The Alberta Journal of Educational Research *Vol. 53, No. 1, Spring 2007, 1-21*

C. Randy Duncan

and

Brian Noonan
University of Saskatchewan

Factors Affecting Teachers' Grading and Assessment Practices

Teachers' classroom grading and assessment practices are important elements of assessment reform. The purpose of this study was to examine the effect of classroom learning factors such as class size, subject area, and school size on teachers' classroom assessment practices. The results of a survey of 513 high school teachers showed evidence that teachers had implemented some aspects of assessment reform and that one classroom learning factor (i.e., subject) had a modest effect on teachers' practices. The results also indicate a need for more research on factors that affect teachers' assessment decisions and the effect of these decisions on implementing assessment reform.

Les pratiques d'évaluation et de classement des enseignants constituent des éléments importants de la réforme de l'évaluation. Cette étude porte sur l'influence qu'ont les facteurs d'apprentissage tels l'effectif d'une classe, le domaine d'études et la taille de l'école sur les pratiques d'évaluation des enseignants. Les résultats d'un sondage auprès de 513 enseignants du secondaire révèlent, d'une part, que les enseignants avaient mis en application quelques aspects de la réforme de l'évaluation et, d'autre part, qu'un facteur d'apprentissage (le domaine d'études) avait un léger effet sur les pratiques des enseignants. Les résultats indiquent également qu'il faut faire davantage de recherche sur les facteurs qui jouent un rôle dans les décisions que prennent les enseignants quant à l'évaluation et sur l'effet de ces facteurs sur la mise en pratique d'une réforme de l'évaluation. @H1 = Introduction

Assessment reform, an important element of school reform, has been characterized by two approaches: (a) the rise of large-scale assessment, and (b) changes in teachers' classroom assessment practices. Large-scale assessment as a form of accountability has been the subject of controversy and concern for some time (Hargreaves, Earl, & Schmidt, 2002; Shepard, 2000; Taylor, 1994). Nonetheless, large-scale assessment has been used to measure and compare student achievement results internationally, nationally, and in Canada provincially. Canadian students, for example, participate in the Programme for International Student Assessment (PISA), an international assessment program, and the School Achievement Indicator Program (SAIP), a Canadian assessment

Randy Duncan is a doctoral candidate in educational psychology whose research interests are in applied measurement in education, cognitive development in preschoolers, and assessment reform and its application to classroom instruction and assessment.
Brian Noonan is an associate professor of education whose research interests are in research methods, applied measurement in education, and the principles and practices of assessment reform and their application to classroom instruction and assessment.

program. In addition, most provinces in Canada have developed and used some form of provincial large-scale assessment (Taylor & Tubianosa, 2001).

Whereas large-scale assessment has been systematically implemented, as an element of assessment reform, teachers' classroom assessment practices have been examined from several perspectives. For example, proponents of assessment reform have advocated for fundamental changes in teachers' classroom assessment strategies based in part on the idea of assessment *for* learning, rather than assessment *of* learning. Stiggins (1999, 2001, 2002) and Guskey (1994, 2003), among others, have been leaders in advocating for this fundamental change in the purposes and processes of teachers' assessment practices. Similarly, there has been considerable interest in examining and promoting the role of formative assessment as an element of teachers' classroom practices. Black, Harrison, Lee, Marshall, and Wiliam (2004) and Boston (2002), for example, emphasized the role of formative assessment in improving student learning. This approach has also focused attention on teachers' use of alternative assessment practices such as peer- and self-assessment (Rolheiser & Ross, 2000).

One of the important contributions of research on assessment reform has been the focus on documenting teachers' classroom assessment practices. Gullickson (1985); Cizek, Fitzgerald, and Racher (1996); Brookhart (1993); Pilcher (1994); and Frary, Cross, and Weber (1993) provided some of the first analyses examining teachers' classroom assessment and grading practices. For example, Frary et al. (1993) found that secondary teachers used objective assessments most frequently as a classroom strategy, but had not examined assessment trends based on the influence of subject area. More recently McMillan (2001, 2003) and Duncan and Noonan (in press) have produced profiles of high school teachers' assessment practices. The profiles were aligned in an internal structure of grading practices, assessment strategies, and cognitive levels of assessment as first reported by McMillan in his 2001 study. Results of these studies show that high school and secondary teachers (i.e., McMillan's study with grades 6-12) have implemented some of the expectations of assessment reform while maintaining some traditional practices in others.

The focus on teachers' assessment strategies and grading practices has generated interest in research on teachers' assessment decision-making (Brookhart, 2003; McMillan, 2003). This research has pointed out that the various aspects of assessment reform have put pressure on teachers to accommodate classroom realities, internal policies, and large-scale testing. One of the results of this pressure was the need to develop measurement theory for classroom assessment (McMillan).

Why there has been an apparently uneven implementation of assessment reform is a complex question. Brookhart (2003) suggests that "assessment must be integrated with instruction, which implies that the meaning of the items or assessment tasks will be dependent on the environment" (p. 8). McMillan (2003) asserts that teachers' decision-making in the classroom is influenced by a variety of external factors (e.g., accountability testing, parents' expectations) and classroom realities (e.g., absenteeism, disruptive behavior, heterogeneity). Clearly teachers differ considerably on how they assess their students and weigh factors in determining grades. Earlier research by Cizek et al. (1996)

suggested that an important characteristic of classroom assessment and grading practices is that they are highly individualized and may be quite different from one teacher to another. Pilcher (1994) suggested that mathematics teachers perceive themselves as grading students more on cognitive abilities (i.e., measurable achievement factors) than teachers in other subject areas that integrate non-cognitive abilities (e.g., student effort and motivation) more in their decision-making on grading.

From a research perspective, it is important to know how teachers' grading practices and assessment strategies are influenced by types of classroom learning conditions (i.e., classroom size, teachers' training, teaching experience, grade level, and subject area). The measurement community has not been successful in influencing day-to-day teaching of assessment practices because their focus has not been on day-to-day life in the classroom or the connections between assessment and instruction (McMillan, 2003; Stiggins, 2001). Measurement specialists are typically those who carefully articulate standards of assessment quality and produce standardized tests based on measurable achievement factors.

