Lyn M. Shulha
Queen's University

Understanding Novice Teachers' Thinking About Student Assessment

This study captures novice teachers' efforts to mark, grade, and report on the language arts achievement of a fictional grade 8 student named Chris. Although it was not mandated, commenting on Chris's work was seen by them as a professional responsibility. Although comments were numerous, relatively few were specific enough to provide guidance in how to improve writing. Grading was influenced by the prescribed grading policy, but also the perceived appropriateness of the assessment instruments and by a holistic impression participants developed about their particular Chris. Student assessment emerges as a complex process influenced by the teacher, the learner, the learning environment, and the principles of measurement. Findings from this study suggest that professional knowledge about assessment developed in isolation from the complex contexts of classrooms will probably remain unconnected and underused.

Cette étude décrit les démarches d'enseignants débutants qui ont corrigé, évalué et analysé les réalisations d'un élève fictif de 8e année nommé Chris qui suivait un cours des arts du langage. Bien qu'on ne n'ait pas demandé aux enseignants de commenter le travail de Chris, ils l'ont fait par sens de responsabilité professionnelle. Alors que les commentaires étaient très nombreux, relativement peu d'entre eux étaient suffisamment détaillés pour constituer des conseils quant à la rédaction. L'évaluation à été influencée par la politique officielle sur les notes, par la perception des enseignants quant à la justesse des outils de mesure et par une impression holistique qu'ont développé les participants au sujet de Chris. L'étude démontre que l'évaluation des élèves constitue un processus complexe qui est influencé par l'enseignant, l'apprenant, le contexte d'apprentissage et les principes docimologiques. Les résultats laissent croire que les connaissances professionnelles sur l'évaluation qui sont développées à l'écart des contextes complexes des salles de classe demeureront probablement sous-utilisées et sans liens avec la réalité.

Whether the purpose is to diagnose learning, provide feedback to students, make decisions about next steps, or report to parents and other stakeholders in education, the assessment of student growth and achievement is central to the ongoing activity of teacher practice.

Academic resources intended to help teachers with evaluation tasks often argue that without the specialized knowledge found in the conventions of educational measurement (frequently renamed *the principles of classroom assessment*) teachers lack the expertise necessary to assess students effectively. "Teachers, as a group, from preschool instructors to university professors, have little formal training in classroom assessment techniques. They are forced to function with partial knowledge, and feel most comfortable with procedures that can easily accommodate their busy days" (Airasian, 1991, p. 262). To be adequate for the task of assessing students, teachers are encouraged to learn

what measurement specialists know about reliable and valid measures and to structure their work such that this knowledge can be put into practice. In short they are urged to become the "responsible, well-prepared classroom assessment professional" (Stiggins, 1997, p. ix).

From a measurement perspective the quality of information provided by a student on an assessment task should be the best predictor of the mark assigned to that performance. Some aggregate of those marks should then account for the final grade. This study had the potential to evoke this purer type of assessment and grading practice, as scoring of Chris's work was not distorted by multiple exposures to the same task, personal involvement with the student, or by the crush of daily routines—conditions said to diminish the validity of regular classroom teachers' assessment practices (Airasian, 1991; Joint Advisory Committee, 1993; Linn & Gronlund, 1990).

The notion that measurement specialists are in the best position to shape classroom assessment practice has not gone unchallenged. "As measurement experts, as academics, we have to begin to see teachers as individuals who make choices among plausible courses of action, routinely and spontaneously, by among other things, balancing the competing forces operating on them" (Wilson, 1994, p. 21). Wilson argues that many classroom-based evaluation decisions reflect teachers' beliefs not only about what is feasible to assess but what is appropriate to assess.

If researchers are willing to acknowledge that it is legitimate for teachers to make assessment decisions based on many factors, only some of which are directly related to the principles of measurement, then analysis of teachers' assessment practices must investigate the role of the competing forces teachers face in their work life. These forces include but are not limited to (a) the expectations of the system; (b) the contexts of their schools and classrooms; (c) the needs, interests, and abilities of the students; and (d) the values and constructs teachers carry with them concerning teaching and learning. The principles of reliable and valid measurement may be applicable only to the extent that they work in consort with these mediating factors.

This article investigates the strength of this competing assumption about classroom assessment by drawing on data that were collected in a near-natural context. By analyzing the behavior and talk of novice teachers both during and after their review of a student's work, it was possible to learn more about the complexity of thinking that contributed to their assessment practices. Questions guiding this inquiry included: How did these preservice teachers conceptualize and carry out the tasks of assessment and grading? What meanings were ascribed to the numbers they used to summarize Chris's performances? What role did measurement principles play in shaping participants' thinking and decision making?

Using Student Portfolios

The study was introduced to 147 preservice teachers as an opportunity to learn about portfolios and how they might help to assess student growth and achievement. Participants were told that the assignments they were to mark had been administered to everyone in Chris's class by the regular classroom teacher and that results from these assignments would be used to report on

Chris's progress and achievement in language arts. The portfolio, assembled over 10 weeks, provided novice teachers with evidence not only of Chris's performance on a set of instruments, but other information that is a regular part of an elementary school classroom teacher's professional knowledge: (a) Chris's socioeconomic background; (b) scores on a recent administration of the Canadian Test of Basic Skills; (c) the grading policies of the school; (d) the structure of Chris's classroom; (e) the preferences, expectations, and instructional style of Chris's teacher; (f) Chris's self-description as provided by an unmarked "Meet Me" essay (apparently written by Chris during the first week of school); (g) Chris's voice when reading aloud (and therefore cues about Chris's gender); (h) the expectations and involvement of Chris's parents; (i) Chris's ability to complete writing folder assignments; (j) Chris's performance in group work; and (k) Chris's academic performance in other subjects compared with that of classmates.

