Jurnal Pendidikan Malaysia 36(2)(2011): 9-17

Self-Perceived Assessment Skills of Pre-service and In-service Teachers

(Kemahiran Pentaksiran Berdasarkan Pengamatan Kendiri Guru Pra Perkhidmatan dan Guru dalam Perkhidmatan)

HUSSAIN ALKHARUSI

ABSTRACT

One of the main responsibilities of the teachers is assessing student learning. As such, without adequate skills in educational assessment, it is unlikely that teachers' assessment practices will yield desired outcomes of student learning. Using a descriptive survey research design that is cross-sectional in nature, the present study investigated differences between pre-service and in-service teachers' self-perceived assessment skills as a function of gender. Using a convenience sampling method, participants were 180 pre-service teachers and 150 in-service teachers from Oman. Using principal component analyses, a 47-item Self-Perceived Assessment Skills Questionnaire was developed and used in the study. Using a 2 × 2 multivariate analysis of variance, results indicated significant gender differences between pre-service and in-service teachers in the self-perceived assessment skills in analyzing test results, constructing and administering tests, communicating assessment results, using performance assessment, and grading. Implications for professional preparation of teachers in educational assessment as well as recommendations for future research are discussed.

Keywords: Pre-service teachers, in-service teachers, self-perceived assessment, skills, gender

ABSTRAK

Satu daripada tugas utama guru-guru adalah menaksir pembelajaran pelajar murid-murid mereka. Tanpa kemahiran pentaksiran yang mencukupi, amalan pentaksiran guru tidak akan dapat memberikan hasilan yang sepatutnya berkaitan pembelajaran murid. Dengan menggunakan reka bentuk kajian tinjauan deskriptif irisan lintang, kajian ini meneliti perbezaan pengamatan kendiri tentang kemahiran pentaksiran antara guru pra perkhidmatan dengan guru dalam perkhidmatan dengan berfungsikan aspek gender. Peserta kajian terdiri daripada 180 guru pra perkhidmatan dan 180 guru dalam perkhidmatan di Oman yang dipilih secara rawak mudah. Instrumen kajian merupakan soal selidik pengamatan kendiri berkaitan kemahiran pentaksiran yang dibina berdasarkan analisis prinsipal komponen. Dapatan kajian yang dianalisis menggunakan analisis varian multivariat menunjukkan perbezaan signifikan berdasarkan gender antara guru pra perkhidmatan dengan guru dalam perkhidmatan untuk aspek analisis keputusan ujian, pembinaan dan pentadbiran ujian, penyampaian keputusan ujian, penggunaan pentaksiran berasaskan pencapaian, dan pemarkahan. Implikasi terhadap penyediaan guru-guru secara profesional dalam pentaksiran pendidikan dan cadangan untuk kajian lanjut dibincangkan.

Kata kunci: Guru pra perkhidmatan, guru dalam perkhidmatan, pentaksiran berdasarkan pengamatan kendiri, kemahiran, gender

INTRODUCTION

Classroom assessment aims at improving student learning and motivation to learn (Gronlund 2006). It has become a tool for improving classroom teaching and learning (Shavelson et al. 2008). In this regard, Gronlund (2006) suggests that a sound classroom assessment requires a clear conception of all intended learning outcomes of the instruction and a variety of assessment procedures that are relevant to the instruction, adequately sample student performance, and fair to everyone. In addition, a sound assessment requires the specifications of criteria for judging successful performance and timely and detailed feedback to students emphasizing strengths of their performance and weaknesses to be corrected (Gronlund 2006).

Teachers are required to develop classroom assessment that aligns with practices recommended by experts of educational assessment. For example, assessment experts have recommended that students should clearly be informed about the grading procedure in advance and involved in the assessment process (Stiggins & Chapuis 2005; Stiggins, Frisbie & Griswold 1989); student personal characteristics such as ability, effort, motivation, interest, and neatness of work should not be incorporated into grading due to the lack of objective measurement (Stiggins et al. 1989); a final grade for borderline cases should be determined using additional academic achievement data rather than nonachievement data (Stiggins et al. 1989); and students should be given continuous and informative assessment

feedback rather than judgmental feedback about their academic performance (Brookhart 1994).

Moreover, teachers are encouraged to use more than one assessment method in order to have enough, accurate evidence of student learning (Nitko 2001; Stiggins et al. 1989). It is emphasized that the assessment should match the learning targets and provides meaningful feedback to students (Nitko 2001). In addition, the American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA) (1990) have jointly defined seven *Standards for Teacher Competence in Educational Assessment of Students*. The standards emphasized that teachers should competently be able to choose and develop assessment methods appropriate for instructional decisions; administer, score, and interpret results of externally produced and teacher-made assessment; use assessment results when making educational decisions; develop valid assessment-based grading procedures; communicate assessment results; and recognize unethical, illegal, and inappropriate methods and uses of assessment (AFT, NCME & NEA 1990).

