

Dan G. Bachor  
University of Victoria

and

Markus R. Baer  
British Columbia Ministry of Education

## An Examination of Preservice Teachers' Simulated Classroom Assessment Practices

*The comments made by 127 preservice teachers (PTs) in the Faculty of Education, University of Victoria as they compiled portfolios on three hypothetical grade 5 children are examined. The PTs were asked to record their comments in the form of a journal throughout the term. At the end of the term the PTs' comments were collected, transcribed, and the resultant data analyzed using Atlas/ti. The data were analyzed to examine the types of decisions the PTs made about the hypothetical children. Two main patterns of decisions were taken by the PTs. Most seemed to follow a fairly logical set of procedures, formulating criteria to evaluate the assignments and then applying them. A few appeared to make judgments that may have been unsound, however, commenting on the children's quality of life or designating them as having special educational needs when the evidence presented did not support such conclusions.*

*L'article traite des commentaires émis par 127 stagiaires de la faculté d'éducation de la University of Victoria qui rédigeaient un portfolio à partir d'enfants hypothétiques qui seraient en 5<sup>e</sup> année. On avait demandé aux stagiaires de noter leurs commentaires dans un journal pendant tout le stage, après quoi ceux-ci ont été recueillis et transcrits pour ensuite être analysés à l'aide de Atlas/ti. Les données ont été analysées dans le but d'identifier le genre de décisions prises par les stagiaires au sujet des enfants hypothétiques. L'analyse a révélé deux principaux types de prise de décisions. La plupart des stagiaires adoptaient une stratégie assez logique qui consistait à formuler des critères d'évaluation dans un premier temps, et à les appliquer aux devoirs dans un deuxième temps. Toutefois, quelques stagiaires ont pris des décisions qui ne semblaient pas fondées: ils ont fait des commentaires sur la qualité de vie des enfants ou ont désigné certains enfants comme ayant des besoins éducationnels spéciaux alors que l'information qui leur avait été présentée ne justifiait pas ce genre de conclusions.*

The evaluation of children's learning progress and achievement is a fundamental component of instruction. Over the past few years, a better understanding of how teachers conduct assessment in the classroom context has emerged

---

Dan Bachor is a professor in the Department of Educational Psychology and Leadership Studies, Faculty of Education. He teaches courses in assessment and instruction for students with learning problems, current issues in special education, research design, and developmental psychology. His research interests are in classroom assessment practices, individualized diagnostic assessment, and instruction for children with special educational needs.

Markus R. Baer was a graduate student studying in the Department of Educational Psychology and Leadership Studies, Faculty of Education when this study was conducted. He has taught at the elementary and secondary levels in various districts in Canada and Japan. He is currently a coordinator for the Student Assessment and Program Evaluation Branch working in the area of literacy examination and assessment, special education, and qualitative analysis.

(Bachor & Anderson, 1994; Broadfoot, 1992; McCallum, McAlister, Brown, & Gipps, 1992; Stiggins, Conklin, & Bridgeford, 1986). Bachor and Anderson found, for example, that teachers viewed classroom assessment as time-consuming. In spite of the time required to depict student achievement more fully, teachers placed a high value on the strategies that encompass classroom assessment, for example, direct observation of students as they complete work and, if working in groups, interact with each other; collecting portfolios, which include samples of student work that represent some of their achievements; and student self-assessment whereby students evaluate their own work and set instructional goals for themselves based on this analysis, thereby assuming great responsibility for their own learning. Less clear, however, is how preservice teachers develop an understanding of this process and interpret classroom assessment information.

As part of an ongoing investigation into classroom assessment practices (the Classroom Assessment Project, National—CAPNat), we have been examining the latter issue (Bachor, Shulha, Anderson, Wilson, & Muir, 1998; Bachor, Wilson, Shulha, Muir, & Anderson, 1997; Shulha, Anderson, Muir, Wilson, & Bachor, 1997; Wilson, Shulha, Anderson, Bachor, & Muir, 1997). As a first step in the process of examining novices Anderson (1999), Shulha (1999), and Wilson and Martinussen (1999) had 147 preservice teachers at Queen's University evaluate eight portfolio products over a 10-week period. Three hypothetical children, each named Chris, completed these products. Each version of Chris varied in the quality of the work that he or she completed. Shulha (1999) reported that the process of classroom assessment for novice teachers was multidimensional and called for continued exploration of the conditions that facilitate or hinder the process of classroom assessment.

