

# JUDGING COMPETENCE: OBSERVING STUDENT TEACHERS IN DIVERSE FIELDS

Marie A. Lynch, Ph.D.  
Linda Capalbo, Ph.D.

Rhode Island College  
Providence, Rhode Island  
United States of America  
[mlynch@ric.edu](mailto:mlynch@ric.edu)  
[lcapalbo@ric.edu](mailto:lcapalbo@ric.edu)

## 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 a Northeastern state, 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) examine the language used by the observer that both describes and evaluates the student teacher's performance, particularly as it differentiates levels of expertise; 2) explore the level of agreement offered within and across (multiple) reports about the same teacher candidate; and 3) outline the development of meaningful processes for evaluating teacher candidates during the student teaching experience.

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

## 1 PURPOSE

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". [6, p. 412]. While all teacher preparation programs have developed or selected tools to aid in the evaluation of student teachers' performance [2; 12], we have little insight as to how or how well those tools are used by supervisors to capture practice and to provide meaningful feedback for the candidate as well as the institution.

The National Council for Accreditation in Teacher Education (NCATE) has offered guiding standards that help shape what Schools of Education must consider when assessing teacher candidates in the field. NCATE, Standard 3, describes the importance of field experiences and clinical practice in the evaluation of teacher candidates, while emphasizing the need for partnerships between school districts and higher education institutions in teacher preparation. Additionally, Standard 3 addresses the conditions of clinical practice for which candidates and institutions must prepare and for which they will be held accountable, including practice in diverse settings [14].

Although the parameters for gauging effective practice delineated by NCATE and other accrediting bodies 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, [18], many or most of whom are adjunct faculty, the inherent imperfections of any particular observational tool [19], and the widely varying circumstances in which teacher candidates find themselves practicing [9; 8], how *do* professionals gauge the levels of expertise in planning, acting, and reflecting demonstrated by their novice practitioners? Further, how do supervisors, in institutional documents, account for, explain, or inform stakeholders of the decisions made?

This study analyzed written records created by college clinical supervisors of student teaching observations carried out during the Fall 2008 & Spring 2009 semester. Observations, conducted in public schools in a Northeastern state, reflect 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 serving students with special educational needs. The 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; 2) explore the level of agreement offered within and across (multiple) reports about the same teacher candidate; and 3) outline the development of meaningful processes for evaluating teacher candidates during the student teaching experience.

### **1.1 Theoretical Framework**

Current research on effective teachers has come to focus, not only on the acts of effective teaching, but the results of effective teaching in terms of student learning [19]. Increasingly, teacher education programs have concluded that generating (and venerating) lists of teacher competencies [18], or collecting samples of course artifacts [9] fails to provide evidence of candidates' ability to effect positive outcomes for learners. Performance-based assessment of candidates during field placements, particularly student teaching or internship experiences, provides a powerful complement to portfolios and teacher tests [22].

Historically, numerous performance-based assessments have been used to assess teacher candidate development that represents the positions for which they are preparing. These include: successful coursework completion, specific field or practica experience, lesson development, special education planning, and the clinical practice of student teaching or an internship [2; 12].

For many teacher candidates, student teaching is often the culminating experience in teaching programs, which requires formal observation of the candidate actually teaching and instructing students by clinical faculty and experienced teachers in schools [8; 16]. During the student teaching experience, assessors include university supervisors who are required to call upon their own clinical expertise to assess the progress of the novice teacher, usually using tools designed or adopted by the accrediting institution. Researchers [6; 21] suggest that the nature of the support given to student teachers during their clinical experiences is key.

Some researchers suggest that the teacher education models of supervision, may be more ecologically appropriate for candidates in general education classrooms than for candidates in special education settings [1; 2; 5]. In fact, many of the protocols and processes that define what a clinical supervisor does in the field have been created and developed by general educators [1; 16].

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 their particular content strand [16; 17; 10]. Thus, many higher education institutions have provided general and special education program options (e.g. dual certification) to help candidates respond to the "increasing diversity and inclusiveness of public school classrooms" [20, p209].

Paris & Gespass [15] 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). Such thinking requires supervisors to move beyond a tendency to evaluate procedure rather than process [13].

Cochran-Smith [3] 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 within and across diverse field placements may provide one important link in that chain.

### **1.2 Methodology & Data Sources**

This mixed-method pilot study analyzed at least 66 written observational reports of 12 student teachers created by 6 elementary and 6 special education clinical supervisors, carried out during the Fall 2008 and Spring 2009 semesters. Each candidate was observed at least 6 times, 3 times in

either a traditional elementary or middle level classroom, and 3 times in a setting (e.g. general education class, self-contained class, or small group instruction) focused on teaching children with special educational needs. The protocols used, developed within the college's School of Education, were standard across placements and participants.

Data were analyzed quantitatively using descriptive statistics to determine how much agreement existed between raters of the same student teacher. The raters in this case are the elementary education supervisor and the special education supervisor. For the sake of these preliminary analyses, frequency data (e.g. percentage of agreement) will be offered.

