



**Universiteit
Leiden**
The Netherlands

Assessment for learning to teach: Appraisal of practice teaching lessons by mentors, supervisors, and student teachers.

Tillema, H.H.

Citation

Tillema, H. H. (2009). Assessment for learning to teach: Appraisal of practice teaching lessons by mentors, supervisors, and student teachers. *Journal Of Teacher Education*, 60(2), 155-167. Retrieved from <https://hdl.handle.net/1887/14997>

Version: Not Applicable (or Unknown)

License: [Leiden University Non-exclusive license](#)

Downloaded from: <https://hdl.handle.net/1887/14997>

Note: To cite this publication please use the final published version (if applicable).

Assessment for Learning to Teach

Appraisal of Practice Teaching Lessons by Mentors, Supervisors, and Student Teachers

Harm H. Tillema
University of Leiden

Supporting student teachers in learning to teach is a collaborative effort by mentor teachers, teacher education supervisors, and student teachers. Each of the participants appraises effort and progress in learning to teach from different perspectives, however. This study explores how practice lessons are assessed by multiple raters. Teacher educators, mentor teachers, and student teachers (51 participants in total) were asked to appraise a practice lesson given by the mentored student. Alignment in rating was analyzed in 17 triads and compared with respect to purpose of assessment, object of appraisal, preferred methods, and focus of the appraisal as well as on the criteria used by the various assessors. Shared problems encountered during the appraisal were also gauged. Our findings indicate considerable variation in purposes and multiple perspectives in criteria among the different assessors. Differences and similarities among the stakeholders were interpreted as contributing to a multifaceted appraisal of accomplishments. Nevertheless, a shared, common ground is also needed to value the different aspects that should be included in an integrated or encompassing approach for assessment of learning to teach.

Keywords: *student teaching; assessment for learning; teacher quality; mentoring; teacher education; learning to teach*

Appraisal of practice teaching lessons is an important vehicle for informing the student teacher about accomplishments and prospects in teaching. It is for this reason that learning to teach from practice lessons is at the core of student teacher preparation programs (Abernathy, Forsyth & Mitchell, 2001; Furlong & Maynard, 1995). One of the key elements in learning to become a teacher is sharing and learning from experiences in close cooperation with practice teachers and teacher educators (Dall'Alba & Sandberg, 2006; Day, 1999; Edwards, Gilroy & Hartley, 2002). Teacher educators, student teachers, and practice teachers are all involved in this process in different ways. Whereas teacher educators seem more inclined to look at a student teacher's practice teaching from the perspective of program standards, and teacher mentors look at a student teacher's classroom performance and how it benefits pupils, the student teacher (as a learner) is more concerned with coping with the direct demands of teaching a class (Loughran, 2003, 2007; Grossman, 2006). It is important to gauge how these different perspectives can merge in an appraisal for supporting and stimulating a student teacher's learning and, more specifically, to determine how different stakeholders operate and appraise teaching prac-

tice lessons and how the assessment is understood by those involved in this assessment-for-learning process (Havnes & McDowell, 2007).

Assessment for Learning to Teach

Assessment is increasingly recognized as a valuable tool to promote learning (Assessment Reform Group, 1999, 2002; Black & Wiliam, 1998; Shephard, 2000). This learning-oriented, (in)formative assessment—that is, in the sense that formative assessment should be informative to the learner—needs to be distinguished from a summary or mandated assessment, which documents and appraises work performance in relation to external evaluation standards (Delandshere & Arens, 2003). Assessment in the latter instance focuses on establishment of marked achievements that may be appreciated and judged according to preestablished standards (Zuzowsky & Libman, 2002; Heilbronn, Jones, Bubbs, & Totterdell, 2002). As such, it has its own legitimized function in teacher education (i.e., serving an accountability warrant; Cochran-Smith & Fries, 2002).

Formative assessment, however, tries to document and illuminate the cyclical and extended process of professional growth and the building of relevant practice experiences (McMillan, 2007; MacLelland, 2004). This occurs through continuous monitoring across an extended period and is mainly aimed at student-oriented goals and individual learning needs (Edwards & Collison, 1996; Wang & Odell, 2002). Viewed this way, assessment aims at providing (in)formative feedback to help the student teacher gain insight into performance so that it is valuable to his or her professional growth (Boshuizen, Bromme, & Gruber, 2004; Brown & Glasner, 1999). Thus, assessment information is collected and communicated for its potential to change or direct the (student) teacher's development (Feiman Nemser & Remillard, 1996). Several framing factors have been identified (Kwakman, 2003; Smith & Tillema, 2003; Tigelaar, Dolmans, Wolfhagen, & van der Vleuten, 2002) that directly relate to the impact of assessment information on professional learning, for instance, type of assessment evidence collected, criteria used with respect to performance appraisal, or whether a relational or situational approach to feedback delivery is used (Tillema & Smith, 2003). These framing factors may variously affect what is acquired from practice experiences by the student teacher.

To complicate matters further, typically, several stakeholders are involved in assessment of learning to teach; they either implicitly or explicitly use these framing factors differently. These include mentor or practice teachers from practice schools, supervisors from teacher education institutes, and as is more often the case, (peer) student teachers (Darling Hammond, 2000; Wilson & Berne, 1999). Assessment in this case is a complex process of joint appraisal and judgment. Several framing factors play an intricate role in this process. It includes not just several assessors and their rating of practice teaching, but different assessment targets or goals may compete as well, along with various appraisal criteria, sources of performance evidence, and diverse intents to deliver informative feedback. A simple model of the isolated, impartial assessor who grades performance undisputed, on mutually accepted criteria, does not correspond to reality (Snyder, Lippincott, & Bower, 1998; Zeichner & Wray, 2000). Instead, several studies indicate that the different stakeholders hold a wide variety of perspectives on appraising student teachers during practice teaching (Atwater & Brett, 2005; Tillema & Smith, 2006; Wilson & Youngs, 2005; Zuzowsky & Libman, 2002). Mentoring practice teachers and supervising teacher educators differ in appreciation of teaching preparation and contents addressed in teacher education programs (Edwards et al., 2002), in mentoring approaches adopted for practice

teaching (Loughran, 2003; Nijveldt, 2007), and in applying criteria for successful teaching (Wang & Odell, 2002; Yinger & Hendriksen-Lee, 2003). Even student teachers disagree with their mentors and supervisors on the amount of support they need to regulate their own learning (Kremer Hayon & Tillema, 1999) or the feedback they need for learning to teach (Zeichner & Wray, 2000).