In this study we revisit the question of the decisions taken by preservice (novice) teachers as they learn the process of classroom assessment. Whereas others (Bachor & Anderson, 1994; Broadfoot, 1992; Lock, 2000; Shulha, 2000; Wilson, 2000) have examined the ongoing decisions made by teachers as they grapple with assessment practices in the classroom context, we chose to focus on portfolio products that had been derived from classroom material but were compiled into sets of language arts materials completed by three hypothetical children. The advantage of incorporating portfolio products completed by these hypothetical children is that we were able to compare novice teachers' assessment decisions as they examined consistent sets of language arts material. Such comparisons are not possible in the context of a series of classrooms where the assessment dynamics change continually. Thus our purpose in this article was to examine as far as possible some of the decisions made by novice teachers as they grappled with the portfolio products of three hypothetical children to try to ascertain their assessment logic.

### *Methods*

#### *Real and Hypothetical Participants*

In designing this study, we invented three hypothetical children to serve as the focal point of the classroom assessment process. They are discussed below. We begin by describing the novice teachers.

*Participants.* One hundred and twenty-seven preservice teachers (PTs) who were attending the Faculty of Education, University of Victoria took part in this

investigation. They were completing the fourth year of university, a professional year of which a key component is the completion of an in-school practicum. At the time they responded to the portfolio material, they had completed three to four language arts courses, which prepared them to address the pedagogical issues raised in the portfolios.

*The hypothetical participants.* Three hypothetical children enrolled in grade 5 were created for the purposes of this investigation. Each of these hypothetical children differed in his or her classroom performance in two ways: (a) all three differed in their performance at the outset of the year; and (b) once a performance pattern was established, each student remained consistent throughout the year in the performance he or she displayed. The first child's assignments were poorly done; this person was struggling to master the fundamentals of language usage such as sentence structure. The second child consistently completed quality work, performing in the top portion of the class. The third child's assignments were at an acceptable level, approximately mid-point between the assignments completed by the other two students. These three children were not assigned names or a sex; rather, they were referred to as Student A, B, and C to try to prevent the PTs from ascribing attributions unrelated to student performance.

#### *Nature of the Portfolios Examined by the PTs*

Each portfolio contained six language arts assignments that were variations on the materials used originally by Anderson (1999), Shulha (1999), and Wilson and Martinussen (1999). They were modified for this investigation to be age- and grade-appropriate. Each PT was told to treat each of the portfolio materials as if it were requested by their sponsor teacher and to grade and comment on it in their journals accordingly. PTs were also provided with a set of instructions from this hypothetical sponsor teacher whereby they were given some background information on the assignments completed by the children and the total mark value of each assignment. Grading instructions were intentionally left ambiguous; thus PTs were not provided with marking criteria or other assessment guidelines such as rubrics.

The six language arts tasks examined by the PTs are given below.

1. *A Trip to the Mall* consisted of a short essay in which the grade 5 children describe a visit to a mall. This essay was completed using a word-processor.
2. *Did I Order an Elephant?* was a cloze exercise in which the children were given a reading passage that contained 15 missing words.
3. *A Salmon for Simon* consisted of a modified cloze reading task. Children had to select from five embedded multiple-choice alternatives a phrase to complete the provided text. These questions were accompanied by four multiple-choice comprehension items.
4. *The New Kid on the Block* was a reading worksheet. Children read a passage and answered a few short-answer questions. They were required to take the perspective of one character, interpret a phrase, and then translate a quote from the story into their own words.
5. *The Mending Wall* required children to read a poem and write a passage describing the personal meaning they derived from the poem.

6. *Final Exam* involved completing a formal written test. It contained four components: (a) word classifications items (which words are noun, adjective, etc.); (b) a paragraph in which the student had to extract some nouns and verbs; (c) some proofreading items to correct capitalization and one form of punctuation (commas); and (d) a reading passage that required children to answer multiple-choice comprehension items and some written responses.

#### *Procedure*

*Data gathered.* The preservice teachers were asked to track the three grade 5 students over the course of an academic term and to examine the language arts assignments given above, which were compiled as portfolios of the children's work. In addition, PTs were asked to assign marks to each completed assignment, to record their comments about this work, and to submit a final grade for each child. These data are reported in Anderson (2000). Finally, these preservice teachers were asked to keep a journal where they were to record their comments about their assessment process. These journals served as the data for this article.

The journal entries themselves varied in length from several lines to numerous pages. The original journals were transcribed, translated into a text file, and then stored as a single primary document as an *Atlas/ti* (Muhr, 1997) file. The journal data were then analyzed for patterns.

#### *Method of Analysis*

*Preliminary coding.* As a starting point, preliminary codes were developed from informed practice and the assessment literature (Bachor & Anderson, 1994). After the establishment of these initial categories, following Glaser and Straus's "constant comparison" method (Tesch, 1990), data from the first three participants were repeatedly coded with the goal of refining and reestablishing codes. Following the establishment of these preliminary codes, data from the first three cases were coded several times using *Atlas/ti* (a) to verify that the codes could be consistently applied across cases by both authors; (b) to ensure that codes were inclusive enough to allow the evidence to be classified comprehensively; and (c) to make sure that the codes did not contain redundancies.

