

March 2001

## The Development Process of an Assessment Instrument for the Supervision of Teacher Education Interns

Tuula Heide  
*Humanities Polytechnic, Kuopio Unit, Finland*

Follow this and additional works at: <https://pdxscholar.library.pdx.edu/nwjte>



Part of the [Education Commons](#)

Let us know how access to this document benefits you.

---

### Recommended Citation

Heide, Tuula (2001) "The Development Process of an Assessment Instrument for the Supervision of Teacher Education Interns," *Northwest Journal of Teacher Education*: Vol. 1 : Iss. 1 , Article 1.

DOI: 10.15760/nwjte.2001.1.1.1

Available at: <https://pdxscholar.library.pdx.edu/nwjte/vol1/iss1/1>

This article is brought to you for free and open access. Creative Commons CC-BY-NC-SA: This article is distributed under the terms of the Creative Commons Attribution 4.0 License ( <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

# The Development Process of an Assessment Instrument for the Supervision of Teacher Education Interns

Tuula Heide, Ph.D.

Humanities Polytechnic, Kuopio Unit, Finland

## ABSTRACT

*The purpose of this article is to describe the development process of an assessment instrument for the supervision of teacher education interns. The assessment instrument was developed in several phases. After literature review in the area of supervision and evaluation of teacher education interns, a steering committee and the researcher developed the first draft of the instrument. The second draft of the instrument was developed based on the survey results asking for respondents' perceptions of the first draft. Finally, the final draft of the instrument was constructed after a pilot study, based on the results of the surveys administered at the end of the testing period.*

## INTRODUCTION

Evaluation of teacher education interns is an essential part of teacher education programs. Final evaluations "purport to distinguish among competent and incompetent, effective and less effective, talented and less talented, outstanding, average, and below-average students regarding their potential as teachers" (Guyton, & McIntyre, 1990, p. 525). Feedback should assist interns to identify strengths and weaknesses in order to improve their teaching practices. Evaluation standards should follow the actual standards utilized.

Typically teacher education interns gain feedback from their mentor teachers and university supervisors. Effective feedback from mentor teachers is an essential part of student teaching (Blocker & Swetnam, 1995; Guyton & McIntyre, 1990; Metcalf, 1991; Ramanathan, 1996; Wilkins-Canter, 1997). Feedback that constructively helps the intern grow will have a positive impact on the intern's professional development. Studies have found that university supervisors tend to provide more evaluative

feedback to interns than do mentor teachers (O'Neal, 1983; Reiff, 1980; Zimpher, DeVass, & Nott, 1980). This difference in supervisory roles has been explained by rater bias and the close relationship between the mentor teacher and teacher education intern.

The evaluation process itself is often referred to as formative or summative evaluation (Acheson, 1989; Sweeney, & Manatt, 1986). Formative evaluation is a continuous process, in which the evaluator provides specific feedback. The purpose of formative evaluation is to provide teachers opportunities for improvement. Summative evaluation, on the other hand, serves as an indicator of accountability, identifies strengths and weaknesses, and sets goals for the future. It utilizes the performance data from the whole period of observation. The summative evaluation for teacher education interns is usually conducted at the end of the student teaching period. At the institution of this study, the University of Idaho, the teacher education intern usually receives two formal teaching evaluations: at mid-term and at the conclusion of the student teaching period (Schmidt, 1996).

## ASSESSMENT INSTRUMENTS

Assessment instruments of teaching can be generally categorized into two types: high or low inference measurement system (Andrews, & Barnes, 1990). A high inference system requires the evaluator to operate at a high level of abstraction. A low inference system is based on actual data that supports the evaluator's operation. The main difference between the two systems is the demands that are placed upon the observer. Even though there are advantages and disadvantages for both of the systems, the bottom line is that the scoring system should serve the overall philosophy of the evaluation system.

Teacher education intern assessment instruments should be reliable and valid because of their major function as screening devices for entry to the profession (Guyton, & McIntyre, 1990). The criteria for evaluation should be the same for both the mentor teacher and the university supervisor. The evaluation criteria should be related to teacher education program's goals.