This variety of perspectives need not necessarily be detrimental to a valid and (in)formative appraisal. On the contrary, a multirater or multiple-perspective viewpoint may even enhance such an appraisal, because it can enrich the nature of informative feedback given to the learner (Atwater & Brett, 2005; Byham, 1996; Thornow, 1993; Darling Hammond & Bransford, 2004). Multirater assessments, such as 360-degree feedback (Waldman & Atwater, 1998), have been successfully used, for instance, in workplace learning and performance appraisal to provide an in-depth and multidimensional view on acquired expertise in practice settings (Boshuizen et al., 2004; Dall'Alba & Sandberg, 2006; Kirby, Knapper, Evans, Carty, & Gadula, 2003).

Shared appraisal is now being widely adopted in work-related settings in many professional fields (i.e., nursing, hospitality management; Baum, 2002). As an assessment tool, this multirater assessment has been found to motivate learning, augment follow-up on feedback recommendations, and advance favorable attitudes toward the improvement of future performance (Jellema, 2003; Maurer, Mitchell, & Barbeite, 2002). Appreciation of multirater assessment predominantly derives from the recognition that no single source in the appraisal of performance has ultimate legitimacy or warranty (Byham, 1996; Cochran Smith & Fries, 2002; Shephard, 2000). Moreover, to arrive at a balanced and multidimensional weighting of the many-faceted nature of professional expertise (Ericsson, 1996), a combined overview of several dimensions in appraisal is needed. Multiperspective assessment in mentored learning and in tutorial relations may have been undervalued in teacher education. What has been stressed is assessment that supports a single, conclusive, if not summary, rating (Cochran Smith & Fries, 2002; Ben Peretz, 2001). But receiving feedback from multiple perspectives, even to the extent that it entails descriptive, judgmental assessment information, can indeed foster the learning process of beginning professionals (Tillema & Smith, 2006; Loughran, 2007). Certainly, relations among mentors, supervisors, and student teachers should be conceived primarily as learning partnerships (Baxter Magolda, 2004; Edwards et al., 2002). Therefore, bringing in multiple perspectives from different sources to provide informative feedback (by peers, supervisors,

and teachers) can help the student teacher in various aspects of his or her performance.

Appraisal Processes With Multiple Raters

It is no small matter to organize such a concerted, fine-tuned arrangement of a multiperspective assessment (Gijbels, Watering, Dochy, & Van den Bossche, 2005; Lievens, 1998). First it is important to acknowledge which framing factors may cause divergence or variance in shared appraisals. It can be maintained that, when not explicated and shared, these framing factors may cause variance in orientation to the appraisal task among assessors and should therefore be scrutinized in a multirater appraisal. As a framework to review appraisal processes, the following framing factors can be identified (Falchikov, 2005; Smith & Tillema, 2003; Tigelaar et al., 2002; Topping, 1998; Zeichner & Wray, 2000):

- (a) the *purpose* of bringing together assessment information (the nature of the information to be collected), that is, why, for what purpose?
- (b) the *object* of evidenced assessment information (the practice teaching performance), that is, what is being appraised?
- (c) the way *evidence* is appraised (the type of information that will be regarded as relevant), that is, what counts as evidence?
- (d) the *focus* on further development (the support for learning that an assessor is willing to provide or the mentoring orientation involved), that is, is the information informative to the learner?
- (e) the *criteria* by which performance is appraised (i.e., the standards used to evaluate what has been accomplished), that is, what measures are gauged?
- (f) the *involvement* of different types of raters to appraise the performance, that is, who is being rated by whom?

Based on these framing factors, it becomes possible to establish what actual convergence or alignment is reached in a joint appraisal by different raters. A deliberate design of framing factors may avoid a situation in which functional feedback becomes distributed and dispersed, or even conflicting in nature, so that no learning consequences may be drawn. When agreement exists on the framing factors, alignment or congruence among raters can be achieved. Unanimity in the process and purpose of the appraisal or, otherwise, a deliberate and orchestrated variance may be striven for by having a balanced review with different evidence.

To gauge the practice of assessing learning to teach, we studied in the context of teacher education how student teaching is appraised by different raters and which framing

factors are used in appraising the performance of student teachers. This study focuses on the joint appraisal (in triads) of a shared, single practice teaching event to contextualize and focus on the different perceptions and experienced problems of the stakeholders.

Method

Study Design

To determine whether there was any alignment among the different raters, we performed an explorative study to gauge actual ratings of student teacher lesson performance. For this purpose, data were collected in triads (17 of them). Triad members rated a particular teaching performance in a lesson given by the student teacher. Triads were used to detect alignment in perceptions of concerned stakeholders on actual teaching, which might otherwise be lost when generically analyzing group data. The assessors ($n = 51$) in each triad were the mentor or practice teacher, the supervisor or visiting teacher educator from the teacher education institute, and the student teacher. All participants volunteered to take part in the study and were affiliated with a large teacher education institute in the Netherlands that had several branches (and practice teaching locations). The institute maintained a core teaching program for all student teachers in which practice lessons were integrated and evaluated against the same standards. Practice teachers were teachers in primary education affiliated with the program as mentors and were informed about course objectives and standards for practice teaching.

Procedure and Instruments

Triads consisted of a cooperating mentor, supervisor, and student teacher; they were formed on the basis of the practice teaching lesson schedule issued by the teacher education institute. Members of the triad were asked to pick one particular lesson recently given by the student teacher (within the current practice teaching period) and to provide further information on the appraisal of that practice teaching lesson through the questionnaires. The following information was requested from each triad member.

