term) for a large number of observations, especially the first, at which point such recommendations might be expected to most help the candidate make progress. At least in terms of the cooperating teacher, it appeared that the better a student teacher's performance was deemed to be, the less specific feedback he or she received. Comments tended to be more generic in nature. this may be

because candidates and cooperating teachers generally have more time for discussion of specific details during the work day or week or, it may signal a cooperating teacher's belief that an "E" (exemplary) rating is self explanatory.

Finally, a lack of clarity in terms and/or lack of specificity in examples given, may have limited the usefulness of the document to serve its purposes: to inform the candidate of strengths and weaknesses in her/his performance, while providing guidance and support in developing or remediating such, and to provide documentation of the candidate's overall readiness to be certified and assume a position in the teaching field. It was difficult to tell, without examples or specific details, precisely what was meant by "well-planned" or "appropriate"; consequently, it was not clear how to understand the character of a lesson that was "extremely well planned", although, clearly, the supervisor or the cooperating teacher meant to confer exemplary status on the second lesson. Overall the lack of specific details or examples could make it difficult for the observer, as well as the institution, to defend an overall rating of competent rather than exemplary, without a clear distinction in language.

The actual ratings themselves also proved problematic in many instances as supervisors/cooperating teachers sought some other score, such as C+ or C/E on the observation protocols. This seems to point to the need for an expanded rating system because the narrow parameters of a 4-point scale seemed problematic.

Although the supervisors'/cooperating teachers' ratings did not demonstrate any remarkable pattern of inter- and intra-rater agreement, the statistically significant agreement may indicate areas of comparable programmatic lenses. For instance, the Plan section indicated *substantial* agreement (68%) regarding the supervisors' ratings of the candidates' need to plan for Developmentally Appropriate Instructional opportunities for the students they taught. In another example, under the Act strand, the Positive (Learning) Environment (69% agreement) refers to whether the candidate has created an environment that fosters student involvement. This might point to a particular set of expectations that each rater sought and observed (or not) in each of the candidates' practice despite the difference in setting. It should be noted that agreement among raters strictly means that they gave the student teacher the same rating.

The Supervisor/Cooperating Teacher agreement comparisons may illustrate some degree of programmatic commonality. As indicated, *moderate* to *substantial* agreement was noted across programs (Formal/Informal Assessment; Implements Suggestions, Accepts Constructive Criticism).

On the other hand, more field specific findings were evident in particular areas. For instance, in special education following policies and procedures is particularly stressed in pre-service programs as teacher candidates must know and adhere to federal/state special educational laws and specific regulations. Failure to do so in the special education field would indicate grounds for dismissal. Conversely, elementary education programs may not share this emphasis. Elementary educators may instead focus more on specific content and overall unit instruction than special education candidates as demonstrated in the *moderate* agreement mentioned.

Limitations

Several limitations must be noted as we unpack our understandings of this investigation. Although important, time did not permit in-depth questions of supervisors/cooperating teachers to determine the nature of their comments and/or ratings. Hence, the overall meaning of specific supervisors'/cooperating teachers comments/ratings about individual candidates, the expectation of the viewed lessons, or the comments made directly to student teachers about their lessons is not known at this time. Also, no clarity was garnered about the meaning of supervisors'/cooperating teachers' comments/scores that indicated an in-between rating of C+ or C/E.

Post observation conferences that took place with teacher candidates after each lesson, where the clinical supervisor made specific suggestions for the next lesson, are also missing from these forms. It is not clear what information was shared with the candidates at that time, and thus excluded from the written record possibly because it was viewed as redundant.

The small sample size does not seem to yield enough information about differences across programs. Our assumption about good lesson implementation, that is that candidates plan, act, and reflect on their lessons in clearly identifiable and similar ways across settings, may be faulty. In fact, special education and elementary education personnel may not view "good lessons" through the same lenses in different settings and across content areas.

Implications

Several implications are evident from these data analyses. First, it appears that the structure of the observational tool, which includes both open-ended sets of questions and a 4-point rating scale, may influence both the kinds and extent of written responses in ways unanticipated by the institution. Currently, our School of Education is using a new observation protocol with a 6-point rating scale that will allow clinical supervisors to better incorporate the variation found in this analysis regarding C+ or C/E ratings.

Second, the internal consistency (inter- and intra-rater agreement between open-ended and closed-ended items) is tenuous. Supervisors and/or cooperating teachers, thus, may need more preparation in the use of the observational protocol generally and per department to understand the specific purposes of the tool, both for individual (candidate) and institutional expectations. Issues about the instrument's purpose (formative or summative), the nature of the raters' task, and their explication of the student teachers' progress, need clarity.