## RATING SCALES

Rating scales are commonly used when evaluating teacher education interns. There are many recommendations related to the development of rating scales. Worthen, White, Fan, & Sudweeks, (1998) suggest several ways to strengthen scales and making them appropriate. First of all, the authors stress the importance of providing definitions for each of the performance level. Secondly, the definitions should be so clearly spelled out that it does not leave space for interpretations. Thirdly, the authors suggest that in order to prevent a difficulty of deciding between two numbers on the scale, a graphic scale should be provided. On a graphic scale, evaluators are given the freedom of choosing to mark any point in between the two numbers.

Some additional recommendations that Worthen et al. (1998) provide include focusing on specific observable behaviors. Furthermore, the motivation and training of the evaluators is highlighted. The traits of characteristics of the

person being evaluated should not be included in the evaluation, because those features cannot be defined or proved. It is emphasized that the motivation of evaluators will produce more accurate ratings, and along with that, the supervision procedure will provide an opportunity to explain the purpose and answer questions that might arise.

Few overall recommendations for teacher evaluation forms deal with similar issues explained above. One article (The five essentials of a teacher evaluation form, 1995) suggested that all the rating scales should be accompanied by space for qualitative comments and there should be space provided for overall comments and self-evaluation. Signature lines for each parties involved should be included in the evaluation instrument.

Many of the suggestions and recommendations that the research provides concerning the assessment tools and rating scales are also supported by a study, in which five mentor teachers and five university supervisors were interviewed regarding their perceptions of an assessment tool currently used at the institution of the study (Heide, 1999). The study indicates that the rating scale criteria were perceived as inadequate and the construction of the questions was not clear. These findings further support the need for a new assessment tool for the supervision of teacher education interns at the University of Idaho.

## DEVELOPMENT OF ASSESSMENT INSTRUMENTS

The literature contains only a few descriptions of the development process of assessment instruments for the supervision of teacher education interns. Hartsough, Perez, & Swain (1998) developed a new assessment instrument using the Behaviorally Anchored Rating Scales (BARS) technique. Twenty-eight preservice teacher supervisors identified teacher education interns by giving concrete examples that would fit the categories that were developed from the six domains of the beginning teacher support and assessment program and five central concepts of the National Board for Professional Teaching

Standards. Furthermore, the study identified the most valid descriptors for teacher education interns and then 64 participants indicated a numerical value for each descriptor. The results indicated agreement among supervisors on the allocation of descriptors to concepts and high interjudge reliability for the assigning of number values to descriptors. The study suggests that teacher education programs should use the scaling process when they develop rating components for their assessment systems. Assessment procedures should also be closely examined due to reforms in teacher education.

Brooker, Muller, Mylonas, and Hansford (1998) developed an assessment instrument for assessing final-year practice teaching. For their first version, six criteria (planning and preparation; communication and interaction; teaching for learning; managing the learning environment; assessment and evaluation; and professionalism) were selected from local (Australia) competencies for beginning teachers, existing practicum assessment frameworks, and hiring criteria. The first construction of the instrument was piloted with 18 teacher education interns. The supervising teachers received guidelines for using the instrument in writing. After the piloting, interns were interviewed about their perceptions about the new assessment tool. For the second piloting, modifications were made for the questions and the guidelines, and the tool was used by 24 supervising teachers. Those teachers were also interviewed after the piloting. The tool in question was still under construction, but the authors conclude that the development of an assessment instrument is not merely enough. It was stressed that the development process should be supported by appropriate training of the educators using the instrument.

### **CONTEXT OF THE STUDY**

The University of Idaho, College of Education is redesigning its teacher education program, and as part of that, the student teaching experience changed from a one semester student teaching to a year long internship. The College of Education has adopted the The In-

terstate New Teacher Assessment and Support Consortium (INTASC) standards, and also follows the standards adopted by the state of Idaho (Idaho Core Standards). This study conceptualized, designed, and tested an instrument for mentor teachers and university supervisors to evaluate the performance of teacher education interns. The development process reflected the INTASC standards, the Idaho Core Standards, and the supervisors' opinions and perspectives.

### **METHOD**

The purpose of this article was to describe the development process of a new assessment instrument for mentor teachers and university supervisors to consistently and accurately supervise teacher education interns' performance related to the INTASC standards and Idaho Core Standards.