The relationship between teachers' assessment strategies and grading practices and subject has not been well articulated. Bol, Stephenson, O'Connell, and Nunnery (1998) reported that among high school teachers, only the assessment practices of mathematics teachers differed from those of other teachers. They suggest that mathematics teachers rely less on traditional assessments such as tests and essays because they tend to focus on the process for arriving at an answer. On the other hand, McMillan (2001) found some evidence for a difference between mathematics teachers and all other teachers for grading practices, but not for assessment strategies or cognitive levels of assessment. Specifically, mathematics teachers reported using academic enablers, extra credit/borderline cases, graded homework, and the use of zeros less frequently than English, science, and social studies teachers. In addition, English teachers have reported more frequent use of assessment strategies showing more emphasis on constructed-response and teacher-developed assessments compared with both mathematics and science teachers (McMillan).

Although there is considerable research on the effect of class size on student achievement, there is little empirical information on the relationship between class size and teachers' assessment and grading practices (Locastro, 2001; Reynolds, Reagin, & Reinshuttle, 2001). It is important to know if there is a relationship because some alternative assessment practices (such as performance assessments or observation strategies) can be time-consuming for teachers compared with administering and scoring multiple-choice tests. Classes are typically considered too large when they do not "allow teachers to perform a variety of important teaching tasks such as developing lesson and unit plans, managing time, space, and materials to keep students productively involved in learning" (Reynolds et al., p. 31). The range at which class sizes become too large to meet the academic (and nonacademic) needs of learners is suggested to fall between 15 to 24 students. Thus large class sizes may be a factor in teachers' decisions as to whether they implement reform-oriented strategies such as alternative assessment.

It is also important to know to what extent school size may be a factor in teachers' assessment decisions because some research indicates a relationship between class size, school size, and achievement. Monk and Haller (1993) have suggested that greater school size may enhance efficiency and that there is a relationship between school size and curricular offerings in United States high schools. Efficiency in larger classes, typically those in larger schools, is suggested to be mediated by subject area and the degree to which class size and subject have minimal influence on student outcomes. However, the effect of school size remains controversial. Welsh (1989) has argued that a negative relationship exists between school size and achievement, suggesting that the larger the school size, the lower the achievement. Further, Welsh has suggested that lower achievement may be linked to administrative capability and the degree to which a principal's span of control is too large for effective instructional leadership.

If assessment reform is to be more fully implemented into classroom practice, it will be important to know what factors influence their assessment strategies and grading practices. The purpose of the study was to examine factors that may affect teachers' assessment strategies and grading practices, specifically subject matter, class size, and school size.

Methods

Data Collection

Data were collected using a survey questionnaire adapted from one used by McMillan (2001). The survey questionnaires were distributed to secondary teachers (i.e., grades 9, 10, 11, & 12) based on a stratified, random sample of 66 high schools in one Western Canadian province. A total of 513 secondary teachers responded, with the distribution by subject, class size, and school size identified in Table 1.

To examine the validity of the survey questionnaire, it was reviewed by a panel of high school teachers and administrators (N=15). The panel agreed that items were valid measures of teachers' grading practices, assessment strategies, and cognitive levels of student assessment. There was consensus that the items were sufficiently comprehensive (i.e., representative) in this internal structure of assessment practices, and the panel recommended no additional items or modifications. The three sections of the survey combined for a 34-item rating scale and a fourth section included several open-ended questions (Appendix). The survey also included a section for demographic information about grade level taught, subject area, number of students in the class, and total number of students in the school. The types of subject area (mathematics, English, science, social studies, practical arts, and other) were similar to those used in McMillan's (2001) study. However, the tendency for broad groupings of subject area (e.g., placing physics, chemistry, and biology in one subject area like science) raised the possibility of limitations and suggested that some caution may be necessary in interpreting the results. Teacher self-report on the fixed-response items contributing to factors in the internal structure of assessment practices (i.e., grading practices, assessment strategies, and cognitive levels of assessment) were measured by their response to a six-point Likert-scale anchored with endpoints of *not at all* (1) to *completely* (6).

Table 1
Number of Teachers by Subject, Class Size, and School Size

	Subject						
	Math	*Sci*	*Eng*	*SocSt*	*PrArts*	*Other*	*Total*
Class Size							
<15	26	11	24	7	27	16	111
15-25	33	40	44	24	26	39	206
26-35	33	32	48	28	9	34	184
>35	2	2	3	4	1		12
Total	94	85	119	63	63	89	513
School Size							
<100	21	26	26	13	16	19	121
101-250	21	13	30	14	9	12	99
251-400	17	11	15	11	10	10	74
401-600	4	7	6	5	8	6	36
>600	31	27	41	19	20	45	183
Total	94	84	118	62	63	92	513

A conceptual framework for the 34-item instrument was developed by McMillan (2001) in which the assessment practice items were reported as having an internal structure of grading practices, assessment strategies, and cognitive levels of student assessment. The items contributing to each of these three constructs were based on earlier research about teachers' assessment and grading practices (Brookhart, 1993; Frary et al., 1993; Stiggins & Conklin, 1992). These studies examined various aspects of assessment that teachers considered in assigning grades (e.g., student effort, improvement, academic performance), types of assessment strategies used (e.g., authentic assessments, major exams, essay-type questions), and the cognitive levels of assessment (e.g., recall knowledge, understanding, reasoning, etc.). McMillan examined the internal structure of assessment practices in greater detail and ultimately identified 12 factors based on principal component extraction. He had suggested six factors for grading practices (i.e., academic enabling behaviors, use of external benchmarks, academic achievement, use of extra credit and borderline cases, use of graded homework, and use of homework not graded), four factors for assessment strategies (i.e., constructed-response assessments, focus on assessment developer, grouped quizzes with objective assessments, and emphasis on the use of major exams), and two factors for cognitive levels of student assessment (i.e., higher-order thinking, and recall knowledge).