The assessment instruments used to measure Chris's growth and achievement in language arts were identical for all participating novice teachers, but the set of responses marked by each participant varied depending on which profile of Chris they were assigned. One third of participants marked work from a Chris whose performances improved throughout the term. Another third marked work where quality remained consistent or steady. For the final third of the participant group Chris's performances steadily declined compared with the quality of initial performances and to those of Chris's classmates.

The preservice teachers knew that Chris was a composite student, not a real person, and that there would be no personal or academic consequences for Chris or themselves based on their performance during the study. The task for participants was simply to mark Chris's work as it was deposited into the portfolio on a weekly basis. There was no overt pressure to attend to any of the contextual information that was provided. After submitting a grade and report card for Chris, participants were asked to speak candidly through an anonymous open-ended survey about their experiences throughout the study.

Analyzing the Data

The investigative categories selected to organize the qualitative data were intended to be inclusive and functional. Bogdan and Biklen (1982) suggest that in any natural situation people (a) develop ways to think about others, (b) use particular processes and strategies to shape their thinking, and (c) invoke particular methods of decision-making and problem-solving. A comprehensive investigation of novice teachers' assessment experiences would need to uncover how participants thought about their student, what shaped these perceptions, and how this thinking led to decisions about Chris's achievement. The closing survey was designed to elicit this information and to answer questions related to the study as a whole.

An initial reading of responses revealed that three activities captured much of these novice teachers' attention: responding to Chris's writing, constructing grades, and making sense of assessment. Each of these behaviors became a theme for data analysis.

Comments related to each theme were coded and placed into categories. Before any further analysis began, the themes were tested for their inclusiveness and stability. Samples of participants' comments were independently coded by collaborators who used the themes developed by this researcher. For all but one of the initial themes consistency in coding was high. For the exception a more precise label was generated. This revision had the effect of reducing the ambiguity in collaborators' assignments. The remaining discrepancies among the coders were resolved by breaking apart participants' compound sentences into data phrases. These phrases were then coded separately. All data phrases were numbered for reference.

For the first two themes the coded data were analyzed using a variation of analytic induction (Krathwohl, 1993) producing first a description of the regularities in the comments and then an explanation of participant behavior that could be traced and grounded in the data. For the third theme a cluster analysis was performed and a content-analytic dendrogram display of the findings was produced (Miles & Huberman, 1984).

Findings, Analyses and Inferences

Responding to Chris's Writing

Of the 147 preservice teachers participating in the study, 136 chose to comment on Chris's writing assignments. It is important to remember that comments on Chris's work were not requested. There were no incentives to comment, nor were there any repercussions for any participant who simply complied with the instruction to assign a mark to each performance. Comments to Chris appeared consistently across the three profiles of growth (*improving, steady, falling behind*). An average of three discrete comments were made on each piece of writing Chris was required to place in the portfolio during the term. Writing samples ranged from one paragraph to one page of word-processed material.

Comments on writing samples varied in their purpose: (a) relationship building, (b) specific improvement, (c) communicating the marking criteria, (d) encouragement, and (e) general improvement. Figure 1 displays how these categories were used by participants with different profiles of Chris.

Encouragement. Teachers seek out many ways to encourage and reinforce learning in their students. Comments such as "Great work Chris!" (# 027) and "Keep up the good work" (# 093) were typical of feedback coded in this category. It is interesting to note, however, that encouraging comments appeared consistently whether Chris was improving or steady and were only slightly less prevalent when Chris was actually falling behind in performance.

Two possibilities for this pattern are worth considering. A closer look at these statements reveals a form of teacher approval. Approval and encouragement in this study may have more to do with perceived work habits, attitude, and effort than with the task of writing. It may also be that novice teachers saw comments such as "A great effort!" (# 089) as a way to enhance Chris's motivation and thus indirectly to improve performance.

Relationship building. It was not surprising to find that novice teachers used assessment tasks as an opportunity to provide encouragement. It was surprising, however, to discover comments where the purpose was to build or extend a working relationship with a fictitious student. These comments were as
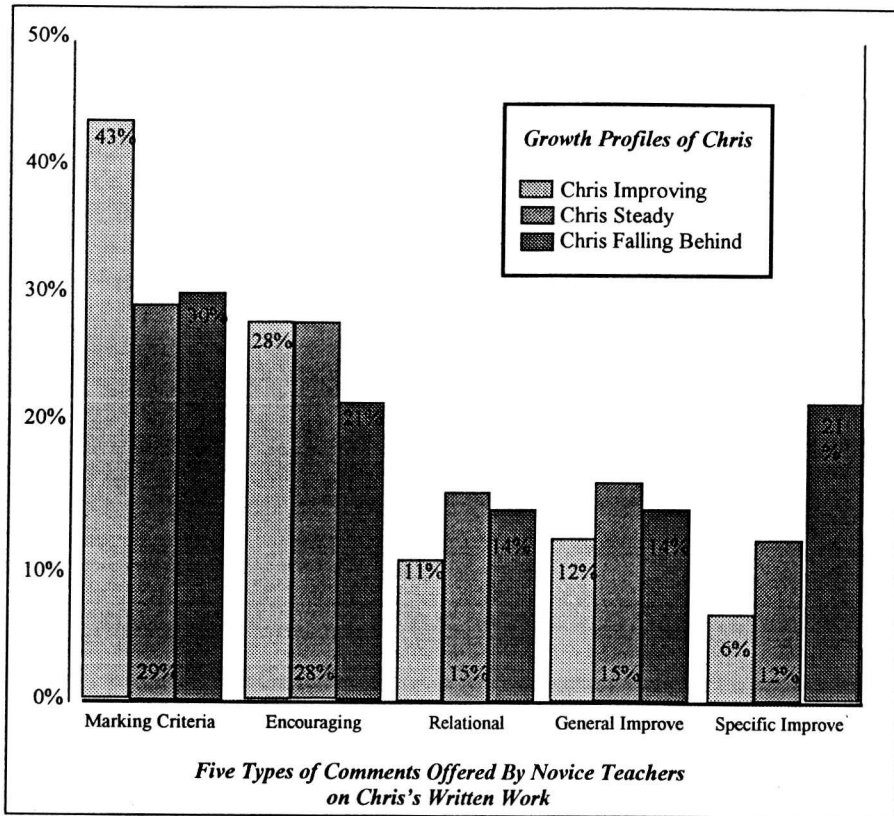
Figure 1. Type of written feedback offered for each growth profile of Chris.