It was determined, as a result of a careful examination of the program and the standards adopted, that there was a necessity for a new assessment instrument. The instrument would have to reflect the standards adopted and meet the goals of both the public school program and the university program.

### **PARTICIPANTS**

A panel of experts was established for the development of the assessment instrument. This steering committee consisted of 12 educators and the researcher. More specifically, one member of the steering committee was a current student teacher, three were mentor teachers, five were practicing teacher educators, and three were administrators and/or university supervisors in higher education. All members of the committee were selected based on recommendations and their contributions to the field of Education.

### **INSTRUMENTS AND PROCEDURES**

The development process of the assessment instrument, after an extensive literature review, was initiated by introducing the study to the Dean, the Assistant Dean, and to the Administrative team at the College of Education, the

University of Idaho. Consent to conduct the study was granted. Following approval from the College of Education, the researcher invited 12 well-known educators to form a steering committee to design the first two drafts of the instrument. All of those invited expressed their interest to participate in the study. The researcher chaired two full day meetings approximately seven weeks apart from each other.

### FIRST STEERING COMMITTEE MEETING

The purpose of the first meeting was to review research on supervision of interns and to develop a first draft of a potential assessment instrument. Ten members of the steering committee were present in this initial gathering. After introductions, the researcher gave a presentation of the study by reviewing research related to the development of an instrument, explaining the goals and objectives of the study and the purpose of the steering committee. The participants of the steering committee unanimously expressed that there is a need for a new assessment tool to supervise teacher education interns.

Next, various examples of different types of rating scales and assessment forms were provided, and a working copy of a possible new instrument developed by the researcher was shared. The purpose of the rough draft was to furnish participants with an instrument that would serve as a starting point for brainstorming activities and that would also promote their thinking process. The developed draft reflected the deemed standards, but the performances taken from the Idaho Core Standards were considerably condensed from the original ones. For the development of this tool, the researcher used all ten INTASC principles as they are and, under each of those, provided criteria from Idaho Core Standard's performances for new teachers.

Following the presentation, the group was divided into two smaller groups, in which the members reviewed the rough draft of the instrument, brainstormed new ideas and prepared for reporting their perceptions to the committee as a whole. Both groups reported their findings and consensus was reached on the first draft of

the instrument.

As a result of the first steering committee meeting, a first draft of the assessment instrument was developed. More specifically, it was suggested that there would be two separate forms: a one-page form for final evaluation and a longer form for the ongoing intern evaluation. The committee felt that there is a need for two individual forms, so that the final tool would provide brief information to determine if the intern met the requirements for INTASC standards. The purpose of the ongoing assessment form is to hold interns accountable for providing and recording evidence of their accomplishments. At the same time, the ongoing tool would serve as a progress report of the professional development of the interns. The tools developed were named Final Intern Evaluation and Ongoing Student Teaching Evaluation.

It was recommended in the steering committee meeting that the researcher would, after designing the draft of the instrument based on the recommendations, send the draft via e-mail to each of the committee members for critiquing. The mailing was conducted four days after the meeting, and by the due date given (seven days), four members responded to the draft. Responses, which were mainly minor editorial comments, were taken into consideration and implemented if applicable.

### SURVEY OF THE FIRST DRAFT OF THE INSTRUMENT

The first draft of the instrument, along with a survey asking for respondents' perceptions of the Final Intern Evaluation and the Ongoing Student Teaching Evaluation tools, were mailed to 287 educators in Idaho, Montana, Washington, Oregon, California and Arkansas. Both mentor teachers (n=170) and university supervisors (n=36) were selected based on intern supervision experience in the previous semester (fall 1999). The entire faculty (n=81) at the College of Education, University of Idaho, received the survey as well. In addition to the total population, 54 College of Education students were surveyed regarding their perceptions of the tools.

After the initial mailing, a follow-up letter

was sent to all the non-respondents, and the researcher called most of them reminding to return the survey. By the due date given, 42.5% (n=122) of the participants had responded to the questionnaire. Of those respondents, 28,9% (n=23) were College of Education faculty members, 44,7% (n=76) were mentor teachers, and 65,1% (n=23) were university supervisors. The return rate was accepted as sufficient for the study.