Although the codes have their origins in existing theory and practice, they are grounded in the data to represent accurately and comprehensively the journal entries. A secondary purpose of repeated coding and comparison was to train for consistency. The relatively open-ended nature of the journal task resulted in responses that were at times vague or ambiguous. Thus code category boundaries required revision and refinement in order to deal with textual uncertainties. In turn, redundancy and overlap between categories were reduced.

*Code development.* Based on the literature and an initial examination of the *Atlas/ti* data, three superordinate categories were identified (see Table 1). Initially, comments were divided into those that were primarily *assignment-based* (dealing with the context, of the work completed, responding to the assignment criteria, or reacting emotionally to the assignment itself) and those labeled *person-based* (describing the competence, quality of life, or other comments directed specifically at the theoretical student as a person). Sub-



sequently, a third category termed *intervention* was added to parse out intervention suggestions that took the form of either comments or directives aimed at specific students. The three core codes of assignment-based, person-based, and intervention proved to fit the data on subsequent reworkings of subordinate categories. Eliminating redundancy, overlap, and ambiguity in lower-order code categories required several further revisions before 14 final codes were established. The final 14 codes classified into three superordinate categories are given in Table 1.

*Journal entry ambiguity.* Despite reworking the codes to reflect and adequately represent the complexity of the journal data, ambiguity and vagueness in the

Table 1  
Codes Assigned to Participants' Journal Data

<i>Code</i>	<i>Definition</i>	<i>Superordinate Category</i>
<i>Assignment-Based</i>		
<i>Context</i>		
Classroom	Points raised about the task, teacher, classroom, etc.	
Subject's background	Comments made about the preservice teacher's own background	
<i>Criteria</i>		
Establishing	Process of establishing assessment criteria	
Reviewing/Refining	Subsequent reviewing and refining of initial criteria	
<i>Questions/Comments</i>		
Concerns	Queries raised about the assignment/task	
Positives	Comments made about the assignment/task	
<i>Intervention</i>		
Comments	Hints of an intervention such as suggestions directed at task, class, teacher, etc.	
Student	Specific suggestions for an intervention, directed at either student A, B, or C	
<i>Person-Based</i>		
<i>Competence</i>		
Performance on Task	Statements about performance on task, directed to Student A, B, or C indicating how well he or she did on an assignment	
Student	Statements directed at the student going beyond task comments, designating the student, e.g., Student A is poor speller	
Classification	Statements directed at the student going beyond assignment comments. Designating one of the students as having a special educational need, e.g., learning-disabled, gifted, etc.	
Quality of Life	Statements directed at the student's family, such as commenting about their social economic status	
<i>Comments</i>		
Knowledge of	Comments indicating that knowing the student was important to participant's understanding of his or her progress as a learner	
Affective State	Statements made about the emotional state of either Student A, B, or C	

language of some participants' journal entries remained. For example, regarding one student's assignment, a participant wrote, "Watch for comprehension in other areas." It is unclear whether the comment is a reminder to the teacher-participant or a word of advice—suggesting an intervention—to the student. In another example, a participant wrote, "Student needs to work on context of her statements." Again, it is uncertain whether this suggests an intervention, merely advises the student where he or she erred, or is simply an effort to justify the grade assigned for the task. In such cases, face validity of the text was assumed and comments were taken at the textual level. The large data set rendered verification of codes with participants impractical, and thus textual inferences were kept to a minimum. Lower-order or broader code categories were applied when there was uncertainty. In both of the above cases, for example, the comments were coded with the larger category of person-based competence-performance on task.

*Inter-judge agreement.* Reliability checks for the code categories were conducted. We each independently coded three randomly selected sections of text consisting of between 100 and 150 lines per section on two separate occasions. A random number table was used to select the text segments. The independently coded sections were compared for consistency of code application using point-by-point agreement ratios (Kazdin, 1982). Reliability rates were checked twice; inter-judge agreement for the first check was 72%, and for the second was 96%. The average inter-judge agreement was 78%.

Using the categories given in Table 1, codes were applied to the collected text of all 127 participant journals. The amalgamated data were treated as one primary document, which was coded in its entirety prior to any analysis. On completion, participants were each given their own code in order to examine differences both across and between this group of preservice teachers. Throughout the data entry we met to check for coding agreement and to ensure consistency of coding.