frequent as those offering guidance for the general improvement of Chris's writing.

> Chris, I am very impressed with your thoughts on this poem. Do you like poetry? (# 119)

> Thanks for the piece of writing about the dance. Did you enjoy the dance? It sounds like it. I love dances but still feel nervous and funny about dancing when not too many other people are. I guess you don't always grow out of it. (# 006)

The emergence of this category was the first indication that the quality of the perceived relationship being developed with Chris might play a significant role in these teachers' thinking about assessment.

Specific improvement. The specific improvement category captured comments that gave detailed feedback on a particular quality or characteristic of the current work. Often these comments included suggestions on how to improve future work. In some cases the comments focused on mechanics.

> Remember to watch out for unnecessary repetition of words. Check the second sentence, could you change it somehow? In the third sentence you might be able to find a word slightly more appropriate than "done" as generally one doesn't "do a custom" but rather "maintains a custom" or "dislikes a custom." (# 006)

Other times participants were interested in extending Chris's thinking behind the writing.

> Chris, you have described the scene in the poem quite well and have shown insight into some of the issues which exist for the author. What do you think about the wall? Do you believe in keeping up traditions for their own sake? I'd love to hear more about how this poem speaks to your own life. (# 005)

It is interesting to note in Figure 1 that the improving Chris group did not receive the amount of specific constructive feedback that was provided to the falling behind group. One possible explanation for this difference is the sophistication of language arts skills needed by teachers intent on improving what could already be considered high quality writing. An alternative explanation is also worth consideration. Figure 1 displays evidence that improving students were continuously provided an abundance of feedback about the criteria used to award marks. If there is an interaction between these two findings, it may be that for novice teachers assessing quality work the primary purpose of written feedback was to help Chris get better marks rather than to help Chris become a better writer.

General improvement. Many comments written in the margins or at the bottom of pages gave general notice about how Chris could improve performance: "Watch your sentence structure and punctuation" (# 072) and "Linking sentences more closely will make this piece flow a little better" (# 117). The worry with these types of comments is that they assume Chris knows where to identify these errors in his or her own work and how to correct them without further instruction. This kind of feedback also has the potential of allowing teachers to feel sanguine about their efforts to promote improvement.

Marking criteria. Participants were free to establish their own criteria when judging the quality of Chris's responses to the assignments and final exam. Figure 2 illustrates that the element of writing most valued by participants was the quality of the content. Content comments referred to the richness of description, the thoroughness of observations, the clarity of ideas, and the development of theme. The mechanics of writing such as grammar, spelling, sentence structure, and punctuation were attended to less than content but as closely as personal style: creativity, insight, writing style, and the degree of personal meaning communicated through the writing. The structure of the assessment tasks did not allow for much room for inventive response formats. This may account for the small emphasis these participants placed on presentation as a criterion for earning marks.

The novice teachers marking a Chris who was falling behind tended to make more comments on the mechanics of writing (see Figure 3). This tendency could mean at least two things: (a) participants felt that a solid mechanical foundation was a necessary prerequisite to support ideas and creativity, or (b) errors in the mechanics of writing were the easiest to detect and the simplest to use when justifying the mark. The data did not allow for further investigation of either of these possibilities.

From Figures 2 and 3 it appears that the comments of these novice teachers served many purposes. Certainly feedback was intended to provide Chris with some explanation of what the assigned mark meant. Almost half the comments
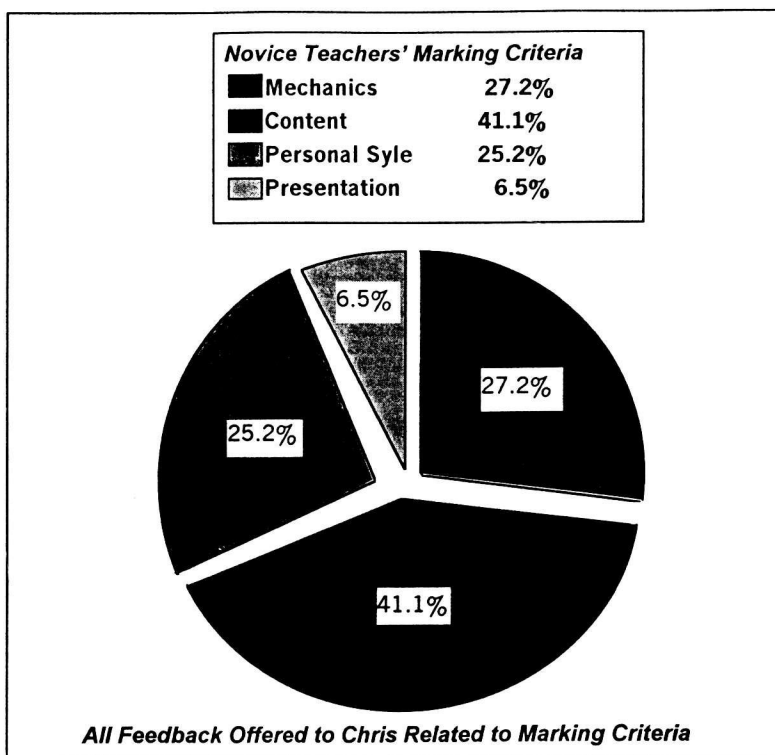
Novice Teachers' Marking Criteria	
■ Mechanics	27.2%
■ Content	41.1%
■ Personal Syle	25.2%
▨ Presentation	6.5%