Each of the observational reports required the clinical supervisor to rate the teacher candidate in terms of a 4-point rating scale (e.g. exemplary, competent, developing, & unsatisfactory). The following key is provided on the observational protocol about these terms:

- Key: E       exemplary/consistent performance above expected proficiency  
 C       competent/performs at expected proficiency  
 D       developing/performance needs improvement  
 U       unsatisfactory/performance is below expected proficiency  
 N.A.   not applicable

Text data in the form of open-ended commentary generated from the observational reports were also compared qualitatively using content analysis. This required the investigators to read, code, and categorize the text, to determine the relationship between/among the categories [11]. Content analysis allowed the authors to explore the relationship between elementary and special education observational commentary for the teacher candidates.

### 1.3 Results

To analyze data quantitatively, 72 observational reports, that documented the candidate teaching three times in each setting (e.g. elementary and special education), were considered. The observational protocol, standard across all observations, uses three strands aligned with the School of Education's Conceptual Framework that focuses on creating teachers that 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	Classroom	Implements
Appropriate Instruction	Management Techniques	Suggestions
Diverse	Oral/Written	Analyzes
Learner Needs	Communication Skills	Teaching
Uses Formal/Informal	Uses Formal/Informal	Follows Policy
Assessment	Assessment	& 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 statements that required the elementary or special education clinical supervisor to rate the student teacher on a 1-4 scale. The ratings ranged from 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.

Although results of this pilot study are still being analyzed, preliminary 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” rating. However, agreement became more random when rating student teachers as either “competent” or “exemplary”. The mean level of agreement was calculated across 6 observations, in each of the three strands, with their respective topics indicating a range from lowest to highest mean score in Table 2.

For example, under the Plan section, the mean agreement between supervisors ranged from 49% to 68%. The Act strand yielded a similar range from 47% to 69%. The Reflect portion indicated a slightly reduced initial score of 37% with an upper score of 66%. Table 2. also disaggregates mean agreement scores and topics below 60%, along with scores and topics starting at least 61%.

Table 2. Mean Agreement between Supervisors Across Observations

Strand	Topic	M	Topic	M
PLAN	Integrated Unit	.49	Appropriate Instruction	.68
	Selection of Content	.56	Formal/Informal Assessment	.66
	Diverse Learner Needs	.57		
ACT	Projects Professional Image	.47	Positive Environment	.69
	Instructional Opportunities	.50	Oral/Written Skills	.68
	Management Techniques	.53		
	Formal/Informal Assessment	.56		
REFLECT	Follows Policy & Procedures	.37	Accepts Constructive Criticism	.66
	Implemented Suggestions	.57	Analyzes Teaching	.62
	Works Collaboratively	.58	Interaction with Colleagues	.61

The open-ended commentary, generated from 66 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. The frequencies for each are shown in Table 3.

Table 3. Most Frequently Used Evaluative Phrases

Evaluative Phrases	SPED/ELED Competent		SPED/ELED Competent plus		SPED/ELED Exemplary	
Well planned	3	4	4	8	2	5
Well organized	2	1	3	1	0	0
Well implemented	2	1	2	1	0	0
Appropriate	5	4	6	2	3	1
Consistent	3	0	1	1	1	0
Works well with...	3	2	4	3	0	2
Professional	0	2	2	2	2	2
Clear	2	0	1	3	1	1
Engaging	1	1	0	2	0	1

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.

## 1.4 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. We had also hoped to be able to embark on the development of more meaningful processes by which to evaluate student teachers in the field. This seems premature to do as more data analyses are necessary.

### *Types of Language*

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 students 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.” (#\$SP09)

“She is open to suggestions to improve on her skills.” (#(SP09)

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

“Too ambitious...”. (#4F08)

“ A good job, but...” . (#\*SP09)

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.

### *Levels of Expertise*

Candidates were rated by both sets of clinical supervisors as either competent or exemplary by the time they completed the student teaching experience. It is also assumed that each candidate in this pilot study completed their clinical experiences during the respective placements noted.

The “competent” and “exemplary” ratings varied markedly and less agreement 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 exemplary, but 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 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.

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 meant to confer exemplary status on the second lesson. Additionally, it could be difficult for the supervisor/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 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 did not demonstrate any remarkable pattern of inter- and intra-rater agreement, the level of response above 60% may indicate areas of comparable programmatic lenses. For instance, the Plan section indicated higher agreement (68%) regarding the 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 clinician is seeking 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 Follows Policies & Procedures (37%) strand was the most discrepant among raters, which may indicate the difference in the two fields of education. For example, in special education following policies and procedures is particularly stressed in pre-service programs as teacher candidates must know and adhere to federal and 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 do not share this emphasis.