## DATA ANALYSIS AND RESULTS

The data from surveys was analyzed to identify patterns related to how participants perceived the Final Intern Evaluation and the Ongoing Student Teaching Evaluation tools. In this study, a pattern is viewed as a recurring behavior that is apparent and occurs at least three times or more. The responses were treated separately for supervisors (including College of Education faculty, mentor teachers and university supervisors) and Education students. In the following sections, the survey results regarding the Final Intern Evaluation-tool, Ongoing Student Teaching Evaluation-tool, and the rating scale, are reported. Furthermore, overall comments related to the instrument are presented as well.

### FINAL INTERN EVALUATION FORM

Many of the respondents, both supervisors and students, were concerned if the form is appropriate for all the different content areas. It was deemed the content knowledge should be more specifically addressed, especially in the secondary level. It was suggested that the ten general principles should be broken down to reflect strengths and weaknesses by subject area. Some of the students wished there was more space for comments and that the criteria were more specific.

### ONGOING STUDENT TEACHING EVALUATION-FORM

Supervisors overwhelmingly declared that the form did not address classroom management: "I don't see anything on the Ongoing Student

Teaching Evaluation about classroom management. This is a key skill for all successful teachers." Further, another university supervisor reported:

In addition to these straight-forward and comprehensive evaluation instruments, it is recommended that the area of classroom management be spelled out, e.g., 1) To what degree does a teacher intern establish and clearly communicate parameters for the learner's self-management?, 2) Does the intern endeavor to search out and solve the causes of inappropriate learner self-management?, 3) Does the intern use various techniques to maintain appropriate learner self-management?

Some of the participants considered that the form did not reflect enough of teacher's personal attitudes. A mentor teacher described: "Not much of character traits, collegial relationships, professional attitude, etc."

Numerous participants, both supervisors and students, declared that there was not enough space to report each of the four sets of evidence. Besides, students proposed that adding an area for improvement might strengthen the ongoing tool. Some students were concerned about the length of the instrument, although it was also indicated, that the length is appropriate if the university supervisor is adequately involved (visits frequently and provides quality feedback) in the supervisory process.

### RATING SCALE

Final Intern Evaluation form: The rating scale on the Final Intern Evaluation-form was designed to provide information regarding the intern's performance in relationship to INTASC standards. The scale, principle met/not met, resulted mixed responses; it was conceived as either positive or negative. Half of the respondents liked the 'principle met/not met' criteria, whereas the rest thought that the "cut and dry – not enough options" was not very beneficial. A student stated: "I feel it would be more useful to the teacher, and to future employers, if it were

rated on a scale. I would want to know how well I met these principles.” A mentor teacher reported: “I would like to see a likert or rubric type scale instead of met/not met. That way interns can see their strong and weak areas.” Furthermore, a faculty member stated:

I like the categories better than the ones on the current evaluation, which I felt overlapped. I like the idea of the number rating on the current form. The new proposal is a pass/fail which I feel does not indicate where improvement is needed even though the minimum expectations were met.

Respondents also suggested that if the criterion was going to stay as it was, comments should accompany both the principle met and principle not met rating.

**Ongoing Student Teaching Evaluation** form: The rating scale on the Ongoing Student Teaching Evaluation form identifies three levels of performance; 1) limited evidence, 2) clear evidence, and 3) clear, consistent, convincing evidence. Many respondents thought that the rating scale was not clear enough. Generally, they did not understand what the differences between the levels were. Especially differentiating level 2 more from level 3 was strongly recommended. It was commented, however, that the scale was “good – leaves plenty of room of improvement by intern.” Also, supervisors liked the fact that the interns would be evaluated four times during the internship year: “I like the idea of four ratings. Makes it clear and concise”.

## OVERALL COMMENTS

Many of the supervisors felt that the new tool was better than the one they had used in the past. They commented that the instrument was concise, well-worded and easy to follow. The organization of the form was perceived as efficient and relevant. Several participants liked the fact that the interns were being held accountable for providing evidence of their accomplishments. As one of the respondents put it: “I am thrilled to see the responsibility is with the intern for providing/recording evidence of accom-

plishments.” Also a need for training was brought up by many participants: “I think a workshop would need to be conducted so that everyone using the form(s) understand how to complete.”