Data Analyses

To explore some of the underlying dimensions of assessment practices better, the current sample was subjected to two methods of factor analysis. Both solutions were based on a principal component analysis, the first orthogonal using Varimax and the second oblique using Oblimin. Whereas the current study still reflected the same general internal structure as the McMillan (2001) study (i.e., grading practices, assessment strategies, and cognitive levels of

student assessment), the data supported only a five-factor solution based on principal component extraction (see Tables 2 and 3).

Descriptive statistics were computed to summarize aggregated mean scores for the five factors of teachers' assessment practices. A multivariate analysis of variance (MANOVA) was conducted to examine the effects for subject, classroom size, and school size and to account for the relationship between multiple dependent variables.

Results

The relative frequencies of use for the five factors of assessment practices were determined by examining their respective mean scores (Table 4). Overall, the use of higher-order thinking (i.e., measures of student understanding, reasoning, and application), a factor *within the structure of cognitive levels of assessment* was the most frequently used cluster of assessment practices by secondary teachers (M=3.97), which reflected use of *quite a bit*. The most commonly used aspect of this factor was assessments that require understanding (M=4.08), which reflected use between *quite a bit* and *extensively*.

The factor *within the structure of assessment strategies* reported as having the highest frequency of use by secondary teachers was grouped quizzes with objective assessments (M=3.72), which reflected use between *some* and *quite a bit*. The most commonly used aspect within this factor was emphasizing the use of major exams (i.e., a summative practice for calculating final grades or at least a unit end test, M=3.81), which also reflected use between *some* and *quite a bit*. Even teachers in practical arts (M=2.95) and the subject area described as other (M=3.16) reported emphasizing the use of major exams *some.*

The factor *within the structure of grading practices* reported as having had the highest frequency of use by secondary teachers was academic enabling behaviors (M=3.28), which reflected use between *some* and *quite a bit*. The most commonly used aspect within this factor was student effort (i.e., how much the student tried to learn, M=3.79), which reflected use between *some* and *quite a bit*.

Factors Affecting Assessment Practices

The effects for subject, classroom size, and school size were examined by conducting a MANOVA to account better for any differences between groups when the five dependent measures (i.e., factors) were examined simultaneously. The factor correlations were all low to moderately low (see Table 3). The

Table 2
Reliability Coefficients for Factor Structure

Structural Component	*Factor*	*No. of Items*	*Alpha*
Grading Practices		11	0.83
	Academic enabling behaviors	7	0.86
	Use of external bench	4	0.64
Assessment Strategies		9	0.67
	Constructed response assessments	5	0.83
	Grouped quizzes / objective assess.	4	0.70
Cognitive Levels of Assessment		3	0.73
	Higher-order thinking	3	0.73

Table 3
Dimensions of Assessment Practices by Principal Component Analysis

	Varimax rotated components from principal component analysis						*Oblimin rotated components from principal component analysis*					
Items	*1*	*2*	*3*	*4*	*5*	h^2	*1*	*2*	*3*	*4*	*5*	h^2
Specific learning objectives mastered	.125	–.087	.554	–.004	–.057	.333	.095	–.095	.206	–.041	–.609	.333
Academic performance as opposed to other factors	–.233	.001	.375	.469	.048	.464	–.287	.390	–.016	.136	–.373	.464
Ability levels of students	.518	.297	.085	.008	–.087	.377	.492	–.001	–.210	–.182	–.004	.377
Student effort—how much the student tried to learn	.785	.135	.053	–.020	–.062	.644	.853	–.003	.035	–.179	.026	.644
Degree to which the student pays attention and/or participates in class.	.828	.056	.008	–.046	.064	.697	.922	–.011	.135	–.049	.058	.697
Effort, improvement, behavior, and other nontest indicators for borderline cases	.753	.081	.045	.053	.098	.589	.831	.074	.089	.006	.027	.589
Improved performance since the beginning of the year	.774	.166	.141	–.041	.081	.656	.814	–.046	.005	–.014	–.077	.656
Work habits and neatness	.735	.126	.143	–.031	.078	.587	.771	–.026	.052	–.021	–.086	.587
Performance compared with other students in the class	.258	.131	–.093	–.048	.618	.490	.219	–.021	–.093	.615	.108	.490
Disruptive student performance	.668	–.031	–.086	.056	.248	.518	.767	.113	.190	.169	.143	.490
Extra credit for nonacademic performance (e.g., bringing items for food drive)	.376	.117	–.105	.052	.406	.336	.367	.097	–.044	.358	.153	.336
Formal or informal school division policy of the percentage of students who may obtain As, Bs, Cs, Ds, or Fs	.112	.004	.107	.149	.547	.440	.021	.186	.084	.525	–.109	.440
Performance compare to students from previous years	.041	.030	.074	–.224	.757	.642	–.046	–.230	.004	.798	–.136	.642
Grade distributions of other teachers	–.041	.035	.002	.063	.741	.571	–.148	.096	–.012	.764	–.022	.571

Table 3 (continued)