6.5%

27.2%

25.2%

41.1%

All Feedback Offered to Chris Related to Marking Criteria

Figure 2. Explicit marking criteria used to discuss Chris's writing.

related to the quality of Chris's ideas and ability to develop a theme that could flow throughout the passage. It is also noteworthy that although teachers are often accused of not attending to issues of grammar and spelling, these participants addressed writing mechanics both in their criteria for marking and in their suggestions for improvement.

Constructing Grades

Participants were instructed at the beginning of the simulation to note the school policy limiting the influence of exam results to no more than 35% of a students' final grade. In Chris's class there were seven major term assignments worth a total of 130 marks and a final exam also worth 130 marks. Of the 145 preservice teachers who assigned a final grade, 82 responded to our request for information about how the grade was derived. With only one exception these participants saw grading as a complex task—more than simply adding up scores on tests instruments, calculating an appropriate weight, and then converting this information into a percentage and letter grade.

For the small number of participants (20) who claimed to determine Chris's grade by making such a calculation, few were confident that this method actually served the purpose of grading well. The exception was this participant: "Arriving at the grade was simple. I weighted the exam as 35% (the school's policy) and then divided the remaining 65% between the other assignments" (# 087). More often there was some reluctance in making grading a straightfor-
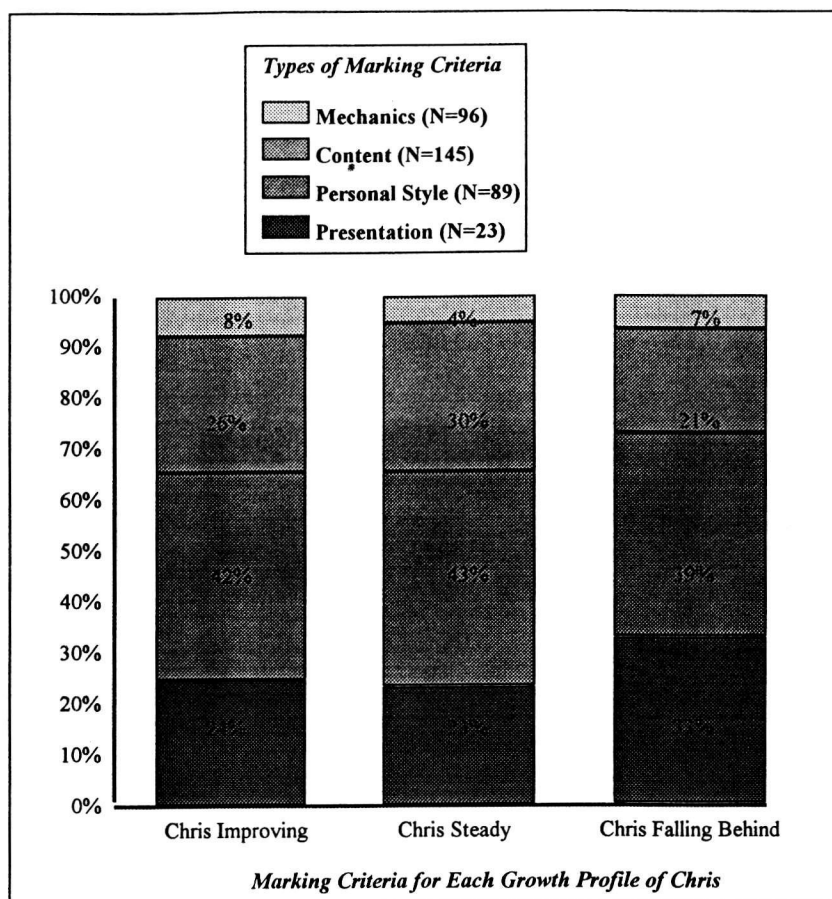
Figure 3. Comparing marking criteria for different growth profiles of Chris.

ward numerical exercise. " I hate to admit it, but the mark really came from the grades earned on all the assignments. This [grade] does not show development. This [grade] shows how Chris averaged out" (# 131).

It was apparent from the remainder of participants' comments that scores on the assessment instruments and the exam provided more of a baseline or reference point from which deliberation about Chris's achievement could begin. For example,

> I added up all the marks available and found a percentage for Chris. I did not include the one poetry assignment that Chris did poorly on. (# 070)

> Aside from the actual material I could apply a grade to, the library resource list was very useful in helping me arrive at grades and comments for Chris.... Near the end of term, his comprehension skills improved and he was now using the resources as much as other students. (# 034)

We expected to hear how participants struggled to combine the component scores of term work and the exam. This was a challenging task given the total marks assigned these two types of tasks were the same, but according to the

school policy the exam was to be given less weight. The data indicate, however, that participants were more interested in assigning unique meanings (weights) to particular assessment tasks and tended to be cautious about folding exam results into Chris's final grade. The only exception to this trend was one participant who appeared to have some strong beliefs about exams in general. "The high mark on this exam greatly influenced me. Great marks on an exam are an accomplishment" (# 047).