*Dendrogram.* In addition to analyzing the journals entries using *Atlas/ti* (Muhr, 1997), the results were incorporated into a dendrogram (Miles & Huberman, 1984). A dendrogram is a "qualitative data analysis procedure that provides the reader with a picture of the increasing abstractions starting with a synopsis of the original evidence" (Bachor, 2001, p. 12). In this case, the dendrogram was constructed by using the categories described above. In the original conception, dendrograms provided only a visual picture of categories of evidence and logical abstractions. Recently, however, a few researchers have quantified the evidence presented in dendrograms (Scanlan, Stein, & Ravizza, 1989; Shulha, 1999). The advantage of this procedure is that the typicality of the reported evidence can quickly be determined (Bachor, 2001). In this case we drew the reported percentages from the data analysis that was conducted using *Atlas/ti*.

### Results

In presenting our results, we began with the evidence collected from all 127 participants and then parsed the data into a number of different groupings based on the conclusions that we deduced the participants made. Two main distinctions were drawn. First, we isolated those individuals whom we called *Task Restricted Participants* (TRP). Second, we pinpointed a second small cluster

of participants, whom we named *Student Elaboration Participants* (SEP). The patterns that each group of individuals appeared to follow are described below.

#### *Participants' Patterns*

Based on the comments that they made in their journals, novice teachers tended to follow one of two main decision paths—the logic the PTs seemed to use in making decisions about the children's achievement—in interpreting children's assignments. Most PTs, the Task Restricted Participants, seemed to be quite conservative in the decision path they appeared to follow in that they drew their conclusions from the evidence provided (see Table 2). However, a minority of individuals, the SEP appeared to make decisions regarding the hypothetical students they assessed that went well beyond the information given in the provided portfolios (see Table 3). In reading these tables, note that we have progressively eliminated an increasing number of participants as we describe the factors that individuals seemed to consider when making decisions. For example, in Table 3 we begin by presenting the decision path of

Table 2  
Task Restricted Participants' Decisions

<i>Decision Path</i>	<i>Number of Participants</i>
All Participants	127
Participants developing criteria (assignment-based criteria-establishing)	124
Participants making person-based competence-performance-on-task statements	105
<i>Excluded Participants</i>	
Participants who made classification comments	16
Participants who made quality-of-life comments	17
<i>Task Restricted Participants (TRP)</i>	
Participants who made no classification or quality-of-life comments	100
TRP who made no affective-state comments and no intervention statements	
TRP who made no affective-state comments and no intervention-comments	52
TRP who made no affective-state comments and no intervention-student comments	56
TRP who made no affective-state comments and no intervention-comments or intervention-student comments	33
TRP who made no affective-state comments and no intervention statements or person-based competence-student comments	
TRP who made no affective-state comments and no person-based competence-student comments	46
TRP who made no affective-state or intervention comments and no person-based competence-student comments	27
TRP who made no affective-state or intervention-student comments and no person-based competence-student comments	36
TRP who made no affective-state intervention comments or intervention-student comments and no person-based competence-student comments	22

Note. All numbers refer to participants who coded at least once or more with the specified categories.

all 16 SEP, hence the reduction in number of participants noted above. As one reads down the table, progressively more assessment comments—*quality of life* and *affective state* in the first instance—are added to note the decreasing number of SEP who included other factors in their decision-making about the three hypothetical children.

The vast majority of participants (see Table 4) established some criteria to judge the assignments they received (124 out of 127 participants). To illustrate, typical comments by PTs are the following two where the focus is on establishing guidelines for marking.

Each response is out of 3. There are 6 questions so task is out of 18 marks. 1 mark is given for each criteria [sic]:—is idea relevant to story & character 1—express ideas as Jimmy (I or me) 1—sentence thoughtful & clearly expressed 1.

Basically, I marked the answer correct if it seemed to reasonably fit into the context of the sentence. Although there were several instances where one student gave a much more appropriate response than another, I marked both of them right because they both were reasonable answers.

Some individuals elaborated the criteria they proposed, commenting extensively about the assignment they were assessing. For example, one person noted,

As I marked this assignment, I specifically looked for reading and writing comprehension. I read each student’s answer in context with the sentence and the story. In Part 2 I had trouble deciding what was the right answer for #3. I kept

Table 3  
Student Elaboration Participants

<i>Decision Path</i>	<i>Number of Participants</i>
All Participants	127
Student Elaboration Participants (SEP) (Person-based competence-classification)	16
SEP who made quality-of-life comments	6
SEP who made affective-state comments	8
SEP who made quality-of-life and affective-state comments	5
SEP who made intervention statements	
SEP who made intervention-comments	11
SEP who made intervention-student comments	10
SEP who made intervention-comments and intervention-student comments	7
SEP who made intervention-comments OR intervention-student comments	14
SEP who made quality-of-life, affective-state, and intervention-comments	4
SEP who made quality-of-life, affective-state, and intervention-student comments	4
SEP who made quality-of-life, affective-state, intervention-student, and intervention-comments	3
SEP who made quality-of-life, affective state, intervention-student, OR intervention comments	5

Note. All numbers refer to participants who coded at least once or more with the specified categories.

marking it wrong then right, so I decided to give everyone a mark for their answers. I do believe that Student B's answer was the most thought out and appropriate, but I also saw how Student A and C might have interpreted the questions and answered accordingly. Each answer was marked out of 1 mark for a total of 9 marks.