(a) *A questionnaire on lesson appraisal.* A questionnaire addressed each of the identified frame factors (purpose, object, method, focus, and criterion) relating to the arrangement of appraisal processes (see above). We inquired about the rater's own position with respect to the appraisal of the lesson. For each frame factor, several alternatives were given for execution of the rater's task. These were rated on extent of application (scale ranging from 1 to 100). For

Table 1
Frame Factors With Regard to Appraisal of Student Teachers' Practice Lessons

Degree of Application	Supervisors	Mentors	Students	Overall Mean	Congruent Triads ($n = 17$)
Purpose of assessment					
Determining progress in development	80	80	55*	72	4
Promoting learning	85	68*	92	82	2
Giving feedback on performance	85	72*	84	80	3
Determining actual competence level	90	68*	84	80	5
Prime object of appraisal					
Based on written lesson protocol	90	95	90	92	11
Based on all available information	90	92	92	92	14
Based on agreed criteria	95	92	68*	85	7
Based on planned targets	100	95	95	97	15
Based on students' needs, questions	85	75	85	82	9
Way of appraisal					
Face-to-face conversation (dyads)	75	80	75	77	8
Individual student self-assessment	55	60	75*	63	3
Independent supervisor rating (program based)	80	75	65	73	13
Applying a rating scale and giving comments	90	85	90	88	8
Using a fixed entry registration protocol; no comments given	75	90*	75	80	8
Focus in assessment is primarily . . .					
Student oriented	65	70	75	70	11
Assessor directed	75	80	90	82	6
Assessee-assessor agreement	75	80	65	73	12
Criterion in appraisal					
Using a fixed set of standards	85	55*	80	75	9
Depending on the mentoring/learning orientation	75	90	90	85	13
Depending on a personal style or approach to teaching	65	75	55	65	15

Note: Rating scale range = 1 to 100 on applicability; $n = 51$.

* $p < .05$, analysis of variance F test, using Scheffe comparison between groups.

example, for purposes of appraisal, several options were offered:

1. Determining progress in development
2. Promoting learning
3. Giving feedback on performance
4. Determining actual competence level

The alternatives given for each frame factor were derived from a previous study on assessment in teacher education that identified assessor perceptions in appraisal (Smith & Tillema, 2003). These were slightly adapted (rephrased) for this study (see Table 1 for the alternatives). In addition to this questionnaire, the assessors were asked to rate and indicate in greater depth (by providing a comment) the role they adhered to as an assessor (i.e., extent to which the assessor viewed herself or himself as assessor, reflector, guide, critical friend, or performance consultant). These roles were explained briefly.

(b) *Written appraisal review.* This was to be a narrative appraisal of the lesson (a review of half a page) specifying evaluations that were most applicable to decisions on the quality of the lesson. The content of each assessor's written evaluation was analyzed with respect to encountered problems during appraisal and criteria mentioned with respect to the quality of the appraisal process. Content analysis consisted of a propositional analysis of the narrative review (Hitchcock & Hughes, 1998) to arrive at a subject-predicate relationship, that is, student performance-judgmental evaluation or criterion statements (see Analysis). Furthermore, each assessor was asked to indicate in the narrative specific problems that were encountered when appraising the lesson.

(c) *A questionnaire on identified problems in assessing practice lessons.* Based on individually encountered problems mentioned in the written appraisal review, a questionnaire was developed on those problems and administered

a few weeks later (i.e., the second data collection occasion). Each triad member was asked to rate the impact (on a scale from 1 to 100) and prioritize the list of shared problems that were present in assessing the practice lessons.

(d) *A questionnaire on competencies appraised.* This questionnaire consisted of two parts and was administered together with the questionnaire on identified problems as part of the second data collection. It was meant to determine whether the five key competencies of lesson quality identified in the teacher education program actually played a role during appraisal. The teacher education program stressed particular competencies to be addressed in practice teaching:

- interpersonal competence: promoting cooperation among pupils and providing a positive climate;
- pedagogical competence: providing a safe learning environment that fostered self-determination in students (having two statements);
- curriculum competence: covering content within a powerful learning environment;
- management competence: maintaining an orderly, task-oriented atmosphere (having two statements); and
- reflective competence: showing pedagogical reasoning and understanding of task performance.

A rating of these competencies according to priority was applied to gauge whether a shared conception was present among assessors on the importance of the to-be-appraised student teaching competencies. Such a priority rating is of interest to determine the common ground or alignment in multiple assessors' appraisals with reference to the stated teacher education program goals. The assessors could give a priority score, ranging from 1 to 100, to each of the five competencies.

In addition to the competencies offered, triad members were encouraged to list their own competencies that they considered relevant to evaluate practice teaching lessons. Assessors could indicate them in the written appraisal review as relevant to judge the lesson performance.

Analysis

The analysis of data focused on finding commonalities and differences among the raters with respect to the problems encountered when appraising the lessons; these were organized according to several frame factors of assessment. Preferences regarding the criteria used were analyzed among the assessors as well as the consistency and alignment among actual appraisals of practice lessons by assessors who preferred different roles.

Frequencies on established categories (frame factors of appraisal, appraisal problems encountered, and criteria

used) were analyzed for congruence among triads, using analyses of variance to determine differences between groups (ANOVA F test for differences between multiple assessor groups and Kruskal Wallis H tests of differences between categories having ranked scores). As a measure of congruency for each category, scores obtained from all three participants in a triad falling within the same 25% frequency range were considered similar.

The content of the written reports were analyzed using a propositional analysis to identify the categories to be used in the questionnaire on identified problems that was administered. Following an iterative text analysis procedure (Bovair & Kieras, 1985), kernel sentences were obtained identifying subject-predicate relations, which were subsumed under topical labels that could be used as categories. For example, "I saw her hesitating when getting pupil reactions to the questions she poses" was coded as "getting pupil reactions—hesitation to act" and subsequently subsumed under "giving feedback to students." Agreement in coding of the content analysis was measured for 10 triad data sets with interrater reliability of $k = .89$.

To contextualize the empirical findings of this study, a vignette is presented (in the appendix) in which one of the teacher educator supervisors reflects on her experiences in the appraisal process. It may offer a perspective on our data. Assessors were invited to provide a reflective account of the appraisal that could illustrate concurrent thoughts or reflections on their role.

Results

The findings of this study deal with several sections of data with respect to the raters' appraisals: (a) evaluations regarding application of identified frame factors in the appraisal, (b) encountered problems in assessing the practice teaching lesson, (c) focus in appraisal on competencies divided into program key competencies and rater-defined key competencies as relevant for judging lesson quality, and (d) specification of applied assessor's role during appraisal.