Also students liked the organization of the form, hence they thought the forms were clear and easy to understand. Some respondents, however, stated that the language used was too complicated: “Too much teacher jargon” and “use less confusing words” were comments from some of the students. Furthermore, students addressed the issue of intern accountability. One of the respondents stated that evidence provided is “a great idea” and “much more helpful than just numbers”. Overall, especially the students, felt that this tool was better than the current assessment tool. One aspect that the students were concerned about was if the supervisors would have sufficient time allocated for supervision in order for them to provide accurate evaluation.

Overall, participants liked the final form, mostly because of its comprehensiveness and conciseness. Also the formative assessment instrument was considered as a viable tool. A university supervisor said: “Bottom line is feedback to improve our ‘teachers in the making’ – these forms are a step in that direction.” Furthermore, a mentor teacher assured: “I would feel comfortable using both evaluation tools.” All the responses from the surveys were organized for the second steering committee meeting that held briefly after the data collection was finalized.

## SECOND STEERING COMMITTEE MEETING

The second steering committee meeting was devoted to do the revisions needed based on the survey results, and to form a second draft of the assessment tool. Seven committee members were present in this meeting. The researcher presented the results and, following the 30-minute opening, two smaller groups reflected on the discussion questions provided. The discussion questions were quotes and issues for discussion drawn from the data. The purpose of the discussions was to provide recommendations as to how to deal with the issue in question that

the groups were assigned to talk about. Finally, the goal was to reach large group consensus on recommendations.

As a result of the second steering committee meeting, the researcher modified and altered some details on the instrument. On the Final Intern Evaluation (FIE) tool, there was space added to identify the internship site and the semester. Furthermore, a line explaining the purpose of the signatures on the form was included. Also, on the Ongoing Student Teaching Evaluation form a location for demographic information was added. Some editorial suggestions were implemented, words added for clarification in the “directions” section, more space was granted in the “evidence provided” section, and space for signatures for each of the conference times were added. Also, for the clarity of the instrument, descriptive titles for each of the principles were added. In addition, a feedback form was provided for a written summary of goals agreed upon. It was intended that after discussing the intern’s performance during scheduled meetings, goal setting should be the next step for personal growth. The name of the tool was changed to the Growth Assessment Instrument (GAI). It was decided that the entire instrument, including the FIE and be named as Intern Performance Assessment Tool (IPAT).

## DISCUSSION AND FUTURE STEPS

The IPAT was developed in four phases. First, the researcher did an extensive research of studies in the area of supervision and evaluation of teacher education interns. Second, a steering committee and the researcher developed a first draft of the instrument, after which other stakeholders’ feedback was asked for to further develop the tool. Third, the steering committee and the researcher made modifications and alternations based on the feedback acquired and produced the second draft of the instrument.

Finally, the second draft of the instrument was piloted during the fall 2000. Mentor teachers (n=9) and university supervisors (n=7) reviewed interns’ (n=9) performance using the GAI twice during the eight-week student teaching (for the purposes of this study, only the first

half of the semester was taken into account). The final evaluation was conducted using the FIE tool with the support and assistance of the GAI. To assure that all the participants used the instrument as accurately as possible, a training workshop for the use of the instrument was held at the beginning of the student teaching period. At the end of the pilot study, the participants were surveyed regarding their perceptions of the IPAT.

The final draft of the IPAT was modified based on the descriptive data given by the participants of the pilot study. The following issues were addressed when implementing the changes: The length of the IPAT, time, repetition of performances and met/not met category.

**The length of the IPAT:** The sections for evidence provided were made shorter in order to be able to reduce the length of the form. The purpose is to use a computerized version of the instrument. Consequently, based on the space needed for comments, the boxes can be altered. In addition, a WEB-site that would facilitate cooperation between the intern, mentor teacher, and university supervisor, should be constructed to allow opportunities for ongoing discussions and time to reflect over those deliberations.

**Time:** It was identified that the use of the IPAT is too time consuming. When the standards are fully implemented in the Teacher Education Program, perhaps the process will not be viewed as time consuming. This assumption is based on the students being introduced to the standards throughout the program in relationship to the IPAT. As a result, expectations will be known before internship assignment and a student can start collecting journals, observations, lesson plans and other appropriate items and use them in creating of the portfolio

**Repetition of performances:** Although the participants of the pilot study suggested that the performances on the GAI were repetitive, performances were not altered due to the fact that the State Board of Education has approved the standards and therefore any changes were not viewed as appropriate.