A small number of novice teachers (13/127 participants, see Table 4) were not satisfied with the initial criteria they established. They revisited the criteria they established either before or during the process of assessing assignments. For example, one person noted,

This is a rather difficult assignment. I wasn't even sure of some answers. As such, I modified my original marking scheme. I started out thinking that it would be smart to mark the first 5 either right or wrong, but I ended up giving 1/2 marks if it was semi-relevant, 0 if not consistent with the story, and 1 for the best choice. That way, the marks weren't so low.

In addition, 50 PTs (see Table 4) made comments about the context of the assignments they were asked to assess. These remarks centered on the artificial nature of the assessment, as the PTs were not setting the assignments but were judging work given by a hypothetical grade 5 teacher who is not well described in the context of the study because the focus is on the three hypothetical students. For example, one novice teacher commented, "Because I do not know exactly what the teacher has discussed with the students before doing the assignment it is more difficult to mark on what they actually wrote about (content)," whereas another was concerned about previous student learning, writing, "I wonder if students have worked with poetry before. I hope so cause this is a heady poem to interpret." Further, 21 individuals expressed discomfort

Table 4  
Participant Count by Code Categories

<i>Code Categories</i>	<i>Participant Count</i>
Assignment-Based Context-classroom	50
Assignment-Based Context-subject's background	21
Assignment-Based Criteria-establishing	124
Assignment-Based Criteria-reviewing/refining	13
Assignment Based Questions/Comments-concerns	112
Assignment Based Questions/Comments-positives	71
Intervention-comments	56
Intervention-student	59
Person-Based Competence-classification	16
Person-Based Competence-student	74
Person-Based Competence-performance on task	105
Person-Based Quality of Life	17
Person-Based Student Comments-affective state	28
Person Based Student Comments-knowledge of	21

Note. Participant count includes all participants who contained one or more instance of the specified category.

in assessing some components of the assignments given due to weaknesses in their own background. For example, one person noted, "Because I do not have much experience with marking I tend to question what I am doing."

*Task Restricted Participants.*

*Characteristics.* As shown in Table 2, most individuals (100/127), whom we term TRP, made no comments beyond judging the hypothetical children's work. That is, they tended to confine their comments to those related to the assignments, such as establishing criteria, without making any classification or quality-of-life comments regarding the student personally.

*Excluded individuals.* Twenty-seven participants (see Table 2) were excluded from further analysis in this category because they did not meet the criteria for task-restricted. Of this total, 17 participants made quality-of-life statements, and 16 participants concluded that some children had special educational needs. There was an overlap between these two sets of comments, as six participants made both types of statements. Some of these individuals are examined below under the category of student-elaboration.

*TRP patterns.* An examination of Table 2 reveals that some of the task-restricted participants were cautious about the statements they made. Some individuals (52/127) made no comments regarding the children's affective state (affective state comments ranged from neutral statements about not wanting to hurt a child's feelings to those indicating that a child was unhappy at school), nor did they make any general intervention suggestions (such as an assignment may need to be rethought). Fifty-six TRP did not make affective state comments or make student-specific intervention comments, such as "Student A needs help in spelling." A smaller subset of participants (33) made neither type of intervention statement, nor did they offer affective state comments.

The most conservative group (22/127), in addition to following the above pattern, further restricted their comments. They made no judgments about individuals' general abilities (person-based competence-student), such as "student cannot spell." As shown in the last part of Table 2, other variations on this pattern of the type of comments were made.

*Student Elaboration Participants (SEP)*

A small number of individuals, referred to as SEP, however, appeared to be willing to make judgments that exceeded the evidence provided. Seventeen novice teachers (see Table 4) made quality-of-life comments; that is, they commented about the quality of the family home or student's social life and how it was thought to have influenced the hypothetical child's school performance. For example, one SEP stated, "Student A seems to have a poor family life and it's reflected in his/her work." Another small group made comments about the affective state of some of the hypothetical students they were assessing. An example of this kind of comment is as follows: "Hard worker and likes to do many things at once—I'm hoping this won't be a detriment (pressure>stress)."