Lesson Appraisal and Assessor Roles

Table 1 presents the overall ratings of all three triad members with respect to the factors identified as relevant to student teaching appraisal (purpose, object, instrument, focus, and criteria of appraisal). In addition, Table 1 indicates the level of congruence among the assessors for the particular practice lesson.

Table 2 presents findings with respect to the assessors' roles during the appraisal process of the particular lesson.

Table 2
Assessors' Role Perceptions in Appraising Student Teachers' Practice Lessons

Preferred Roles of Each Triad Member	Supervisors	Mentors	Students	Difference ^a $p < .05$
Assessor	+	++	++	
Reflector	+++	+	+++	$H = 7.82$
Guide	++	+++	++	
Critical friend	+++	+	++	$H = 7.01$
(Performance) consultant	++	+++	+++	

Note: Rating on a 5-point Likert scale; + indicates a full scale point.

a. Kruskal Wallis one-way analysis of variance H test.

The last column in Table 1 shows the number of triads that were congruent, that is, falling within same 25% range of scoring (range 1 to 100). Furthermore, Table 1 indicates (in bold) which assessor group diverged most from the others (using ANOVA Scheffe tests). Table 2 provides similar information on discrepancies about the assessor roles applied (based on Likert-type ratings). The findings shown in Table 1 illustrate that assessors differed most in the domain of purpose of appraisal (based on the number of significant differences and the number of poorly congruent triads). In particular, the mentors' opinions were more divergent than were the other two assessor types. The relatively low rankings in the list of purposes for mentors seems to indicate their main purpose in appraisal was not included among the alternatives, although they did agree with the supervisors (not the students) on having looked at the progress made by the student teacher. Supervisors and student teachers agreed that determining actual competence levels (in reference to program standards) and promoting learning are the prime purposes of appraisal. Strong agreement as well as congruence between all rater groups was found with respect to the objects of appraisal, most notably, with respect to using all available information and deploying planned targets. Considerable variance in viewpoints (although not significant), however, was found with regard to the way assessment was conducted or the appraisal instruments used.

Notable is the preference of mentors for a fixed-entry, checkbox type of instrument, in which a clear rating of performance is possible (a system that seems to have been disliked by both supervisors and student teachers). The instrument everyone preferred was a clear (inconspicuous) rating scale with sufficient opportunity for giving (judgmental) comments. Furthermore, there was great congruence on having independent supervisor ratings (although not with high preference). With respect to the applied focus of appraisal, we note that most adherence was found for an

assessor-directed assessment process, in which congruence exists in the triads when both assessor and assessed party (student teacher) arrive at the same opinion in the appraisal process, that is, striving for mutual agreement but initiated or directed by an assessor.

The findings from Table 2 intensify this by showing significant divergence in the triads about preferred roles: Those more closely connected to the teacher education program preferred a reflective role, acting as a critical friend, whereas mentors preferred a steering and performance-oriented advisory role. Despite the divergence in preferred interpretation of assessor roles by the stakeholders, Table 1 shows that they adhered to an assessor-initiated or an assessor-directed appraisal process. This finding is further supported by information found under criterion as a frame factor, which highlights an overall learning-oriented approach to the appraisal process, when it is viewed as an "assessment for learning."

Overall, considering the number of congruent triads, notable is the lack of full agreement among all three groups on frame factors relating to appraisal (with a mean of 8.17 congruent triads out of 17 in total). This implies a considerable variation among stakeholders in the direction they take in the appraisal processes.

Problems Encountered

To highlight possible differences in perceptions among stakeholders, the problems encountered during the appraisal process were identified. Table 3 presents these difficulties based on (a) priority given by assessors, (b) agreement found, and (c) congruence established in triads.

What is clear from the table (having top priority, great agreement, as well as congruence) is the jointly experienced lack of guidelines and clear procedures on how to work as assessors. In addition to this common ground in perception, each stakeholder has his or her own perspective on issues of further concern: Students are predominantly concerned about alignment in appraisal among stakeholders

Table 3
Priority List of Identified Problems in Assessment of a Practice Teaching Lesson

Problems Encountered	Supervisors	Mentors	Students	Overall Mean	Congruent Triads ($n = 17$)
Lack of guidelines and grading rules for assessors	100	95	90	95	14
Managing multiple perspectives in appraisal	90	75	75	80	4
Using different appraisal sources/information	90	85	55	77	7
Not having clear criteria in appraisal	86	80	75	80	13
Conducting a supervision conversation	82	85	83	83	8
Structure of supervision conversation	80	80	62	74	14
Maintaining supervision standards	80	85	62	76	5
Using observations of practice teachers	75	78	65	73	5
Students' influence on ratings	73	75	80	76	8
Giving directions for future learning	72	89	84	82	9
Giving feedback to students	63	82	88	78	11
Use of observation data in conversations	62	84	62	69	8
Alignment in ratings among assessors	50	60	80	63	5
Mean					8.54

Note: Figures in boldface indicate ANOVA significance testing at $p < .05$ using Scheffe comparisons among groups; congruency = all triad scores in same quartile.

Table 4
Agreement Among Raters on Competencies

Priority Given	Teaching Program Competence Domain	Supervisors	Mentors	Students	Kruskal Wallis Test at $p < .05$
1	Maintaining order in the classroom = management competence	++ ^a	++++	+++	$H = 6.46$
2	Clear presentation of lesson content = curriculum competence	++++	++	+++	$H = 7.61$
3	Well conducted introduction of lesson = curriculum competence	+++	+	++	
4	Adequate guidance and interaction with individual students = interpersonal competence	++	+++	+	
5	Showing an interest in student reactions = pedagogical competence	++	+++	+	
6	Creating a positive learning atmosphere in the class = pedagogical competence	++	++	++	
7	Being aware of one's position in the classroom = reflective competence	++	+	+	

Note: + indicates full scale point.

a. Mean rating on 5-point scale.

and the way they receive feedback, whereas mentors more readily focus on how to give directions (for learning) and maintaining standards. Supervisors, however, seem more concerned about maintaining multiple perspectives (or conflicting voices) in the appraisal but value the different sources of information that come into play. Both supervisors and mentors seem more aligned in their perceptions than do their students, given the number of significant differences found among the groups for each problem. With respect to congruence, a mixed picture emerges: Strong congruence was found on problems with respect to clarity of procedure (lack of guidelines, criteria, and structure in appraisal); however, there was diversity in perspectives on

clarity of purpose (allowing for multiple perspectives, using different observations, maintaining standards and alignments of ratings).