For a quarter of the participants the writing assignments were deemed to generate the most trustworthy evidence for the task of grading Chris in language arts. "Chris's writing [is represented] the most because that's what I think it is that's most important" (# 060). Some chose to feature other instruments from the set. "The greatest influence [on the final grade] was Chris's responses to the many comprehension questions. He displayed qualities of being logical" (# 066).

Still others went well beyond the marked assignments for assistance in calculating grades. "I looked to the circulation loans and the writing folder entries to see the effort Chris was putting in" (# 028). "I took everything into account in comments and grades: his book borrowing, his writing pieces, his final exam and exercises based on reading comprehension. I broke it all down equally to get a percentage" (# 53). This last example is revealing because participants were directed to use exercises such as the "Meet Me" essay and the reading miscue analysis for diagnostic purposes only.

As expected Chris's perceived growth also had an influence on final grades. One participant tracking a Chris who was falling behind wrote, "The greatest influence on me was Chris's progress. I was disappointed in watching her grades fall.... To me this suggests something was wrong" (# 054). In contrast, a participant with an improving Chris told us, "The most influential [factor] ... was the drastic difference between the first assignment and the last one. Her improvement was outstanding.... It seemed actually rewarding when Chris improved her achievement level" (# 114). For these two participants, and several others, the final grade was some measure of change.

Finally, the data uncovered a small group of participants who became so frustrated with the parameters of the portfolio and their grading responsibilities that they created their own rules for the study. These participants lamented:

> I did not feel that the material given was suitable to give Chris a mark. I gave him one based on the fact he handed in all his assignments and they were completed to my satisfaction (# 130).

> I assessed in a vacuum basically.... My grade was based on an assessment of general accomplishments but I don't feel I could successfully argue the [grade] I gave (# 134).

That novice teachers admitted to establishing grading criteria based on what felt appropriate for them confirms the presence of within-participant variability in the task of grading Chris.

An initial analysis of participants' decision-making might lead a reader to judge these novice teachers as simply lacking foundational principles in (a) reliability and validity, (b) formative and summative evaluation, and (c) report-

ing for growth versus achievement. Yet this "feeling factor" has been acknowledged to be an influence in real classrooms, even when the creation and implementation of the assessment plan has been under the control of experienced teachers (Bachor & Anderson, 1994). Data from this study suggest that these novices were going well beyond the information that Chris provided when thinking about grading. Their decision-making took into account some combination of (a) evidence of growth, (b) perceived evidence of effort, (c) judgment about the appropriateness of the assessment instruments to gather adequate data, (d) a holistic impression of the student based on multiple sources of information, and (e) a judgment about the utility of the school grading policy.

Making Sense of Assessment

Of the 82 participants who volunteered to talk more extensively about their assessment experiences, 11 responded with nothing but positive comments. " I really enjoyed this assignment. It was really interesting to see a student's work and try and evaluate it" (# 95). For the others, however, there were abundant questions and comments to be made about the assessment process in which they had been key players. They talked about what was good about portfolios, what was difficult about this particular task, what information was essential in assessment but missing in this case, what they brought to the process, and what they felt teachers needed in order to do assessment properly.

Figure 4 represents a cluster analysis and content-analytic dendrogram display of these findings. The analysis involved first grouping similar comments and labeling them as a primary cluster. There had to be at least two comments with a similar theme to create a primary-level cluster, and a participant's comments could contribute only once to each of the 27 primary clusters (i.e., a single respondent who made four statements about the significance of Chris's classroom behaviors contributed only once to the *work habits* cluster). A second and third level of clustering occurred when subsets of the data appeared to have a relationship. The result was a descriptive depiction of novice teachers' perceptions of optimal classroom assessment. The paths in the dendrogram are in degrees of boldface according to their importance (i.e., the number of participants who contributed to the construction of that path).

Through their comments the novice teachers in this study declared what they were searching for, as well as what they valued, when making assessment and grading decisions. For them the teacher, the learner, the learning environment, and the assessment activities all had a contribution to make if there was to be a comprehensive understanding of Chris's growth and achievement.

The teacher. In working through the raw data, one theme became immediately apparent. Participants felt limited in their ability to assess, primarily due to their lack of any in-depth knowledge of Chris. Background data, recorded classroom performances, a degree of parental involvement, and even a tape of Chris talking—all provided by the design—did not begin to satisfy participants' desire to know Chris as a person.