Third, language used by supervisors and/or cooperating teachers to describe and evaluate candidates' performance may reflect technical constraints of the tool and supervisors' perceptions of institutional norms, as much as their own professional backgrounds and training. Supervisors' and/or cooperating teachers' "voices" often seem to reflect a technical/rational rather than interactive or transformative (Paris & Gespass, 2001) mode of discourse as they judge candidate practice. This might be indicative of how a difference in their perspectives affects overall assessment of candidates' competence. Further, approaches to supervision, including language use, may reflect elementary or special education orientations.

This investigation focused primarily on the procedures and language of clinical observation. Another fruitful avenue to better understand the meaningfulness of clinical supervision would be to explore the nature of the interpersonal relationship that develops between the observer and the student teacher. Such an examination may help us to reframe this critical phase of a teacher candidate's preparation and to account for variability across settings and populations.

Notably, teacher preparation programs are being asked to provide evidence of candidate effectiveness by a variety of constituencies such as, accrediting bodies, public schools, students and parents. This study should add to greater understanding of the role, influence, and perhaps ability of the university/college-based supervisor to provide credible evidence (Cochran-Smith, 2005) of candidate preparedness for beginning teaching.

Moreover, the complex nature of inclusive classrooms makes it imperative for schools of education and teacher preparation programs to streamline their efforts to prepare their candidates well for the diverse fields in which they will work. This study helps to provide guidance to preparing supervisors and cooperating teachers to carry out their responsibilities in a coherent fashion within and across general and special education programs as well. It also contributes to the reexamination of the structure of preparation of dually certifiable teacher candidates.

References

Albi, L.D., Clifford, J.R., & Macy, M.G. (2005). A model of clinical supervision for preservice professionals in early intervention & early childhood special education. *Topics in Early Childhood Special Education, 25*, 167-176.

Blanton, L.P, Sindelar, P.T., & Correa, V.I. (2006). Models and measures of beginning teacher quality. *Journal of Special Education, 40*, 115-127.

Cochran-Smith, M. (2005). Studying teacher education: What we know and need to know. *Journal of Teacher Education, 56*, 301-306.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Coladorci, T., & Breton, W.A. (1997). Teacher efficacy, supervision, & the special education resource-room teacher. *Journal of Educational Research, 90*, 230-239.

Darling-Hammond, L. & Bransford, J. (eds.). (2005). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco, CA: Wiley and Sons.

Denzin, N.K. & Lincoln, Y.S. (2000). *Handbook of qualitative research (*2nd ed.*)*. Thousand Oaks, CA: Sage Publications.

Good, T.L., & Brophy, J.E. (2000). *Looking into classrooms* (8th ed.) New York: Longman

Hamel, F.L. & Merz, C. (2005). Reframing accountability: A preservice program wrestles with mandated reform. *Journal of Teacher Education, 56*, 157-167.

Interstate New Teacher Assessment & Support Consortium. (2001, May). *Model standards for licensing general and special education teachers of students with disabilities: A resource for state dialogue*. Washington DC: Council of Chief State School Officers.

Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage Publications.

Lang, W. S., & Wilkerson, J. R. (2007). *Assessing teacher competency : Five standards-based steps to valid measurement using the CAATS model*. New York: Corwin Press.

Moore, R. (2003). Reexamining the field experiences of preservice teachers. *Journal of Teacher Education, 54*, 31-42.

National Council for Accreditation of Teacher Education. (2002). *Professional standards for accreditation of schools, colleges, and departments of education*. Washington, DC: Author.

Paris, C. & Gespass, S. (2001). Examining the mismatch between learner-centered teaching and teacher-centered supervision. *Journal of Teacher Education, 52*, 398-412.

Prater, M.A. & Sileo, T.W. (2004). Fieldwork requirements in special education preparation. *Teacher and Special Education, 27*, 251-263.

Rademacher, J.A.,Wilhelm, R.W., Hildreth, B.L., Bridges, D.L., & Cowart, M.F. (1998). A study of preservice teachers' attitudes toward inclusion. *Educational Forum, 62*, 154-163.

Raths, J. & Lyman, F. (2003). Summative evaluation of student teachers: An enduring problem. *Journal of Teacher Education, 40*, 206-216.

Sim, J., & Wright, C.C. (2005). The Kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy, 85*, 257-268.

Tucker, P.D. & Stronge, J.H. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: ASCD.

Van Laarhoven, T., Munk, D.D., Lynch, K., Wyland, S., Dorsch, N., Zurita, L., Bosma, J., & Rouse, J. (2006). *Teacher and Special Education, 29*, 209-212.

Wilson, S. M., Floden, R. E., & Ferrini-Mundy, J. (2002). Teacher preparation research: An insider's view from the outside. *Journal of Teacher Education, 53*, 190-204.

Zeichner, K. (2005). Learning from experience with performance-based teacher education. In Peterman, F. P. (Ed.) *Designing performance assessment systems for urban teacher preparation.* Mahwah, NJ: Lawrence Erlbaum Associates.