Items	*Varimax rotated components from principal component analysis* 1	2	3	4	5	h^2	*Oblimin rotated components from principal component analysis* 1	2	3	4	5	h^2
Performance quizzes	.080	–.058	–.010	.668	–.013	.457	.139	.698	.086	–.006	.098	.457
Objective assessments (e.g., multiple choice, matching, short answer)	–.017	.102	–.079	.674	.057	.479	–.030	.712	–.121	.055	.186	.479
Essay-type questions	–.030	.776	.014	.114	.109	.628	–.283	.062	–.906	.078	.099	.628
Performance assessments (e.g., structured teacher observations or ratings of a performance such as a speech or paper)	.351	.743	.070	–.135	.070	.704	.155	–.189	–.775	–.010	.033	.704
Projects completed by individual students	.361	.619	.187	–.195	.029	.588	.182	–.259	–.608	–.050	–.118	.588
Major exams	–.286	–.079	.179	.716	.075	.639	–.319	.708	.061	.138	–.140	.639
Authentic assessments (e.g., real-world performance tasks)	.422	.279	.212	–.117	.019	.366	.334	–.139	–.177	–.072	–.175	.366
Projects completed in teams of students	.166	.703	–.049	–.071	.058	.569	–.051	–.088	–.769	–.029	.153	.569
Oral presentations	.195	.803	–.024	–.029	.119	.703	–.041	–.064	–.885	.043	.145	.703
Assessments that measure student understanding.	–.012	.019	.760	.195	.020	.617	–.126	.063	.082	.066	–.809	.617
Assessments that measure how well students apply what they learn	.205	.167	.767	–.087	–.002	.666	.064	–.228	–.039	–.003	–.815	.666
Assessments that measure student reasoning	.077	.146	.775	.112	.072	.652	–.087	–.019	–.033	.088	–.814	.652
Assessments that measure student recall	.085	–.049	.197	.634	.025	.458	.092	.637	.117	.027	–.133	.458
Percent of variance accounted for	19.139	9.328	6.654	6.230	5.196		19.139	9.328	6.654	6.230	5.196	
Eigenvalue	6.507	3.171	2.262	1.767	1.477		6.507	3.171	2.262	1.767	1.477	
Alpha reliability coefficients	.86	.70	.83	.64	.73		.87	.70	.83	.64	.73	

Table 3 (continued)

Items	Varimax rotated components from principal component analysis						Oblimin rotated components from principal component analysis					
	1	*2*	*3*	*4*	*5*	h^2	*1*	*2*	*3*	*4*	*5*	h^2
Items that did not load at .400 or greater												
Performance compared to a scale of percentage correct	.075	–.021	–.034	.345	–.126	.147	.131	.358	.028	–.127	.091	.147
Inclusion of zeros for incomplete assignments in the determination of final percentage correct	–.026	.317	.125	.221	–.243	.264	–.090	.157	–.382	–.233	–.059	.147
Quality of completed homework (graded)	.181	.369	.292	.254	–.050	.321	.068	.188	–.349	–.074	–.220	.321
Completion of homework (not graded)	.270	.088	.146	.195	.029	.149	.256	.191	–.009	–.039	–.100	.149
Extra credit for academic performance.	.395	.198	.112	.081	.257	.290	.365	.066	–.122	.225	–.064	.290
Assignments designed primarily by yourself	–.021	–.036	.112	.087	–.109	.639	.080	–.002	–.017	–.002	–.118	.639
Assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)	.066	.097	.122	.076	–.016	.623	–.062	.122	–.015	–.122	–.101	.623
Factor Correlations							*1*	*2*	*3*	*4*	*5*	
1. Academic Enabling Behaviors							1.000	–.501	–.268	–.030	–.262	
2. Constructed-Response Assessments (Factor 3 in Oblimin)							–.501	1.000	.334	–.088	–.170	
3. Measuring Higher-Order Thinking (Factor 5 in Oblimin)							–.268	.334	1.000	–.281	–.257	
4. Grouped Quizzes with Objective Assessments (Factor 2 in Oblimin)							–.030	–.088	–.281	1.000	–.079	
5. Use of External Benchmarks (Factor 4 in Oblimin)							–.262	–.170	–.257	–.079	1.000	

a. Varimax rotation converged in 7 iterations.
b. Oblimin rotation converged in 8 iterations.

Table 4
Mean Factor Scores

Component	*Mean*	*Std. Dev.*	*N*
Grading Practices			
Academic enabling behaviors	3.280	0.77	513
Use of external benchmarks	1.500	0.50	513
Assessment Strategies			
Objective assessments with grouped quizzes	3.720	0.70	513
Constructed-response assessments	3.080	0.93	513
Cognitive Levels of Assessment			
Higher-order thinking	3.960	0.67	513

only significant multivariate test was a main effect for subject (Wilks' Lambda: $F[25, 1565] = 9.40$, $p<.0005$). There were no significant multivariate tests for any two-way interactions, the three-way interaction, or the main effects for either classroom size or school size.

Effects of Subject

A main effect was present for subject area taught in four of the five factors. The factor *use of external benchmarks* did not contribute to the contrast of dependent variables that led to a significant multivariate F (Table 5).

The mean factor scores in Table 6 illustrate the relative degree of preference for assessment practices by subject area taught. There was a consistent preference across all subject areas for high use of higher-order thinking strategies (i.e., measures of understanding, reasoning, and application) and low use of external benchmarks (i.e., performance compared with other students, grade distributions of other teachers, school policy on normal curve distributions). More specifically, mathematics and science teachers displayed a relative preference for grouped quizzes with objective assessments (including an emphasis on major exams, M=4.07 and M=4.08 respectively) and equally reported a relative avoidance of the use of external benchmarks (M=1.31 and M=1.48

Table 5
Tests of Between-Subjects Effects Showing Significant Omnibus *F* Tests

Factor	*Environmental Influence*	*F*	*Sig.*	*Equality of Variance*
Grading Practices				
Academic enabling	Subject	5.429	.0005	ns
External benchmarks	Subject	2.186	ns	ns
Assessment Strategies				
Constructed-response	Subject	22.202	.0005	sig.
Grouped quizzes	Subject	12.443	.0005	sig.
Cognitive Levels				
Higher-order thinking	Subject	3.554	.004	ns

Table 6
Statistically Significant Mean Factor Scores by Subject Area

	Subject Area							
Component	*Math n=93*	*Sci. n=85*	*Eng. n=119*	*Soc. n=63*	*P. Art n=62*	*Other n=91*	*F*	*P<*
Enabling behaviors	2.82	3.04	3.30	3.32	3.65	3.67	5.429	.000
External benchmarks	1.31	1.48	1.51	1.55	1.58	1.57	2.186	.055
Constructed-response	1.83	2.96	3.68	3.63	3.18	3.25	22.202	.000
Grouped quizzes	4.07	4.08	3.55	3.90	3.35	3.35	12.443	.000
Higher-order thinking	4.01	4.00	4.03	3.88	4.00	3.84	3.554	.004

respectively). For a more comprehensive ranking of the 34 individual items, see Duncan and Noonan (in press).