Additional substantial judgments were made by 16 SEP who were willing to consider designating one of the children as having special educational needs based on limited evidence. They made comments like "I'm wondering if they are ESL or some type of learning disability—why didn't the teacher offer extra



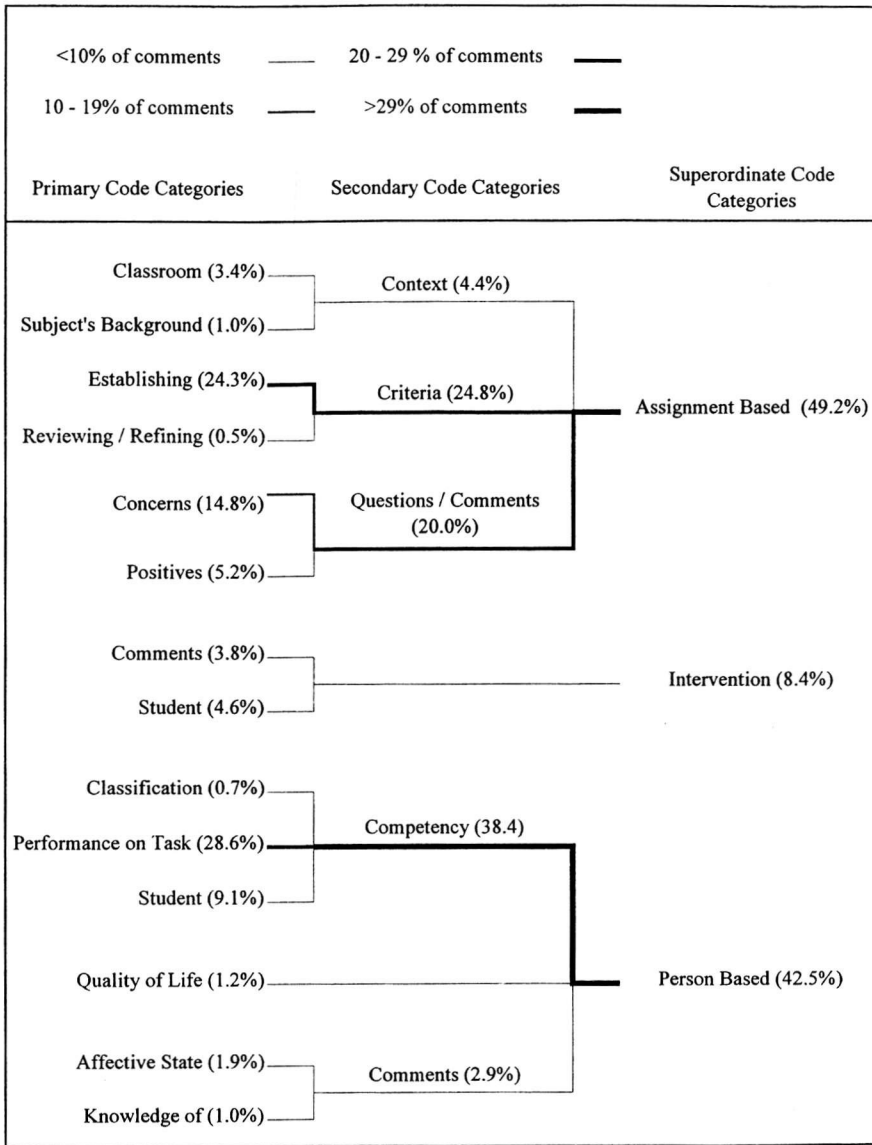


Figure 1. Dendrogram of novice teachers' pattern of assessment response.

assistance at some point?" Six of the 127 SEP also made quality-of-life comments, and a further subset of five of the six SEP made designations not only of special educational needs and quality-of-life concerns, but also went on to offer an intervention directed at the student. For example, one SEP commented,

Student A—needs a great deal of work with grammar [sic], spelling and sentence structure. I am wondering if this student has a learning disability or not one of the greatest home lives ... This student needs a great deal of encouragement and assistance [sic]. I hope that s/he gets it.

### *Interventions*

*Limitation.* Although we were able to isolate comments made by SEP, we were not able to differentiate completely between the types of intervention statements made across participants. Thus there might have been some overlap in intervention comments offered by the various PTs.

*Intervention-comments.* General comments about assignments (see Table 4) were made by 56 of the 127 participants. Prototypical examples include the following three, which illustrate the range of comments made.

One individual suggested, "As follow up, I would ask students to re-read their work for structure problems and make a lesson out of it." Focusing on the educator's role in the assignment, a PT stated, "The teacher should go over components/characteristics of an essay—paragraph breaks indentations etc." Another commented on what they themselves might do, saying, "I would spend much time reviewing this sheet because the students obviously did not understand this concept. Also a follow up lesson was needed to ensure it was learned as was done."

*Intervention-students.* In addition, 59 PTs made intervention comments specific to one of the three students. Examples of this latter type include the following three.