Focus on Competencies Appraised

This study also explored whether agreement existed among the stakeholders on key teaching competencies to be appraised in the practice teaching lesson. Table 4 gives mean ratings and priorities for the key competencies aimed for by the teacher education program. These are merged with the key competencies derived from the content analysis of the written reviews by stakeholders.

Table 5
Agreement on Criteria for Quality of Appraisal Process

Criterion	Supervisors	Mentors	Students	Kruskal Wallis Test at $p < .05$
Clarity of goals to be attained	++++	++	++	$H = 5.9$
Uniformity in grading and scoring rules	+++	++	+++	
Transparency of procedures and rating	+++	++	++++	$H = 6.5$
Recognizable and constructive appraisal conversation	++	++	+++	
Guidance for future activity	+++	++	++++	$H = 6.3$

Note: + indicates full scale point.

What seems to stand out in the appraisal focus is the student teacher's proficiency in presenting to the class and managing the process. These were more important than reflection and pedagogy. But among the triads, there was considerable variation in focus, or at least appreciation of importance, on the highest prioritized key competencies. Testing for differences among groups (Kruskal Wallis's one-way analysis of variance), resulted in significant results for maintaining classroom order and lesson presentation (respectively, $H = 6.46$ and 7.61). Apparently, supervisors differ from students and mentors on issues of classroom management but agree more with students about lesson presentation and guidance of students. But data in Table 4 do not reveal that there was hardly any overall congruence or shared focus on the competencies to be appraised among stakeholders within each triad (counting only a total of 3 congruent triads out of 17).

Criteria in Appraisal

With respect to agreement on criteria relevant for appraisal of student teaching, our findings show (Table 5) that mentors have a different perspective on standards of quality, with an overall lower rating on applicability of the five criteria presented (based on their mean ratings per category on a 5-point scale).

Supervisors ranked clarity of goals higher than did mentors and students, although students felt transparency in the appraisal process and the support and guidance it gives for future action were more significant. Mentors had no clear preference for giving guidance for future action as a criterion in the appraisal process, that is, as part of an assessment for (future) learning of the student teacher. These findings on criteria relevant for student teacher appraisal were corroborated by data obtained from the content analyses of the written reviews. The issues that were mentioned frequently in these reports, in exemplification of the Table 5 findings, show important points of agreement as well as concern about a mutual

appraisal of student teaching. Common criteria or considerations mentioned in all reports on the conduct of an assessor during appraisal were

- working with known criteria,
- using an accepted scoring format to record performance,
- using accepted and shared competencies for appraisal,
- having an opportunity for reflection on performance,
- exchanging information and comments in conversation meetings,
- acknowledging comments and suggestions made by students, and
- using both verbal and written forms of feedback.

Discussion

This study explored the assessment of practice teaching from the perspective of those involved in the process of appraisal (supervisors, mentors, and student teachers) to find agreements or congruence in the approaches and criteria used to appraise lessons given by student teachers. Our argument was that assessment of learning (to teach) is an important vehicle for organizing and supporting the student teacher to achieve competence in teaching. And the way assessment is delivered could very well influence what students learn from (appraising their) practice experiences. This study sought to compare multiperspective appraisals of a shared event by different raters to gauge whether and to what extent they look at the event from divergent or congruent perspectives, because it can be contended that joint (shared and multifaceted) viewpoints on process and criteria of appraisal will support an informative and balanced (e) valuation of the performance. The study findings can be concluded with respect to (a) agreements among stakeholders on identified frame factors in the appraisal, (b) encountered problems in assessing teaching practice lessons, and (c) alignment in the focus on competencies and criteria used in the appraisal process.

Frame Factors in Appraisal

Looking at the viewpoints with respect to arrangement of appraisals, all concerned agree on the following understanding of the appraisal process: (a) It is intended to promote learning (the primary purpose), although there is a difference about whether to determine actual levels or progress in development (students vs. mentors). We noted that the most disagreement was found about the purposes of appraisal. (b) There was high agreement about the object of appraisal; that is, it should be based on written protocols, have agreed-on planned targets, and to a lesser extent, deal with student needs and questions. (c) The most preferred instrument for appraising lessons was a rating scale that allows for adding comments and reflections as opposed to, for instance, student self-assessment (Boud & Falchikov, 1989). (d) Overall, the stakeholders have most trust (i.e., congruence) in an assessor who has a guided and judgmental approach to assessment, where agreement is sought among those involved in the outcomes of the appraisal (also Tillema & Smith, 2006).

In summary, these factors point to a preferred assessment for learning that uses a variety of information sources to provide further opportunity for reflection. But it should also be noted that congruence was not high in this respect. For that matter, mentors differed most from the other assessors in that they may have had a different purpose or one more directed toward a behavior- or action-oriented appraisal of student performance. This finding may suggest guarding against an overly simple equation of formative assessment with process rather than outcomes instead of keeping both perspectives in mind. The process of learning and the outcomes or effects of actions, although covered by different raters, were found in the formative appraisal process.

Problems Encountered

Distinct problems were encountered by all three stakeholders when executing the appraisal process. Most notably, a lack of guidelines and grading rules was recognized as problematic, together with a lack of criteria and structure in supervision meetings. The most congruence (i.e., agreement among all concerned) was found for clarity of process (which needs to be improved). Yet when stressing clarity in performance standards (i.e., competencies to be rated), divergences occurred. For students, what mattered most were the directions given for future performance, whereas mentors and supervisors were more concerned (although they differed on this) with appreciating multiple perspectives and maintaining supervision standards. Mentors were more lenient (or indifferent?) about process aspects and more rigid about

compliance with standards than were the teacher education supervisors. "Technical" aspects, such as applying observation data in supervision meetings, mattered less to all assessors.

In summary, most difficulties seem to have stemmed from the ambiguity of guidelines in the appraisal, both in the process (all point to a lack of clarity) and in content. This ambiguity has various causes, perhaps originating from different views on the purpose of appraisal.