> The biggest thing for me, which is no surprise I'm sure, is that I wish I could've known Chris as a person. Although the "Meet Me" essay helped a bit, as did the creative writing pieces, I felt I was mostly in the dark about a big dimension of
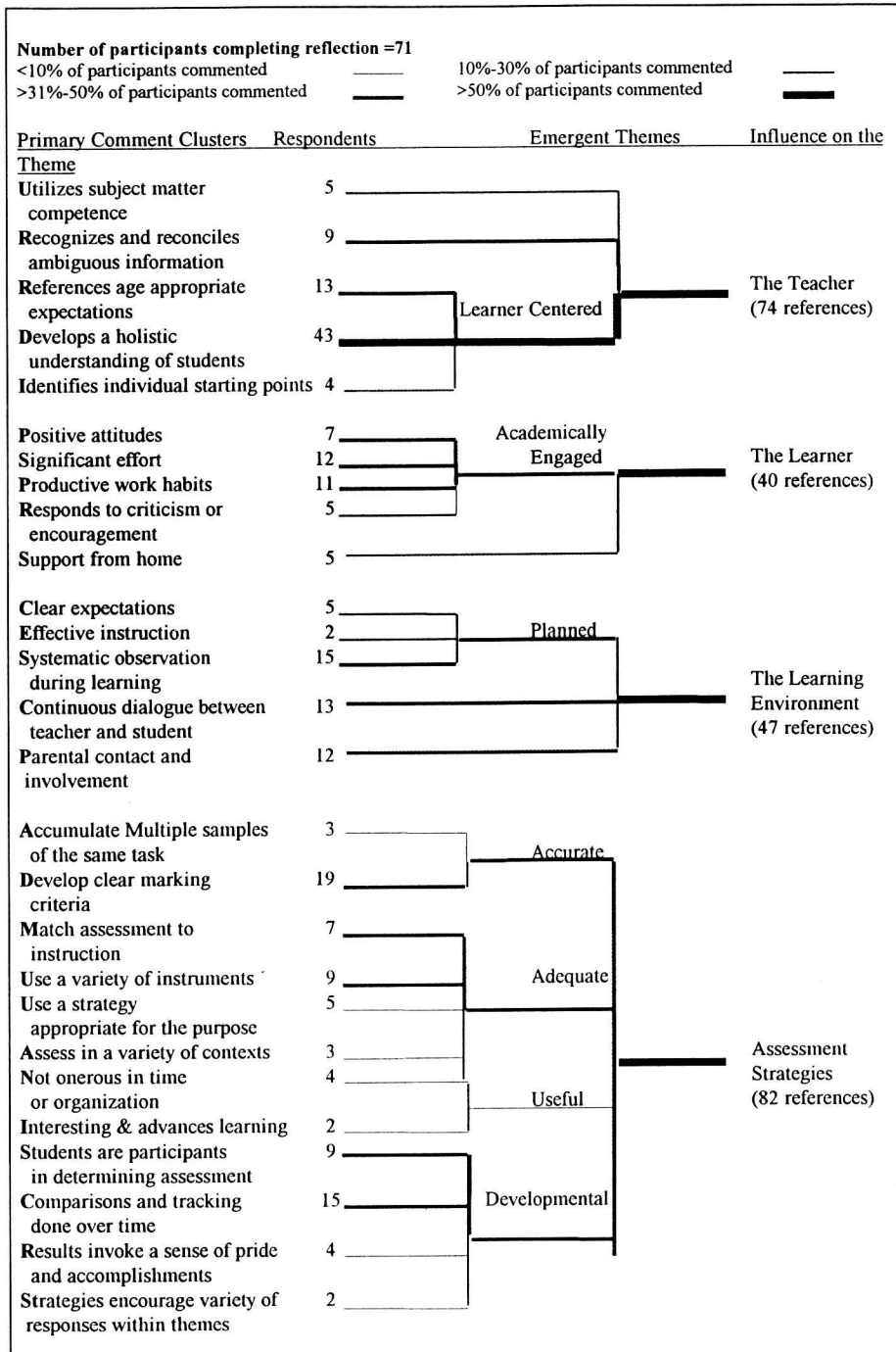
Number of participants completing reflection =71
<10% of participants commented   ———   10%-30% of participants commented   ———
>31%-50% of participants commented   ━━━   >50% of participants commented   ━━━

Primary Comment Clusters	Respondents	Emergent Themes	Influence on the Theme
Utilizes subject matter competence	5		
Recognizes and reconciles ambiguous information	9		
References age appropriate expectations	13	Learner Centered	The Teacher (74 references)
Develops a holistic understanding of students	43		
Identifies individual starting points	4		
Positive attitudes	7		
Significant effort	12	Academically Engaged	The Learner (40 references)
Productive work habits	11		
Responds to criticism or encouragement	5		
Support from home	5		
Clear expectations	5		
Effective instruction	2	Planned	
Systematic observation during learning	15		The Learning Environment (47 references)
Continuous dialogue between teacher and student	13		
Parental contact and involvement	12		
Accumulate Multiple samples of the same task	3	Accurate	
Develop clear marking criteria	19		
Match assessment to instruction	7		
Use a variety of instruments	9	Adequate	
Use a strategy appropriate for the purpose	5		
Assess in a variety of contexts	3		Assessment Strategies (82 references)
Not onerous in time or organization	4	Useful	
Interesting & advances learning	2		
Students are participants in determining assessment	9		
Comparisons and tracking done over time	15	Developmental	
Results invoke a sense of pride and accomplishments	4		
Strategies encourage variety of responses within themes	2		

Figure 4. A dendrogram of novice teachers' perceptions of optimal classroom assessment.

"Chris the student"—I realize as an assessor that I can't judge work in a vacuum very well. (# 005)

This assignment made marking a very clinical endeavour. It removed the personal element from teaching entirely and made me realize that it's definitely this (the human) aspect of teaching that interests me. (# 087)

Respondents struggled to establish reasonable expectations for Chris. Part of this struggle they attributed to their inexperience in teaching and with Chris's age group generally.

One thing that did influence my grading was my practicum placement. I was in a grade 7/8 and therefore witnessed what was standard for this level. This caused me to reevaluate Chris and to realize that he was a pretty steady kid who was slowly improving, although that improvement was slightly inconsistent. (# 001)

After having the opportunity in my teaching round to see other students' level of writing, it allowed me to evaluate with more competence and confidence. (# 033)

Participants also admitted that at times it was their own expertise in language arts that shaped the judgments they made about Chris. This admission supports the suggestion made above that there may have been a ceiling effect on the amount of specific constructive feedback that was available to an improving Chris. As well, participants expressed frustration in facing unexpected performances from Chris and not knowing how to respond.