Grading practices. A main effect of subject area was found for one factor (i.e., academic enabling behaviors) within the structural component of grading practices (Table 7). This was consistent with McMillan's (2001) research where subject area differences were also found for the component of academic enabling behaviors. Mathematics teachers emphasized academic enabling behaviors (i.e., ability level, student effort, paying attention, improved performance, work habits, and level of disruptive performance) less than teachers in English, social studies, practical arts (i.e., industrial arts, home economics), and other (i.e., performing arts, fine arts, religious studies, see Table 7). Mathematics teachers' lower emphasis on this factor was consistent across all seven aspects of academic enabling behaviors. Mathematics teachers did not differ from science teachers on this factor, however, science teachers differed from (i.e., consistently lower emphasis across all aspects) teachers in practical arts and other.

Teachers classified as other (i.e., performing arts, fine arts, religious studies) emphasized academic enabling behaviors more than teachers in mathematics, science, English, and social studies. This higher emphasis on this factor was consistent across all seven aspects of academic enabling behaviors. In addition, teachers classified as other did not differ from practical arts teachers. However, practical arts teachers differed (i.e., consistently higher emphasis across all aspects) from mathematics, English, and social studies teachers.

Assessment strategies. A main effect for subject was found for the two factors (i.e., constructed-response assessments and grouped quizzes with objective assessments) within the structural component of assessment strategies (see Table 5). This was consistent with McMillan's (2001) study where subject differences were also found for constructed-response assessments and grouped quizzes with objective assessments. McMillan suggested that English teachers exerted the largest influence on the size of the main effect across his four factors of assessment strategies. However, the current study has suggested that mathematics teachers are more consistently different than teachers in any of the other five subject areas across the two factors.

Mathematics teachers emphasized constructed-response assessments (i.e., essay-type questions, performance assessments, individual projects, team projects, and oral presentations) less than teachers in all five other subject areas.

Table 7
Statistically Significant Differences on Components between Six Content Areas

Component	*Subject Differences*
Grading Practices	
Academic enabling behaviors	Math < English, Social St., Pr. Arts and Other Other > Math, Science, English, & Social St. Pr. Arts > Math, Science, & English Science < Pr. Arts and Other
Use of external benchmarks	N/A
Assessment Strategies	
Constructed-response	Math < Sci., Soc. St., English, Pr. Arts, and Other English > Math, Science, Pr. Arts, and Other Social St. > Math, Science, Pr. Arts, and Other
Grouped quizzes/objective assessment	Math > English, Pr. Arts, & Other Science > English, Pr. Arts, and Other English < Math, Science, and Social St. Social St. > English, Pr. Arts, and Other
Cognitive Levels of Assessment	
Higher-order thinking	N/A

Based on $p<.05$ using Scheffé post-hoc test.

The lower emphasis was consistent across all five aspects of this factor. On the other hand, English teachers emphasized constructed-response assessments more than teachers in all subject areas other than social studies. This was consistent across all five aspects of this factor. Social studies teachers also reported higher frequency of use across all five aspects of this factor compared with mathematics, science, practical arts, and other teachers.

Mathematics teachers generally emphasized grouped quizzes with objective assessments (i.e., performance quizzes, objective assessments, major exams, and assessments that measure recall) more than teachers in English, practical arts, and other. However, the higher emphasis was based on only three of the four aspects of this factor. For objective assessments (i.e., multiple choice, matching, short answer) mathematics teachers reported a lower frequency of use compared with science and social studies teachers.

Science teachers emphasized grouped quizzes with objective assessments more than teachers in English, practical arts, and other. This higher frequency of use was consistent across all four aspects of this factor (i.e., performance quizzes, objective assessments [e.g., multiple choice, matching, short answer], major exams, and assessments that measure student recall). English teachers emphasized grouped quizzes with objective assessments less than mathematics, science, and social studies teachers. This lower frequency of use was consistent across all four aspects of this factor. Social studies teachers emphasized these assessment strategies more than teachers in English, practical arts, and other. This higher frequency of use was also consistent across all four aspects of this factor.

Cognitive levels of assessment. A main effect for subject was found for the one factor of cognitive levels of assessment. However, for this factor of higher-order thinking (i.e., measuring understanding, reasoning, and application) no significant differences were found using either Scheffé or Tukey post-hoc tests. The largest difference suggested between subject groups was for mathematics teachers (M=4.00) and teachers classified as other (M=3.84). The range for these mean scores, based on the standard error of the mean (0.071 and 0.076 respectively), does not overlap. An interpretation biased in favor of a practical difference between these two subject areas would have suggested that mathematics teachers emphasized higher-order thinking more than teachers in such areas as performing arts, fine arts, and religious studies. Fully demonstrating such an actual difference would require support from future studies, as in good conscience this potential subject difference cannot be reported as a significant finding in the current study. It is more likely that some combination of dependent variables, a combination of groups (as there are more than three subject groups), or a combination of both dependent variables and groups led to the significant multivariate F statistic.

Discussion

Documenting high school teachers' assessment practices in one Western Canadian province has provided insights into the extent to which external conditions (factors) may influence teachers' rationale for their grading practices and assessment strategies. The focus of this study has been to create a profile of assessment strategies and grading practices in high schools and to provide an indication as to the extent to which these elements of assessment reform have affected teachers' classroom assessment practices. Results of this exploratory study are intended for administrators at the school and school board levels, as well as classroom teachers, to examine the effects of teachers' practices. The results helped to identify the limitation of class size and school size as potential environmental influences on high school teachers' assessment practices. Unlike subject area, class size and school size did not produce a main effect for any of the five factors or any interaction effects with subject area.

Influence of Subject Area

A finer distinction was used for subject area in the current study, and instead of four content areas (i.e., mathematics, science, English, and social studies) subject areas were also delineated by practical arts (i.e., industrial arts, home economics) and other (i.e., performing arts, fine arts, religious studies). In this study, mathematics surfaced as the most consistently different subject area across two of the three factors (i.e., academic enabling behaviors and constructed-response assessments) that had significant post-hoc results. This suggests that mathematics teachers may have a different perspective (i.e., less frequent use) on one factor for each of grading practices and assessment strategies when compared with teachers in the four other subject areas.