One PT noted some additional work might be required in rethinking an assignment: "I would perhaps return student A's paper and let him/her redo the assignment." Another person suggested that one of the hypothetical students might need some assistance in writing: "She needs to work on her run-on sentences; look out for these in the future." Finally, another PT offered suggestions to improve spelling: "I would encourage the student to use the dictionary and read over and proofread work for errors. Student may also have a peer read or assist with spelling. I would also encourage the student to slow down when he writes & try to write on the lines. I may have the student complete grammar exercises."

### *Reframing the Evidence: A Dendrogram*

Examining the evidence from another perspective, the PTs comments about the assessments were divided into two main categories. Looking at the dendrogram given in Figure 1, approximately half (49.2%) of the journal entries focused on the assignments the PTs addressed. These were divided into two main subgroups: setting or reviewing criteria (24.8%) and asking questions or commenting on the assignments (20%). The second common cluster of comments was centered on the three hypothetical students (42.5%). As can be seen, the bulk of these comments (38.4%) focused on the hypothetical students' competence. The majority (28.6%), however, were restricted to addressing specific aspects of the children's performance on the language arts assignments. A minority of comments (e.g., quality of life, 1.2%, or classification statements, 0.7%), however, were not supported by the evidence provided in the portfolios.

### *Discussion*

*Limitation.* A key limitation in analyzing the journals of the novice teachers who took part in this study is that we were not able to verify that the comments made actually reflected the decisions that these PTs would make in the class-

room. Each person was asked to comment about the process they were following as they assessed the three hypothetical students, and we took these comments at face value.

In addition, it is important to interpret our findings with caution because even those participants who made seemingly extreme comments often added contextual qualifications to their remarks. Thus we cannot ensure that the decision paths that we traced were the specific paths taken by the various participants.

*Decisions made.* Most of the novice teachers in this study appeared to make cautious decisions, staying close to the evidence they were given over the course of an academic term. When they had concerns, they centered on their own competence or lack of background, on the appropriateness of an assignment for a particular child, or on checking to see if a particular student needed some additional help in mastering some aspect of language arts. These individuals show consistency with the assessment patterns demonstrated by many other teachers (Shuhla, 1999). Following the scheme suggested by McCallum et al. (1992), these novice teachers seem to be moving in the direction of adopting the techniques of teachers classified as *systematic planners*. Systematic planners allow specific time for assessment, incorporating it into their teaching practice. They tend to employ a constructivist approach to learning, while acknowledging that children learn in idiosyncratic ways. Systematic planners also uphold the centrality of teaching, but add that their teaching effectiveness is enhanced by diagnostic assessment. Most of the PTs, then, appear to be adopting such an approach to teaching by systematically incorporating assessment evidence into their teaching practice.

A minority of individuals, however, presumably made assessment decisions that far exceeded the evidence provided. They seemed prepared to base their assessment decisions on some undefined assumptions. They appeared to have an intuitive basis for the judgments they made and speculated willingly about the three hypothetical learners and their families. Others (Bachor & Anderson, 1994; Broadfoot, Abbott, Osborn, Pollard, & Croll, 1993; Stiggins, 1999) have also noted the idiosyncratic nature of assessment. They have urged teachers to be prudent and systematic when conducting classroom assessment, as the cost of teachers using unsound assessment practices is too high.

*Implications for teacher educators.* Although caution is necessary when offering recommendations based on hypothetical cases, there are two general implications for teacher education. The first is that as PTs are learning to make assessment decisions, they need to be explicitly and frequently reminded to examine and articulate the assumptions they hold about learners. More specifically, novice teachers need to review how such assumptions influence (a) their view of learners' potential for progress, and (b) how they evaluate subsequent achievement. As far as possible, teacher educators need to intervene to dissuade novice teachers early in their teacher preparation from developing patterns of drawing conclusions based on "knowing a child well." The second implication for teacher educators follows from the comments made about systematic planners. Preservice teachers should be encouraged as much as is feasible to become systematic planners. That is, teacher educators need to

provide direct instruction (along with opportunities to practice through modeling and rehearsal) until novice teachers systematically incorporate assessment into their teaching practice in a pattern like those adopted by systematic planners.

*Conclusions.* Teacher educators can take some comfort in knowing that novice teachers for the most part have the skills to make fair assessment decisions and appear to be making reasonable decisions. One unanswered question, however, is whether these competences will be utilized in the classroom context where teachers have different levels of commitment to the students with whom they are interacting on a daily basis. In the present case, the presumed impartiality of the majority of participants may be a reflection of judging hypothetical students or other unidentified considerations.