Appraisal Process

A third important issue in this study is the criteria used, both the standards and the quality of appraisal processes. Again, lack of clarity of goals and transparency of procedures were rated high, especially among supervisors and students. The latter also stressed the need for guidance on future action. But the greatest discrepancy was found in the competencies weighted as indicative of teaching performance: Whereas mentors stressed orderly classroom control, supervisors focused more on adequate presentation; students seemed to adopt a middle option here, by recognizing both as being important. All participants regarded competence in individual guidance and interaction with pupils as least important. This outcome should be viewed against the goals of the teacher education program, which stresses attainment of competence in all domains to an equal degree. This finding points out a dissonance in the standards set and the appraisal focus in actual practice lessons.

This study was conducted to investigate assessors' agreement on a specific query: Do assessors and those assessed employ a concerted and aligned assessment in learning to teach? It was assumed that such an agreed-on and shared approach would support student teachers' acceptance of feedback and lead to following up on recommendations. Such an assessment for learning (Assessment Reform Group, 2006; Birenbaum, 1996, 2003; Black & Wiliam, 1998) stresses active involvement of the "learner" in obtaining relevant feedback about performance and supportive guidance by assessors on progress. This (in)formative assessment is said to improve students' motivation and self-esteem, because it adjusts to their need to be able to assess themselves and to understand and improve their learning (Falchikov, 2005; Sadler, 1998).

The study findings partly support such a learning-oriented view of assessment being present in the appraisal of practice teaching lessons. There seems to be agreement at least among supervisors and student teachers on having a learning orientation in appraisal. But from a slightly different perspective, mentors also stress the need to assess performance improvement. More than the other

groups, students ask for a supportive, guidance-oriented assessment rather than an appraisal based on strict standards. These findings indicate the presence of a multiple-assessor rating, but they also point to a need to integrate specific assessor perspectives. It is the combined viewpoints that must be considered in a full appraisal of how a student teacher performs. Two difficulties in adopting such a multiperspective view on assessment, however, are the lack of feasible tools and the lack of clear procedures to enable such a multifaceted appraisal. At present, neither exists in the actual appraisals of practice teaching, and this absence obstructs an integration of viewpoints: There are now both different perspectives *and* diverse or unclear guidelines. Our findings indicated a high variability in criteria among assessors. This is troublesome, because the various orientations play an ambiguous role in actual appraisals (Nijveldt, 2007; Uhlenbeck, Verloop, & Beijaard, 2002). Explication of these orientations, as was carried out in this study, can be a first step for providing a common frame of reference in appraisals. Fortunately, we did find a common ground in the criteria used among assessors, especially with respect to the process of appraisal. Although this allows for organizing multiple perspectives into an integrated appraisal system, clear guidelines still need to be established.

Implications

It can be seen from our findings that assessment is a process closely linked to assessors' intentions and the aspects the assessor considers relevant. In our view, the most important feature of assessments in learning to teach is that they allow students to control their own learning by helping students identify strengths and weaknesses in a continuous, nonthreatening way. In this respect, assessment is a bridge between learning needs and competence levels (Guskey & Bailey, 2001; McMillan, 2007;). Admittedly, student teachers and their "teachers" (both mentors and supervisors) still have great difficulty with this approach to assessment (perhaps because they need to comply with each other's intentions in appraisals, and there are no clear procedures for doing this). Also, the external environment (examinations, success or failure) makes appraisal more of a summary, externally controlled, objective-governed procedure (Falchikov, 1995; Wiggins, 1989). Actively collecting and deliberating appraisal information, however, lies at the heart of assessment for learning. Therefore, developing and using feasible assessment instruments for performance monitoring, such as multirater feedback, would constitute a valuable tool for redirecting learning (Smith & Tillema, 1998; Smith, 2006; Tomlinson & Saunders, 1998).

Caution is needed, however, in arguing that instruments or guidelines for appraisal would be sufficient in themselves to achieve assessment for learning. More important is the way assessment tools provide feedback: "Assessment is all about feedback" (Sadler, 1998; Shute, 2008). The feedback process is complex (Bennett & Ward, 1993; Butler & Winne, 1995). Functional feedback starts with detecting the necessary goal- and learner-related needs for performance improvement. Providing relevant feedback for the learner through assessment essentially means setting the goals for learning and reflection first (Gipps, 1994; Sadler, 1998) and then focusing on a careful diagnosis and monitoring of experiences that offer and scaffold competence-framed knowledge (Landy & Farr, 1983; Redman, 1994). The assessment process (and its tools) needs to offer opportunities for scrutiny in which the stakeholders set mutually agreed-on goals and direct their standards accordingly (Dochy, Segers, & Sluijsmans, 1999; Fisher & King, 1995; Falchikov, 2005). A collaborative or multiperspective feedback process may be more conducive to pursuing the many developmental issues that need to be addressed in student teacher learning. In combining different perspectives, delivery of feedback can complete the appraisal process by integrating collected practice experiences to provide recommendations for further development.

Appendix

Reflective Report of a Teacher Educator Reviewing her Experiences of Appraisal Processes

When I look at my own position I have both wonderful as well as troublesome experiences with appraising my students. Over the years I tried out several approaches to appraisal, and I cannot say I found a solution yet. Let me explain a bit.

First, there is the problem of appraisal tools. In the nineties we experimented with several assessment approaches in our program. I was expected to explain them to our collaborating practice teachers, but I was not sure myself whether they were an improvement or not. Actually, we used quite different approaches at the same time: performance grading, reflective accounts, learner reports, portfolios, and the like, and it was not always clear what purpose they served.

There was quite a debate going on about the level of detail and specificity required in appraising the student's lesson. Also, the criteria for grading practice teaching shifted quite a lot and were not clear to everyone. Often the question was how to react; i.e., as a mentor, as a guide, as an assessor. We had a period of uncertainty but also of meaningful discussion with student teachers.

(continued)

Appendix (continued)

We learned how to converse about performance and discuss about relevant evidence to show in practice teaching. A start was even made in setting an established criteria list on how to appraise practice lessons. This list used a grading system from 1 to 10. The discussions with students were enlightening and provided me with more insights about their learning needs as well as their reflective capability.

We also started to use peer assessment so their fellow students could observe practice lessons as well, which the mentors did not entirely like, I must hasten to add.