As expected, based on earlier research, mathematics teachers demonstrated a practical difference in their lack of preference for using non-cognitive abilities as contributing elements to grading practices (i.e., academic enabling behaviors such as student effort, motivation, work habits, neatness, non-disruptive performance, perception of paying attention) when compared with all other sub-

ject areas. Mathematics teachers also demonstrated a practical difference (i.e., significantly lower use) in their preference for constructed-response assessments such as structured observations and essay-type questions. The inherent subjectivity in these types of assessment strategies may have suggested some similarity to those elements of academic enabling behaviors and a relative avoidance of calculating grades on the basis of classroom activities that are loosely categorized as non-cognitive abilities.

It was also anticipated that mathematics teachers would demonstrate a preference for using assessment strategies based on more explicit cognitive abilities. In the current study, mathematics teachers did demonstrate a preference for using grouped quizzes with objective assessments (i.e., performance quizzes, multiple-choice tests, assessments that measure student recall, and major exams) as key elements of classroom assessment strategies more than teachers in English, practical arts, and those classified as other. However, mathematics teachers were not distinct in their preference for this factor of assessment strategies from either science or social studies teachers. It is conceivable that mathematics, science, and social studies (at least when it comes to the historical elements) teachers perceive that they have relatively less ability to provide opportunities to students for flexible interpretations (i.e., several or many correct answers) than teachers in English, practical arts, and other.

A limitation of this study is the inability to elaborate on how the process focus of mathematics teachers translates into differing frequencies of use within these factors. McMillan's (2001) findings have suggested that a good starting point for examining decision-making about how grading practices are influenced is via the relative desire to use practices that motivate students, encourage student participation, and facilitate understanding. It is important to know how teachers decide on classroom assessment practices and current classroom practices that better facilitate the interests of students (i.e., improve learning), the goals and interests of instructors (e.g., efficiency of instruction relative to numbers of students), and/or manage adequately to meet the needs of both in the classroom environment.

The results of the descriptive analysis indicate the most frequently used factor in the structure of grading practices across all subject areas was the use of grouped quizzes with objective assessments. These results reflect a possible shift in what may influence the grading practices of high school teachers (i.e., grades 10 to 12) because earlier research suggested that academic achievement was the most frequently used factor (McMillan, 2001). McMillan based his findings on secondary teachers (i.e., grades 6-12) and a three-variable factor for academic achievement (i.e., performance compared with a scale of percentage correct, specific learning objectives mastered, and academic performance as opposed to other factors), which did not surface as a factor in the current study. The influence of middle school model (i.e., more focus on authentic assessment) in McMillan's sample may account for some of the differences identified in the dimensionality of the instrument.

The current study indicates that the influence of subject area has a small effect size (i.e., Eta squared range of 0.040 to 0.207) and that power for the current study was adequate (i.e., varied along a range of 0.716 to 1.00). The extent to which subject area as a condition for learning influences teachers'

classroom assessment practices cannot be fully understood from this study due to limitations such as the depth of interpretation available from quantitative data. However, the results from the current study do make an initial contribution by informing teachers, administrators, and policy-makers what assessment decisions are currently being incorporated into the classroom environment and to what extent this represents assessment reform and the idea of assessment for learning.

Influence of Class Size

Teachers have suggested that with smaller classes they are able to vary their teaching methods to provide a richer learning environment for students (Alberta Teachers' Association, 2003). The results of the current study do not support the position that teachers actually vary their assessment practices based on class size alone. However, the current study indicates that the effect size of a learning condition like class size is very small (i.e., Eta squared range of .000 to .018), and the observed power of the study (i.e., a range of .061 to .486) may not have been considered adequate to detect such small differences. In addition, the current study does not purport to account for the peripheral benefits of smaller classes (e.g., classroom management, more time for one-on-one interaction, more constructed-response assessments) that may positively affect more accurate assessment. Therefore, it remains unclear whether teachers with smaller class sizes have demonstrated any preference for classroom assessment practices that would be indicative of an assessment for learning environment. The current nature of high school assessment practices reflect the use of cognitive ability measures as frequently as non-cognitive assessments like ability levels of students and student effort (i.e., academic enablers).

Influence of School Size

Of interest to the present study was the general perception that the larger the school population, the larger the corresponding class sizes were likely to be. A positive linear relationship between school size and class size was considered a possible source of influence on teachers' assessment practices. No significant main effect for school size or an interaction effect with class size was found in the current study. The current study (based on 66 high schools) indicates that an effect size for the influence of school size is very small (Eta squared range of .002 to .011) and that the observed power for the study would not have been considered adequate (i.e., a range of .102 to .375) to have detected any such small differences.

The practical value of an influence like school size when sampled in a single province may have had some limitations for an accurate examination of such a learning condition due to factors such as a disproportionate number of small rural high schools. The delineation of relative school sizes may be somewhat unique to a single province's demographics, and such group characteristics may not be meaningful in other provinces. Further study is warranted relative to the position by Welsh (1989) that the relationship of school size and academic achievement changes (i.e., becomes less negative) when the school size approaches 700 due to the tendency for schools to have more administrative support. Future studies may need to recruit a larger sample of high schools that is more variable (i.e., better representative of larger school sizes).

In Canada school policy, and more specifically classroom assessment practice, is a provincial responsibility and consequently may be different in terms of emphasis and frequency of use across jurisdictions. For example, the Western Canadian province of interest to this study is a jurisdiction where accredited teachers provide the assessment and grade for certain grade 12 subject areas. Most other provinces have developed policy mandating some combination of teachers' assessment and summative evaluation based on large-scale testing to formulate final grades. Taylor and Tubianosa (2001) have suggested that "almost all provinces are turning to different modes of maximizing the use of assessment information" (p. 72) and that plans relative to improving programs based on assessment results are emerging at various levels within and between provincial jurisdictions.