I had difficulty reconciling the Chris I perceived from the Meet Me essay with the work I received from Chris later on. (# 034)

I found it interesting that when Chris's results changed from his usual achievement, I was reluctant to change his mark. (# 052)

Subject matter competence, the ability to attend to and resolve ambiguous information, experience, and, most important, a disposition to develop a holistic understanding of students best describe the contributions these participants thought teachers made to the assessment process.

The learners. Some participants acknowledged that students' growth and achievement was probably a function of their academic starting point (aptitude and prior knowledge) and their home profile (the degree of support and involvement from significant adults). But it was Chris's individual behaviors that most participants felt could have the most influence in the assessment process.

I found [assessment] frustrating ... [In] a final grade I would also include a mark for effort and participation. I arrived at this by a guess since I didn't even know Chris. (# 135)

This particular task was very difficult because normally the teacher would be interacting with the student on a daily basis and therefore know more about his or her work habits. (# 125)

Many participants were looking for some indication that Chris was engaged in the learning process. There appears to be a need to monitor, support, and report on this willingness to learn as this is displayed through attitude, work habits, effort, and responses to teacher comments. Many participants felt that

these learner characteristics should contribute to summary notations of academic achievement. For others information about Chris's orientation to learning would have helped with judgments about the appropriateness of the assessment activities included in the study.

The learning environment. According to these novice teachers, for assessment to be meaningful the learning environment must be characterized by clear goals (in some cases learner-determined), effective instruction, and a partnership among all the people who have a stake in the learning.

> I realized that a lot goes into the whole practice of evaluating. In fact, student-teacher, teacher-parent and teacher-student [links] are extremely important in maintaining a good connection. (# 033)

The optimal learning environment also needs to support opportunities for observation and dialogue between teacher and student.

> I feel that when you know the person as an individual and get to hear and see all of their work in the classroom you may arrive at certain insights that you may not have through strictly paperwork. (# 041)

> Teaching is about establishing a relationship with your students and learning together. (# 40)

> I found it difficult to mark material without being able to communicate with Chris face to face. (# 06)

The quality of assessment strategies. The final issues that were targeted by these novice teachers embodied how well the tasks designed to gather information about growth and achievement actually did the job. The absence of imposed marking criteria for the seven classroom assessment tasks and the exam gave participants much to think about.

> It was tough assessing when I have not set the criteria for marks before asking the student to do the work. I believe students need to know what is being looked at and being assessed in their work. (# 055)

> I don't feel the assessment was valid because I did not know what the exact criteria for assessment was for various assignments. I feel I may have marked Chris's work differently than the other Chrises were marked. (# 129)

Respondents were also conscious of the conditions necessary to have confidence in their judgments.

> There are many different forms of assessment for Language Arts: writing assignments like responses to stories, create your own stories, writing tasks-all of these should make up a final grade. (# 026)

> Research assignments ... would allow you to assess their logic in thought patterns, sentence structures spelling etc. all through writing. (# 051)

> I found the task of assessing Chris a good reminder as to the variety of assessment means that are necessary for accurate assessment. (# 052)

Even though only one participant made reference to the constructs of reliability and validity, intuitively they all seemed to know the importance of both.

Respondents commented a great deal about the advantage of portfolios, especially the opportunities to track students and make some comparative

judgments about performance over time. As well, there was speculation about how portfolios might engage students more in decision-making about their work. It is possible that these comments were artifacts of the study's design. Had we not referred to our work together as the portfolio study, participants might not have focused as intently on this assessment strategy.

Discussion

The feedback offered by participants in this study allowed for an in-depth look at how the complex tasks of responding to student work, constructing grades, and making sense of assessment were tackled by novice teachers. These findings lead to speculation about assessment and grading as these occur in the more intricate contexts of real classrooms.

Responding to Student Work

It is important to note that of all the comments analyzed on all writing samples across all profiles of Chris (N=1,199), only 13% provided Chris with feedback concrete enough to improve the current work or to extend that learning into future performances. This finding could support an argument that the bulk of comments were not useful in advancing Chris's proficiency and, especially as Chris was not a real person, a great waste of time.

Yet making comments on students' work was seen as a professional responsibility. Comments were multipurposed. They were intended (a) to clarify performance expectations, (b) to build and sustain a positive working relationship with the student, (c) to offer constructive feedback, and (d) to let Chris know the extent to which performance expectations had been met. In a classroom of 20-35 students the question of how to accomplish these goals while making the most of a teacher's time and energy is critical.

Brookhart (1999) identifies portfolios, conferences, exhibits, and rubrics as methods of assessment that can engage teachers and students in the joint construction of meaning around both the expectations and the evidence for learning. Yet these newer practices do not automatically yield improvement in assessment and grading. Deliberations on their purpose, their ability to promote instructional intentions, their contribution to judgments about growth and achievement, and their fit with a teacher's orientation to teaching and learning are necessary prerequisites to their valid use.

As Bookhart reminds us, "Teachers who cannot write well, or at least recognize good writing when they see it, will not be able to assess pupil's writing in portfolios" (p. 11). Similarly, teachers who find it difficult to involve students in the creation of expectations and in discussions about performance will not value the opportunities for dialogue that are inherent in these methods. It is possible, however, that inclusion of these newer types of assessment practices into a general assessment plan could satisfy many of the goals novice teachers had for responding to Chris with much less ambiguity and cost in teacher time. They could also help to align the expectations, instruction, observations, and dialogue that were identified as essential if the learning environment was to support optimal assessment practices (see Figure 4).