Unfortunately, findings from past and recent studies of classroom assessment practices have consistently expressed a concern about the adequacy of teachers' skills in assessment. For example, in an earlier survey of statistical analyses of test results for 336 elementary and secondary school teachers, Gullickson (1982) found that a substantial proportion of teachers reported using relatively little statistical information such as means, medians, and standard deviations to describe assessment results. Also, these same teachers did not have an adequate understanding of basic testing concepts such as item difficulty and reliability. Parallel to Gullickson's (1982) study, Mertler (1998 1999) found in two studies of 625 K-12 Ohio state teachers that teachers did not spend much time conducting statistical analyses of the assessment data with no significant differences based on teacher's gender and years of teaching experience. Further, Hills (1991) identified four misuses of classroom assessment in schools including using grades for controlling students' behavior, assigning grades that are contingent on improvement, using tests that are technically inadequate, and deviation from established standardized-test administration procedures.

In a review of literature on teachers' grading practices, Brookhart (1994) located 19 studies that were done since 1985. Seven studies focused on grading practices of secondary school teachers, 11 studies investigated both elementary and secondary school teachers, and one study included only elementary school teachers. Research methods employed in those studies included surveys in which teachers were asked about the components incorporated in term grades, grade distributions, and their beliefs about grading issues and grading scenarios; and observations, interviews, and document analyses. Based on this review, Brookhart (1994) concluded that teachers grading practices deviate from the recommendations of educational assessment experts. In a related survey research study of grading practices for 91 middle/high school science teachers, Feldman, Alibrandi and Kropf (1998) found that teachers primarily used traditional forms of assessment to determine report card grades. Student's work habits, class attendance, and behavior were reported as being rarely used. No statistically significant differences were found by teacher's gender, years of teaching experience, and school's geographic location on forms of assessment used in grading. Feldman et al. (1998) concluded that reform efforts on classroom assessment had little or no effect on the practices of the participating middle/high school science teachers.

Like Feldman et al. (1998) and Alsarimi (2000) found that science teachers indicated using four main sources of information when assigning grades to students in his survey study of 246 third preparatory science teachers from 112 schools in Oman. These sources were final exams, midterm exams, class participation, and oral questioning. Also, these same teachers tended to incorporate some nonachievement factors such as student's effort in grading. The teachers commented that the grades reflect student improvement, effort, and knowledge of the subject matter. Further, Zhang and Burry-Stock (2003) surveyed 297 teachers across teaching levels and content areas about their assessment skills. Zhang and Burry-Stock (2003) found that teachers with training in educational assessment tended to report higher levels of self-perceived assessment skills in performance assessment, standardized testing, test revision and communicating assessment results.

In a survey of assessment skills of 69 pre-service teachers, Volante and Fazio (2007) found that the self-described levels of assessment skills remained relatively low for the pre-service teachers across the four years of the teacher education program, thereby suggesting the need for in-service assessment training to ensure an acceptable level of assessment skills. Along similar lines, Wolfe, Viger, Jarvinen and Linkman (2007) proposed that teachers' self-perceived confidence in assessment should be a vital component in the professional development of in-service teachers. Further, Alkharusi (2009) found that assessment knowledge of pre-service teachers tended to vary as a function of gender. Specifically, in a survey of 211 pre-service teachers, Alkharusi (2009) found that males tended to have on average a higher level of knowledge in educational assessment than females. In a two-week classroom assessment workshop for seven in-service teachers, Mertler (2009) pre- and post-tested teachers' assessment skills. The results showed that the assessment training had a positive impact on teachers' assessment skills as well as on their feelings regarding assessment and confidence in using assessment.

When comparing assessment skills of 62 pre-service teachers and 71 in-service teachers, Johns and Davis (1991) found that both groups need a better understanding of standardized testing procedures and ethical considerations in test-wiseness. Green (1992) examined differences in

opinions about testing between 236 pre-service teachers and 553 in-service teachers. Green (1992) found that pre-service teachers had less favorable attitudes toward classroom testing and more favorable attitudes toward standardized testing than in-service teachers, thereby drawing teacher educators' attention toward the relevance of assessment training to the realities of classroom assessment. Likewise, in a study of 67 pre-service teachers and 10 in-service teachers, Mertler (2004) found that pre-service teachers demonstrated lower levels of assessment literacy in choosing and developing appropriate assessment methods; administering, scoring, and interpreting assessment results; and recognizing unethical assessment practices. Recently, Alkharusi, Kazem, and Al-Musawai (2011) compared assessment skills of 279 pre-service teachers and 233 in-service teachers from Oman. Results indicated that in-service teachers tended to have a higher level of perceived skillfulness in educational assessment than pre-service teachers, thereby testifying the value of including teaching experience when preparing teachers in educational assessment. In conclusion, the classroom assessment literature reveals some contradictions between teachers' practices and recommendations of educational assessment experts regarding issues of classroom assessment and grading, thereby necessitating the need for a better understanding of teachers' skills in educational assessment.