Summary

Given a sample size of 513 high school teachers, the current study provides a useful perspective on the current status of secondary classroom grading and assessment practices. It provides a starting point to inform instruction for new teachers on the paradox of using criterion-referenced grading scales (i.e., rubrics), but in reality still using varying degrees of norm-referenced judgments about grading and assessment. High school teachers in the province use assessment strategies that they develop; large-scale assessments, although used by the province, are not used by teachers for classroom assessments. In the current study, high school teachers used non-cognitive practices such as academic enabling behaviors or traits (i.e., which includes items like *effort, improvement, behavior*, and other nontest indicators for borderline cases) between *some* and *quite a bit* (M=3.28). This is probably an indication that most teachers perceive a need to help and pull for borderline cases. These types of issues require undertaking additional studies to understand why teachers use comparative judgments in assessment and if they think these comparisons are fair.

The present study replicates an examination of three internal structural aspects of assessment practices (i.e., grading practices, assessment strategies, and cognitive levels of student assessment) based on the influence of subject area and has extended the scope of investigation to include an initial examination of class size and school size. However, this research could benefit from increasing the overall sample size of high school teachers and perhaps also including a sample of middle school teachers. This latter recommendation would be prudent as high schools in some provinces include grade 9 (i.e., a middle-years curriculum). This would add to the validity of results (i.e., including those of McMillan [2001] who surveyed secondary teachers from grades 6-12) by examining existing differences to determine if they remain consistent across additional grade levels and permit further investigation of the internal structure of assessment practices based on possible curriculum differences between middle and secondary programs. Also, it would improve the possibility of addressing the issue of small effect sizes for the influence of subject, class size, and school size. Although documenting the existing state of classroom learning conditions is a useful and necessary initial step in understanding and improving teachers' classroom assessment decision-making, it would also be prudent to examine why teachers use specific assessment techniques

and approaches. In addition, an examination of the nature of teacher decision-making should go beyond the influence of external factors to an investigation of the constraints teachers are under (i.e., realities of classroom management), their internal beliefs and values, and their decision-making rationale for using assessment practices.

An underlying rationale for examining the state of classroom assessment practices was that teacher training in classroom assessment may be inadequate based on the expectations of assessment reform (Brookhart, 2003; Gusky, 2003; Stiggins, 2002). This is not an issue only with Canadian teachers, as Stiggins (1999) has claimed that only about half of the US states have a requirement that teachers meet any type of assessment competence standards and/or complete assessment coursework. This is an important aspect of guiding principles for instruction in classroom assessment because as McMillan (2003) has suggested, teachers often find it difficult to give a rationale for their grading and assessment practices. Understanding why teachers think they make classroom assessment decisions and how this interacts with what decisions are actually being made will contribute to developing and modifying assessment principles, which may ultimately improve instruction and student learning. Classroom assessment could benefit from a more rigorous application of measurement concepts (e.g., the gathering, interpreting, and evaluating of evidence based on the meaning of results) to help teachers align instruction with the related cognitive processes that facilitate student learning.

References

Alberta Teachers' Association. (2003). *Class size and its effect on learning*. Retrieved October 25, 2004, from:http://www.teachers.ab.ca/publications/reports/size.html

Duncan, C.R., & Noonan, B. (in press). Is there evidence of assessment reform in teachers' classroom practices? *Brock Education: A Journal of General Inquiry.*

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86*, 8-21.

Bol, L., Stephenson, P.L., O'Connell, A.A., & Nunnery, J.A. (1998). Influence of experience, grade level, and subject area on teachers' assessment practices. *Journal of Educational Research, 91*, 323-330.

Boston, C. (2002). The concept of formative assessment. *Practical Assessment Research and Evaluation.* Retrieved March 23, 2005, from: http://PAREonline.net/getvn.asp?v=8&n=9.

Brookhart, S.M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123-142.

Brookhart, S.M. (2003). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*(4), 5-25.

Cizek, G.J., Fitzgerald, S.M., & Racher, R.E. (1996). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment, 3*, 159-179.

Frary, R.B., Cross, L.H., & Weber, L.J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction in measurement. *Educational Measurement: Issues and Practice, 12*(3), 23-30.

Gullickson, A.R. (1985). Student evaluation techniques and their relationship to grade and curriculum. *Journal of Educational Research, 79*(2), 96-100.

Gusky, T.R. (1994). Making the grade: What benefits students? *Educational Leadership, 14-19.*

Gusky, T.R. (2003). How classroom assessments improve learning. *Educational Leadership, 60*(5), 6-11.

Hargreaves, A., Earl, L., & Schmidt, M. (2002). Perspectives on alternative assessment reform. *American Educational Research Journal, 39*(1), 69-95.

Locastro, V. (2001). Teaching English to large classes. *TESOL Quarterly, 35*(1), 493-496.

McMillan, J.H. (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice, 20*(1), 20-32.

McMillan, J.H. (2003). Understanding and improving teachers' classroom assessment decision-making: Implications for theory and practice. *Educational Measurement: Issues and Practice, 22*(4), 34-43.
Monk, D.H., & Haller, E.J. (1993). Predictors of high school academic course offerings: The rule of school size. *American Educational Research Journal, 30*(1), 3-21.
Pilcher, J.K. (1994). The value-driven meaning of grades. *Educational Assessment, 2*(1), 69-88.
Reynolds, A., Reagin, M., & Reinshuttle, K. (2001). Less is more: What teachers say about decreasing class size and increasing learning. *American School Board Journal, 199*(9), 30-32.
Rolheiser, C., & Ross, J. (2000). Student self-evaluation—What do we know? *Orbit, 30*(4), 33-36.
Shepard, L.A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4-14.
Stiggins, R.J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice, 18*(1), 23-27.
Stiggins, R.J. (2001). The unfulfilled promise of classroom assessment. *Educational Measurement: Issues and Practice, 20*(3), 5-15.
Stiggins, R.J. (2002). Assessment crisis: The absence of assessment for learning. *Phi Delta Kappan, 83*, 758-765.
Stiggins, R.J., & Conklin, N.F. (1992). *In teacher's hands: Investigating the practices of classroom assessment.* Albany, NY: SUNY Press.
Taylor, C. (1994). Assessment for measurement or standards: The peril and promise of large-scale assessment reform. *American Journal of Educational Research, 31*, 231-262.
Taylor, A.R., & Tubianosa, T. (2001). *Student assessment in Canada: Improving the learning environment through effective education*. Kelowna, BC: Society for the Advancement of Excellence in Education.
Welsh, C.L. (1989). *The relationship of school size to student achievement*. Unpublished doctoral dissertation, University of Kansas.