For a small number of novice teachers, teacher educators must be vigilant in addressing the assumptions that seem to be held by any individuals who are prepared to make judgments based on sparse evidence. This concern is particularly justified when we consider the larger context of teachers' classroom assessment decision-making. Concern has been expressed about the basis that some teachers use to make decisions (McCallum et al., 1992). Specifically, some teachers make decisions about children based on their intuitive sense of a child, on the family and school history, or on limited encounters with an individual. These decisions tend to become rigid and are subsequently not readily amended. Whether teacher educators can influence such individuals to shift their assessment practices is unknown; however, every effort must be made to redress assessment practices that do not follow from evidence.

#### Acknowledgments

This research was supported in part by a grant from the Social Sciences and Humanities Research Council of Canada.

#### References

- Anderson, J.O. (1999). Modeling the development of student assessment. *Alberta Journal of Educational Research*, 45, 278-287.
- Anderson, J.O. (2000, May). *The evaluation of student achievement: Preliminary analysis in modeling teacher decisions*. Paper presented at the Canadian Society for Studies in Education Conference, Edmonton.
- Bachor, D.G. (2001). *Increasing the believability of case study reports*. Unpublished manuscript.
- Bachor, D.G., & Anderson, J.O. (1994). Perspectives on assessment practices in the elementary classroom in British Columbia, Canada. *Assessment in Education: Principles Policy and Practices*, 1, 65-95.
- Bachor, D.G., Shulha, L.N., Anderson, J.O., Wilson, R.J., & Muir, W. (1998, October). *Developing and maintaining collaborative research partnerships in classroom assessment*. Paper presented at the Measurement and Evaluation: Current and Future Research Directions for the New Millennium Conference, Banff.
- Bachor, D.G., Wilson, R.J., Shulha, L.N., Muir, W., & Anderson, J.O. (1997, June). *Diversity in approaches to assessment and instruction in the classroom*. Paper presented at the Canadian Society for Studies in Education Conference, St. John's.
- Broadfoot, P. (1992). Assessment developments in French education. *Educational Review*, 44, 309-326.
- Broadfoot, P., Abbott, D., Osborn, M., Pollard, A., & Croll, P. (1993, September). *SATS and teacher assessment: The changing reality of national assessment in primary schools*. Paper presented at the British Educational Research Association Conference, Liverpool.
- Lock, C. (2000, May). *The influence of two teachers' perceptions of knowledge and learning on their classroom practices*. Paper presented at the annual meeting of the Canadian Society for the Study of Education, Edmonton.

- Kazdin, A.E. (1982). *Single case research design: Methods for clinical and applied settings*. New York: Oxford University Press.
- Miles, M.B., & Huberman, A. M. (1984). *Qualitative data analysis: A sourcebook of new methods*. Newbury Park, CA: Sage.
- Muhr, T. (1997). *Atlas/ti: The knowledge workbench* (version 4.1). Berlin: Scientific Software Development.
- McCallum, B., McAlister, S., Brown, M., & Gipps, C. (1992, August). *Teacher assessment at key stage one*. Paper presented at the British Educational Research Association Conference, Stirling.
- Scanlan, T.K., Stein, G.L., & Ravizza, K. (1989). An in-depth study of former elite figure skaters: II. Sources of enjoyment. *Journal of Sport and Exercise Psychology*, 11, 54-64.
- Shulha, L.M. (1999). Understanding novice teachers' thinking about assessment. *Alberta Journal of Educational Research*, 45, 288-303.
- Shulha, L.M. (2000, May). *Understanding collaboration: Lessons from a large scale research project in teachers' classroom assessment practices*. Paper presented at the Annual meeting of the Canadian Society for the Study of Education, Edmonton.
- Shulha, L.M., Anderson, J.O., Muir, W., Wilson, R.J., & Bachor, D.G. (1997, June). *Collaboration in research and development*. Paper presented at the Canadian Society for Studies in Education Conference, St. John's.
- Stiggins, R.J. (1999). Evaluating classroom assessment: Training in teacher education programs. *Educational Measurement: Issues and Practice*, 18, 23-27.
- Stiggins, R.J., Conklin, N.F., & Bridgeford, N.J. (1986). Classroom assessment: A key to effective instruction. *Educational Measurement*, 5, 5-17.
- Tesch, R. (1990). *Qualitative research: Analysis types and software tools*. Philadelphia, PA: Falmer Press.
- Wilson, R.J. (2000, May). *Toward an integrated model of assessment-in-practice*. Paper presented at the Annual meeting of the Canadian Society for the Study of Education, Edmonton.
- Wilson, R.J., & Martinussen, R.L. (1999). Factors affecting the assessment of student achievement. *Alberta Journal of Educational Research*, 45, 267-277.
- Wilson, R.J., Shulha, L.M., Anderson, J.O., Bachor, D.G., & Muir, W. (1997, June). *Data issues in collaborative research*. Paper presented at the Canadian Society for Studies in Education Conference, St. John's.