Although nothing in this study suggests that teachers would want to abandon commenting on student work, it may be prudent for teachers to consider how their comments can affect student performance. This study suggests that

there is some doubt about the power of general improvement comments in bettering performance. Once written, these comments represent a set of teacher expectations about the quality of the students' future work. The learner who may not be skilled enough to apply general comments to the specific task is at risk of appearing reticent about the learning process if performance does not improve. Teachers who see repetitive errors that they have already corrected through general comments may misinterpret this behavior, analyzing it as a learner's rather than a teacher's problem. Regardless, if a teacher is pressed for time it appears that one or two pieces of specific feedback designed to extend the learning in the current product, and described in a way that is transferable to the future, is the best investment of teacher time.

Grading

What is to be done with the feeling factor that was such a large part of the novice teachers' grading of Chris? Is it necessary to train teachers to exclude it from their practice? A closer look at the grading practices constructed in this study reveals an overwhelming desire on the part of novice teachers to be *fair* to Chris. Whittington (1999) argues that fairness can be equated with valid and reliable assessment. What is significant about Whittington's position is her insistence that in teacher education these constructs be examined not merely from a psychometric perspective, but through the decision-making context that classroom teachers confront on a daily basis. "The teaching of validity and reliability involves two elements: (a) exploration of the domain of assessment as it applies to teaching, and (b) insightful examination of each preservice or in-service teacher's way of reasoning and valuing" (p. 18). It might be wise to add Wilson's (1994) notion of analyzing the competing forces imposed by schools to this list as well.

Although presentations and textbooks can facilitate the first of these elements, the second and third activities as well as the integration of all of these elements require opportunities for reflection, dialogue, and analysis of practice. The more schools approach professional learning as active engagement in professional problem-solving, the better teachers will be able to examine their practices in context. The goal would be to have intuition anchored in sound principles, imperfect knowledge augmented by the precepts of the disciplines or the profession, and arbitrariness exposed and diminished. Rather than promoting assessment practices that negate the effects of teachers on student achievement, this approach would help teachers capitalize on their personal or professional trademarks while helping them to generate trustworthy and useful information about student growth and achievement.

Making Sense of Assessment

The participants in this study demonstrated through their comments how difficult it is for teachers to make assessment decisions in isolation from the students and the contexts where learning takes place. They have depicted assessment as an activity that by definition is shaped by the teacher, the learner, the learning environment, and the quality of the assessment strategies selected to generate information. This finding supports the urging of researchers like Brookhart (1999) who argues that "classroom assessment must be taught to aspiring teachers in relation to both instruction and classroom management,

not simply as a decontextualized application of measurement principles" (p. 13). Findings from this study suggest that the context for understanding assessment could be extended even further and include the fields of curriculum planning and learner development (cognitive, social, and emotional).

Conclusions

What are the next steps in learning about teachers' classroom assessment practices? In this study much of the learning about these novice teachers' values, beliefs, and practices came from direct analysis of their judgments and from indirect questioning designed to tap the reasoning that guided their judgments. Efforts must continue to generate insights into teachers' implicit knowledge. Reliance on literal accounts of intentions and practices denies the possibility that teachers will display more about their intentions and express more through their actions than it is possible for them to articulate.

The quality of classroom assessment, at least as it is conceptualized by this group of novice teachers, is multidimensional, yet largely in the control of teachers. The challenge is to develop a deeper understanding of the conditions that help or hinder attention to the various dimensions. Work in this direction would help to better situate assessment in the broader context of teaching, learning, and schooling.

As researchers with an interest in improving classroom assessment practices, we must also be willing to assert our energies in the contexts where assessment is played out. The goal would not be to eradicate current teacher practices or to create a new breed of classroom assessment professionals, but to support teachers in examining how expertise in assessment emerges and finds its rightful place in a multifaceted role.

References

Airasian, P.W. (1991). *Classroom assessment*. New York: McGraw Hill.

Bachor, D.G., & Anderson, J.O. (1994). Elementary teachers' assessment practices as observed in the province of British Columbia. *Assessment in Education, 1*(1), 65-95.

Bogdan, R.C., & Biklen, S.K. (1982). *Qualitative research for education: An introduction to theory and methods*. Boston, MA: Allyn and Bacon.

Brookhart, S.M. (1999). Teaching about communicating assessment results and grading. *Educational Measurement: Issues and Practice 18*(1), 5-13.

Joint Advisory Committee. (1993). *Principles for fair student assessment practices for education in Canada*. Edmonton, AB: Centre for Research in Applied Measurement and Evaluation.

Krathwohl, D.R. (1993). *Methods of educational and social science research: An integrated approach*. Toronto, ON: Copp Clark Pitman.

Linn, R.L., & Gronlund, N.E. (1990). *Measurement and assessment in teaching* (7th ed.). Englewood Cliffs, NJ: Merrill.

Miles, M.B., & Huberman, A.M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Beverly Hills, CA: Sage.

Stiggins, R.J. (1997). *Student-centered classroom assessment* (2nd ed.). Upper Saddle River, NJ: Merrill.

Whittingdon, D. (1999). Making room for values and fairness: Teaching reliability and validity in the classroom context. *Educational Measurement: Issues and Practice 18*(1), 14-22.

Wilson, R.J. (1994). *Back to basics: A revisionist model of classroom based assessment*. Paper presented at the annual meeting of the Canadian Society for Studies in Education, Calgary.