It is understood that while student teaching is the culminating experience in pre-service teacher education programs and deemed an important step toward becoming a professional in the field, the utility of the clinical supervisor's relationship to the candidate about this process is questionable. Only one of the clinical supervisors in this study was considered full-time faculty at this School of Education. Thus, 11 of the clinical supervisors were part-time adjunct faculty, who were all retired school teachers. None of these supervisors had historical relationships with the candidates prior to the three observations they conducted.

### **1.5 Limitations**

Several limitations must be noted as we unpack our understandings of this pilot study: Although important, time did not permit follow up questions of supervisors to determine the nature of their comments and/or ratings. Hence, the overall meaning of specific supervisors' 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.

Post observation conferences that took place with teacher candidates after each lesson, where the clinical supervisor made specific suggestions for the next lesson, are not captured on the forms. Thus, it is unclear what, if any, recommendations were made that influenced the student teaching practice from the first, second, to the third observations. The meaning of supervisors' comments/scores that indicated an in-between rating of C+ or C/E needs clarification as well.

Although comments and ratings by the cooperating teachers who supervised candidates on site have been secured, they have not yet been compared and analyzed. This coupled with the small sample size does not seem to yield enough data for comparisons across programs. Although 12 raters were also indicated, some student teachers shared the same rater in several instances. For example, two elementary and two special education supervisors observed 3 students each over the course of these semesters, while other supervisors observed 1-2 student teachers each. It is unclear what affect this had on the overall findings or on the student teachers themselves.

### **1.6 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 piloting a new observation protocol with a 6-point rating scale that will allow clinical supervisors more precision in their ratings.

Second, the internal consistency (intra-rater agreement between open-ended and closed-ended items) is tenuous. Supervisors, 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), and the nature of the supervisors' task, and their explication of the student teachers' progress, need clarity.

Third, reexamining the relationship between clinical supervisor and student teacher is paramount. Under what circumstances can candidates most benefit from the mentoring opportunities afforded them by their college supervisor, and how are those balanced with the evaluative nature of the interaction? In what ways can or should the process be made more collaborative, given the high expectations for candidate reflectivity embodied in the observation tool?

Fourth, data from these observation protocols need to be used to inform, not only the student teacher about strengths and weaknesses, but by extension, illuminate the strengths and weaknesses of the candidates' preparation programs. It is evident that these data can provide critical information to assist the institution's self-analysis of program effectiveness.



Finally, language used by supervisors 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' "voices" often seem to reflect a technical/rational rather than interactive or transformative [13] mode of discourse as they judge candidate practice. This might be indicative of how a difference in their perspective affects overall assessment of candidates' competence. Further, approaches to supervision, including language use, may reflect general or special education orientations.

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 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 to carry out their responsibilities in a coherent fashion within and across general and special education programs as well.

## 2 REFERENCES

### References

- [1] 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.
- [2] Blanton, L.P., Sindelar, P.T., & Correa, V.I. (2006). Models and measures of beginning teacher quality. *Journal of Special Education, 40*, 115-127.
- [3] Cochran-Smith, M. (2005). Studying teacher education: What we know and need to know. *Journal of Teacher Education, 56*, 301-306.
- [4] Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*, 37-46.
- [5] Coladorci, T., & Breton, W.A. (1997). Teacher efficacy, supervision, & the special education resource-room teacher. *Journal of Educational Research, 90*, 230-239.
- [6] 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.
- [7] Denzin, N.K. & Lincoln, Y.S. (2000). *Handbook of qualitative research* (2<sup>nd</sup> ed.). Thousand Oaks, CA: Sage Publications.
- [8] Good, T.L., & Brophy, J.E. (2000). *Looking into classrooms* (8<sup>th</sup> ed.) New York: Longman.
- [9] Hamel, F.L. & Merz, C. (2005). Reframing accountability: A preservice program wrestles with mandated reform. *Journal of Teacher Education, 56*, 157-167.
- [10] 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.
- [11] Krippendorff, K. (1980). *Content analysis: An introduction to its methodology*. Beverly Hills, CA: Sage Publications.

- [12] 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.
- [13] Moore, R. (2003). Reexamining the field experiences of preservice teachers. *Journal of Teacher Education, 54*, 31-42.
- [14] National Council for Accreditation of Teacher Education. (2002). *Professional standards for accreditation of schools, colleges, and departments of education*. Washington, DC: Author.
- [15] Paris, C. & Gespass, S. (2001). Examining the mismatch between learner-centered teaching and teacher-centered supervision. *Journal of Teacher Education, 52*, 398-412.
- [16] Prater, M.A. & Sileo, T.W. (2004). Fieldwork requirements in special education preparation. *Teacher and Special Education, 27*, 251-263.
- [17] 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.
- [18] Raths, J. & Lyman, F. (2003). Summative evaluation of student teachers: An enduring problem. *Journal of Teacher Education, 40*, 206-216.
- [19] Tucker, P.D. & Stronge, J.H. (2005). *Linking teacher evaluation and student learning*. Alexandria, VA: ASCD.
- [20] 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.
- [21] 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.
- [22] 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.