Appendix: Assessment Practices of High School Teachers

Thank you for taking the time to complete this short questionnaire which is based on assessment practices for one of the courses you teach. In order to ensure anonymity, please do not put your name or the name of your school on the survey. By completing the questionnaire it is understood that you do so voluntarily and that you consent to the use of your responses in the study. The survey consists of three parts:

Part 1 background information

Part 2 34 selected response questions. Please clearly indicate your response by shading the bubble completely. Please do not use checkmarks (✓) or an (x).

Example: In responding to this survey, I am doing so for (select one):

Math ①	Science ②	Soc. Sc. ③	Fine Arts ④

Part 3 your open ended questions

PART 1 **Instructions:** *In responding to the following questions, please do so thinking about any course you are currently teaching, or have taught recently, for example Math 30, ELA 30-1, etc.*

1. In responding to this questionnaire, I am doing so for (select one):

Grade	9	①
Grade	10	②
Grade	11	③
Grade	12	④

2. In responding to this questionnaire, I am doing so for Subject/Course (select one):

Math	①
Science	②
English LA	③
Soc. Sci.	④
Fine Arts	⑤
Other	⑥

3. Number of students in the class

less than 15	①
15-25	②
26-35	③
more than 35	④

4. How many times have you taught this course?

Once	①
Twice	②
Three times or more	③

5. Number of high school (Gr. 9-12) students in your school:

less than 150	①
151-250	②
251-400	③
401-600	④
more than 600	⑤

6. Years of Teaching Experience:

2 years or less	①
3 to 10 years	②
more than 10 years	③

7. Have you taken any university courses in classroom assessment (i.e. measurement and evaluation)?

Yes	①
No	②

PART 2 **Please shade the bubble for the response that best matches your practices, based on the grade and course you identified in Part 1.**

		Not at all	Very Little	Some	Quite a Bit	Extensively	Completely
1.	*Factors you use in determining grades*						
1	Performance compared to a set of pre-assigned criteria	①	②	③	④	⑤	⑥
2	Specific learning objectives mastered	①	②	③	④	⑤	⑥
3	Academic performance as expressed in other tests	①	②	③	④	⑤	⑥
4	Inclusion of zeros for unacceptable assignments in the determination of final percentage earned	①	②	③	④	⑤	⑥
5	Ability levels of the students	①	②	③	④	⑤	⑥
6	Student effort - how much the student tried to learn	①	②	③	④	⑤	⑥
7	Quality of completed homework/ Goals/	①	②	③	④	⑤	⑥

	Not at all	Very Little	Some	Quite a Bit	Extensively	Completely
8. Degree to which the student pays attention and/or participates in class	①	②	③	④	⑤	⑥
9. Completion of homework (not graded)	①	②	③	④	⑤	⑥
10. Effort, improvement, behavior and other "academic" indicators for borderline cases	①	②	③	④	⑤	⑥
11. Improved performance since the beginning of the year	①	②	③	④	⑤	⑥
12. Work habits and neatness	①	②	③	④	⑤	⑥
13. Extra credit for academic performance	①	②	③	④	⑤	⑥
14. Performance compared to other students in the class	①	②	③	④	⑤	⑥
15. Disruptive student performance	①	②	③	④	⑤	⑥
16. Extra credit for nonacademic performance (e.g., bringing items for the food drive)	①	②	③	④	⑤	⑥
17. Formal or informal school or division policy of the percentage of students who may obtain A's, B's, C's, D's, and F's	①	②	③	④	⑤	⑥
18. Performance compared to students from previous years	①	②	③	④	⑤	⑥
19. Grade distributions of other teachers	①	②	③	④	⑤	⑥
B. Types of assessments you use						
20. Assessments designed primarily by yourself	①	②	③	④	⑤	⑥
21. Performance Quizzes	①	②	③	④	⑤	⑥
22. Objective assessments (e.g. Multiple choice, matching, short answer)	①	②	③	④	⑤	⑥
23. Essay-type questions	①	②	③	④	⑤	⑥
24. Performance assessments (e.g. Structured teacher observations or ratings of performance such as a speech or paper)	①	②	③	④	⑤	⑥
25. Projects completed by individual students	①	②	③	④	⑤	⑥
26. Major exams	①	②	③	④	⑤	⑥
27. Authentic assessments (e.g. "real world" performance tasks)	①	②	③	④	⑤	⑥
28. Projects completed by teams of students	①	②	③	④	⑤	⑥
29. Assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)	①	②	③	④	⑤	⑥
30. Oral presentations	①	②	③	④	⑤	⑥

		Not at all	Very Little	Some	Quite a Bit	Extensively	Completely
f.	*Cognitive level of assessment*						
31.	Assessments that measure student understanding	①	②	③	④	⑤	⑥
32.	Assessments that measure how well students apply what they learn	①	②	③	④	⑤	⑥
33.	Assessments that measure student reasoning	①	②	③	④	⑤	⑥
34.	Assessments that measure student recall knowledge	①	②	③	④	⑤	⑥

PART 3 *Please respond to each of the open ended questions based on the grade and course you identified in Part 1.*

1. a) In your view what are the most positive aspects of your current classroom assessment practices?

   b) What do you see as problems or concerns?

2. Recent *Student Assessment Initiative Reports* have found that the academic achievement of girls exceeds boys for most if not all subjects in high school. To what extent do you find this to be true based on the grade and course you identified in Part 1?

3. To what extent do you use peer-assessment or self-assessment for other types of assessments not included in the survey) as part of grading student achievement?

4. General comments or suggestions: (Use reverse side if needed)

Thank you for taking the time to complete this questionnaire.