I guess the difficulty was that we had no way of establishing in an objective way to determine what students accomplished throughout the practice period. Our solution was to assess as a duo i.e., with a second assessor. After having observed a lesson, there was always a supervision meeting in which exchange and sharing of insights was the prime goal. But the mentor of the practice period often had already given a grading report that operated alongside and not in concert with my supervision meeting.

For me, at least, the supervision meeting should be a learning moment for the student, covering strong points to be remembered as well as developmental issues that need attention in the future. I continue to be focused on learning and development as an assessor.

All in all, I can look back on my experiences because I learned that

- having clear criteria is paramount, although I am not sure about what criteria should be prevalent;
- having standards established in order to provide for some norm or objective for appraisal is a problem that still need *[sic]* to be tackled;
- talking with students about their *[sic]* learning needs and linking them to the evidence they bring forward is the key to appraisal. Also, involving peers adds value to the process.

References

- Abernathy, T., Forsyth, A., & Mitchell, J. (2001). The bridge from student to teacher: What principals, teacher education faculty and students value in a teaching applicant. *Teacher Education Quarterly*, 28(4), 109-119.
- Assessment Reform Group. (1999). *Assessment for learning: Beyond the black box*. Cambridge, UK: University of Cambridge, School of Education.
- Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Retrieved August 26, 2006, from <http://www.assessment-reform-group.org.uk>
- Assessment Reform Group (2006). In J. Gardner, *Assessment and learning*, London: Sage.
- Atwater, L. E., & Brett, J. F. (2005). Antecedents and consequences of reactions to developmental 360 degree feedback. *Journal of Vocational Behavior*, 66, 532-548.
- Baum, T. (2002). Skills and training for the hospitality sector: A review of issues. *Journal of Vocational Education and Training*, 54(3), 343-363.
- Baxter Magolda, M. B. (2004). Evolution of a constructivist conceptualization of epistemological reflection. *Educational Psychologist*, 39(1), 31-43.
- Ben-Peretz, M. (2001). The impossible role of teacher educators in a changing world. *Teacher Education*, 52(1), 48-56.
- Bennett, R. E., & Ward, W. C. (1993). *Construction vs. choice in cognitive measurement: Issues in performance testing and portfolio assessment*. Hillsdale, NJ: Lawrence Erlbaum.
- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge* (pp. 3-29). Boston: Kluwer Academic.
- Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascalar (Eds.), *Optimising new modes of assessment, in search of qualities and standards* (pp. 13-36). Dordrecht, Netherlands: Kluwer Academic.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Boud, D., & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: A critical analysis of findings. *Higher Education*, 18(5), 529-549.
- Boshuizen, H. P. A., Bromme, R., & Gruber, H. (2004). *Professional Learning: Gaps and transitions on the way from novice to expert*. Dordrecht, Netherlands: Kluwer Academic.
- Bovair, S., & Kieras, D. E. (1985). A guide to propositional analysis for research on technical prose. In B. K. Britton & J. B. Black (Eds.), *Understanding expository text* (pp. 315-362). Hillsdale, NJ: Lawrence Erlbaum.
- Brown, S., & Glasner, A. (1999). *Assessment matters in higher education: Choosing and using diverse approaches*. Buckingham, UK: SRHE/Open University Press.
- Butler, D. L., & Winne, P. H. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of Educational Research*, 65(3), 245-281.
- Byham, W. C. (1996). *What is an assessment center: Method, application and technologies*. Los Angeles: Development Dimensions International.
- Cochran-Smith, M., & Fries, M. K. (2002). The discourse of reform in teacher education: Extending the dialogue. *Educational Researcher*, 31(6), 26-28.
- Dall'Alba, G., & Sandberg, J. (2006). Unveiling professional development: A critical review of stage models. *Review of Educational Research*, 76(3), 383-412.
- Darling-Hammond, L. (2000). Teacher quality and student achievement: A review of state policy evidence. *Education Policy Analysis Archives*, 8(1), 23-36.
- Darling-Hammond, L., & Bransford, J. (2004). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco: Jossey-Bass/Wiley.
- Day, C. (1999). *Professional development of teachers*. Buckingham, UK: Open University Press.
- Delandshere, G., & Arens, S. A. (2003). Examining the quality of the evidence in pre-service teacher portfolios. *Journal of Teacher Education*, 54(1), 57-73.
- Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A literature review. *Studies in Higher Education*, 24, 331-350.
- Edwards, A., & Collison, J. (1996). *Mentoring and developing practice in primary schools*. Buckingham, UK: Open University Press.