**Met/not met categories:** It was decided that the met/not met categories would remain on the FIE form. There was an additional explanation added to the comments-section. It was recom-

mended that the section would be used to explain what is needed for an unsatisfactory standard to be met. The section should also be used for providing information regarding the factors that might have affected intern's performance, such as inadequate possibilities to perform to the fullest potential in the context of the internship. It should be noticed that it is strongly recommended that comments be given also when the standard was accomplished.

Adding the modifications reported above will strengthen the IPAT. The redesign of the Teacher Education Program in the College of Education at the University of Idaho was based on INTASC standards and implemented in the fall 1998. The first graduating group of students from the redesigned program will be in the year 2003. Therefore, it may be appropriate to conduct a study at that time to determine if goals and objectives of the program made a difference on how respondents perceive the use of IPAT.

## REFERENCES

- Acheson, K. A. (1989). Evaluating, supervising, and analyzing teaching. OSSC Report, 29 (3), 1-7.
- Andrews, T.E., & Barnes, S. (1990). Assessment of teaching. In W.R. Houston (Ed.), Handbook of research on teacher education, (pp. 569-598). New York: Macmillan Publishing Company.
- Blocker, L.S., & Swetnam, L. A. (1995). The selection and evaluation of cooperating teachers: A status report. The Teacher educator, 30 (3). 19-30.
- Brooker, R., Muller, R., Mylonas, A., & Hansford, B. (1998). Improving the assessment of practice teaching: A criteria and standards framework. Assessment and Evaluation in Higher Education, 23 (1), 5-25.
- Guyton, E., & McIntyre, D. J. (1990). Student teaching and school experiences. In W. R. Houston (Ed.), Handbook of research on teacher education (pp. 514-534). New York: Macmillan Publishing Company.
- Hartsough, C. S., Perez, K. D., & Swain, C. L. Development and scaling of a preservice teacher rating instrument. Journal of Teacher Education, 49 (2), 132-139.
- Heide, T. (1999). Teacher education interns' evaluation: Solidarity across difference, University of Idaho. Paper presented at the Northern Rocky Mountain Educational Research Association 17<sup>th</sup> annual Conference, Jackson, Wyoming.
- Metcalf, K. K. (1991). The supervision of student teaching: A review of research. The Teacher Educator, 26 (4), 27-42.
- O'Neal, S. F. (1983). Supervision of student teachers: Feedback and evaluation. Austin: University of Austin, R & D Center for Teacher Education.
- Ramanathan, H. (1996/1997). Perceptions of cooperating teachers, university supervisors and directors of field experiences of the evaluation process in early field experiences. (Doctoral dissertation, The Ohio State University, 1996). Dissertation Abstracts International, 57 (9), 4332.
- Reiff, J. C. (1980). Evaluating student teacher effectiveness. College Student Journal, 14 (4). 369-372.
- Schmidt, W. (1996). Student teaching and internship in teaching handbook. Moscow: University of Idaho.
- Sweeney, J., & Manatt, R. P. (1986). Teacher evaluation. In Ronald A. Berk (Ed.), Performance assessment: Methods and application (pp. 446-468). Baltimore: The Johns Hopkins University Press.
- The five essentials of a teacher evaluation form. (1995, October). Curriculum Review, 35 (2), 10.
- Wilkins-Canter, E. A. (1997). The nature and effectiveness of feedback given by cooperating teachers to student teachers. The Teacher Educator, 32 (4), 235-249.
- Worthen, B. R., White, K.R., Fan, X., & Sudweeks, R. R. (1998). Measurement and as-

essment in schools. New York: Addison Wesley Longman, Inc.

Zimpher, N. L., DeVass, G. G., & Nott, D. L. (1980). A closer look at university student teacher supervision. Journal of Teacher Education, 31 (4), 11-15.

***Dr. Tuula Heide*** is a lecturer at Humanities Polytechnic, Kuopio Unit, Kuopio, Finland.  
Email: [tuula.heide@humak.edu](mailto:tuula.heide@humak.edu)