- Edwards, A., Gilroy, P., & Hartley, D. (2002). *Rethinking teacher education: Collaborative responses to uncertainty*. London: Routledge Falmer.
- Ericsson, K. A. (Ed.). (1996). *The road to excellence: The acquisition of expert performance in the arts and sciences, sports and games*. Mahwah, NJ: Lawrence Erlbaum.
- Falchikov, N. (2005). *Improving assessment through student involvement: Practical solutions for aiding learning in higher and further education*. London: Routledge Falmer.
- Feiman Nemser, S., & Remillard, J. (1996). Perspectives on learning to teach. In F. B. Murray (Ed.), *The teacher educator's handbook* (pp. 63-91). San Francisco: Jossey-Bass.
- Fisher, C.F. & King, R.M. (1995). *Authentic assessment, a guide to implementation*. Thousand Oaks: Corwin.
- Furlong, J., & Maynard, T. (1995). *Mentoring student teachers: The growth of professional knowledge*. London: Routledge.
- Guskey, T., & Bailey, J. (2001). *Developing grades and reporting systems for student learning*. Thousand Oaks, CA: Corwin.
- Gijbels, D., Watering, G. van de, Dochy, F., & Van den Bossche, P. (2005). The relationships between students' approaches to learning and the assessment of learning outcomes. *European Journal of Psychology of Education*, 20(4), 327-341.
- Gipps, C. (1994). *Beyond testing: Towards a theory of educational assessment*. London: Falmer.
- Grossman, P. (2006). Research on pedagogical approaches in teacher education. In M. Cochran-Smith & K. M. Zeichner (Eds.), *Studying teacher education: The report of the AERA Panel on Research and Teacher Education* (pp. 425-476). Mahwah, NJ: Lawrence Erlbaum.
- Haynes, A., & McDowell, L. (Eds.). (2007). *Balancing dilemmas in assessment and learning in contemporary education*. London: Routledge Research in Education.
- Heilbronn, R., Jones, C., Bubb, S., & Totterdell, M. (2002). School-based induction tutors: A challenging role. *School Leadership and Management*, 22(4), 34-45.
- Hitchcock, G., & Hughes, D. (1998). *Research and the teacher: A qualitative introduction to school-based research*. London: Routledge.
- Jellema, F. (2003). *Measuring training effects: The potential of 360-degree feedback*. Doctoral dissertation, Twente University, Enschede, Netherlands.
- Kirby, J. R., Knapper, C. K., Evans, C. J., Carty, A. E., & Gadula, C. (2003). Approaches to learning at work and workplace climate. *International Journal of Training and Development*, 7(1), 31-52.
- Kremer-Hayon, L., & Tillema, H. H. (1999). Self-regulated learning in the context of teacher education. *Teaching and Teacher Education*, 15(5), 507-522.
- Kwakman, K. (2003). Factors affecting teachers' participation in professional learning activities. *Teaching and Teachers Education*, 19, 149-170.
- Landy, F. J., & Farr, J. L. (1983). *The measurement of work performance*. New York: Academic Press.
- Lievens, F. (1998). Factors which improve the construct validity of assessment centers. *International Journal of Selection and Assessment*, 6(3), 141-152.
- Loughran, J. (2003, June). *Knowledge construction and learning to teach*. Keynote address delivered to the conference of the International Association of Teachers and Teaching, Leiden University, Leiden, Netherlands.
- Loughran, J. (2007). Researching teacher education practices: Responding to the challenges, demands, and expectations of self-study. *Journal of Teacher Education*, 58(1), 12-20.
- MacLelland, E. (2004). How convincing is alternative assessment for use in higher education? *Assessment and Evaluation in Higher Education*, 29, 311-321.
- Maurer, T. J., Mitchell, R. D., & Barbeite, F. G. (2002). Predicators of attitude towards 360-degree feedback system and involvement of post-feedback management development. *Journal of Occupational and Organizational Psychology*, 75, 87-102.
- McMillan, J. (2007). *Formative classroom assessment: Theory into practice*. New York: Teachers College Press.
- Nijveldt, M. (2007). *Validity in teacher assessment: An exploration of the judgment processes of assessors*. Doctoral dissertation, Leiden University, Leiden, Netherlands.
- Redman, W. (1994). *Portfolios for development: A guide for trainers and managers*. London: Kogan Page.
- Sadler, D. R. (1998). Formative assessment: Revisiting the territory. *Assessment in Education*, 5(1), 77-85.
- Shephard, L. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), 4-14.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1), 153-189.
- Smith, K. & Tillema, H.H. (1998). Evaluating portfolio use as a learning tool for professionals. *ScandiJournal of EducaResearch*, 41(2), 193-205.
- Smith, K. (2006). The function of modelling: Teacher educators as assessors, students as assesseees. In P. Frenkel & K. Smith (Eds.), *How to assess what? Functions of assessment in teacher education* (pp. 46-67). Tel Aviv, Israel: Tema, Mofet Institute.
- Smith, K., & Tillema, H. (2003). Clarifying different types of portfolio use. *Assessment and Evaluation in Higher Education*, 26(6), 625-648.
- Snyder, J., Lippincott, A., & Bower, D. (1998). The inherent tensions in the multiple uses of portfolios in teacher education. *Teacher Education Quarterly*, 25(1), 45-60.
- Thornow, W. W. (1993). Perception or reality: Is multi-perspective measurement a means or an end? *Human Resource Management*, 32, 221-230.
- Tillema, H. H. & Smith, K. (2007). Portfolio assessment: In search of criteria. *Teaching and Teacher Education*, 23(4), 442-456.
- Tomlinson, P. & Saunders, S. (1995). The current possibilities for competence profiling in teacher education. In Edwards, A. and Knight, P. (Eds.) *The Assessment of Competence in Higher Education*, London: Kogan Press.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 66, 249-276.
- Tigelaar, D. E., Dolmans, D., Wolfhagen, I., & van der Vleuten, C. (2002). The development and validation of a framework for teaching competencies in higher education. *Higher Education*, 48(2), 253-268.
- Uhlenbeck, A. M., Verloop, N., & Beijaard, D. (2002). Requirements for an assessment procedure for beginning teachers: Implications from recent theories on teaching and assessment. *Teachers College Record*, 104, 242-272.
- Waldman, D. A., & Atwater, L. A. (1998). *The power of 360-degree feedback: How to leverage performance evaluations for top productivity*. Houston, TX: Gulf.
- Wang, J., & Odell, S. J. (2002). Mentored learning to teach according to standards-based reform: A critical review. *Review of Educational Research*, 72(3), 481-546.
- Wiggins, G. (1998). *Educative assessment: Designing assessment to inform and improve student performance*. San Francisco: Jossey-Bass.
- Wilson, S. M., & Berne, J. (1999). Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. *Review of Research in Education*, 24, 173-209.
- Wilson, S. M., & Youngs, P. (2005). Research on accountability processes in teacher education. In M. Cochran-Smith & K. Zeichner (Eds.), *Studying teacher education: Report of the AERA Panel on*

- Research and Teacher Education* (pp. 591-645). Mahwah, NJ: Lawrence Erlbaum.
- Yinger, R. J. & Hendricks-Lee, M. S. (1998). Professional development standards as a new context for professional development in the US teachers and teaching: Theory and practice, *4*(2), 273 - 298.
- Zeichner, K., & Wray, S. (2000). The teaching portfolio in US teacher education programs: What we know and what we need to know. *Teaching and Teacher Education*, *17*, 613-621.
- Zuzowsky, R., & Libman, Z. (2002, August). *Standards of teaching performance and teacher tests: Where do they lead us?* Paper presented at the conference of ATEE, Warsaw, Poland.

Harm H. Tillema's main field of interest is professional learning in teaching as well as teacher education with a special interest in the role of assessment and feedback as a tool of professional learning. In his consultancy work in several teaching organizations he is involved in establishing powerful learning environments that make use of assessment.