Research Problem and Purpose of the Study

The research problem being addressed in this study stemmed from the notion that a substantial proportion of teachers' professional time is devoted to activities related to the educational assessment of student learning. Unfortunately, findings from the aforementioned studies investigating assessment practices of classroom teachers (Alkharusi et al. 2010; Hills 1991; Mertler 1998 1999) have consistently revealed a concern about the adequacy of teachers' self-perceived assessment skills. As might be expected, this situation might threaten the educational assessment process of student learning. Therefore, the purpose of this study is to examine differences between pre-service and in-service teachers' self-perceived skills in assessment.

Significance of the Study

Classroom assessment refers to the process used in the classroom by the teacher to obtain information about students' performances on assessment tasks to determine the extent to which students are achieving the target instructional outcomes (Gronlund 2006). It involves various activities including, but are not limited to, developing assessment methods such as paper-pencil tests and performance measures; administering, scoring, and interpreting assessment results; developing grading procedures; communicating assessment results; and using them in making educational decisions (AFT, NCME & NEA 1990). Over the past years, considerable amount of qualitative and quantitative research (Greenstein 2004; Mertler 1998, 1999; 2004) have provided empirical evidence indicating the inadequacies of teachers' assessment practices, which in turn might threaten the quality of student learning. Therefore, an investigation of teachers' skills in classroom assessment is needed. It was expected that the study would provide some insights to improve the assessment skills of the teachers. The study was also expected to reveal some implications for teacher educators who are involved in the professional preparation of teachers in educational assessment.

Methodology

Research Design

This study employed a descriptive survey research design that is cross-sectional in nature to compare male and female pre-service and in-service teachers with regard to their self-perceived assessment skills. Due to time and cost constraints as well as unavailability of a complete list of the population units, a convenience sampling method was utilized in the selection of the pre-service and in-service teachers. A survey in terms of a questionnaire was developed to collect data from the participants. A 2×2 multivariate analysis of variance was used to analyze the data using gender (male versus female) and teaching status (pre-service versus in-service) as independent variables and self-perceived assessment skills as dependent variables. The following sections provide more details about the participants, data collection process, instrumentation, and analyses.

Participants

The participants in this study were 180 pre-service teachers (100 males and 80 females) and 150 in-service teachers (60 males and 90 females) teaching grades 5 to 10 in Oman. The participants were selected using a convenience sampling method due to time and cost constraints as well as unavailability of a complete list of the population units. The majority of the pre-service teachers (80%) were of English language major whereas the rest were of science and math education majors. Their ages ranged from 20 to 26 with an average of 21 years a standard deviation of 1.28. The teaching experience of the in-service teachers ranged from 1 to 15 years with an average of 6 years and a standard deviation of 3. The majority of them (74.67%) were teaching math and science subjects whereas the rest were teaching different subject areas such as Arabic language (10%), English language (8.67%), and social studies (6.67%).

DATA COLLECTION PROCESS

Permission was requested from course instructors, who were teaching educational assessment course for pre-service teachers, to collect data from pre-service teachers and from school principals to collect data from in-service teachers. The participants were informed that a study is being conducted to investigate differences between pre-service and in-service teachers' perceptions about their skills in educational assessment. The teachers were also informed that they were not obligated to participate in the study, and that if they wished, their responses would remain anonymous and confidential. Those who wished to participate in the study were provided a cover letter and a questionnaire along with brief instructions about the information that was requested in the questionnaire, how to respond to the items, and where to find directions that were also included both on the cover letter and the questionnaire.

INSTRUMENTATION

Informed by the literature (Alkharusi 2009; Zhang & Burry-Stock 2003), a 50-item Self-Perceived Assessment Skills Questionnaire was developed and used in this study. The items were initially grouped into three domains identified by examining the content of four textbooks in educational assessment (Gallagher 1998; Gronlund 2006; Nitko 2006; Poham 2000). These domains were test construction and administration (20 items), test and item analysis (15 items), and performance assessment (15 items). The participants were asked to indicate how skilled they are in using the assessment issue described by the item on a 5-point Likert scale ranging from 1 (not at all skilled) to 5 (very skilled). To establish content validity, the items were given to a group of faculty members in the areas of educational measurement and psychology from Sultan Qaboos University. They were asked to judge the clarity of wording and the appropriateness of each item and its relevance to the construct being measured. Their feedback was used for further refinement of the items.

The participants' responses were submitted to principal components analyses to identify their underlying dimensions. No particular number of dimensions was hypothesized and the criterion was set to eigenvalues greater than one (Tabachnick & Fidell 2001). The initial unrotated principal components analyses resulted in a factor model of seven dimensions as suggested by the scree plot and eigenvalues exceeding unity. However, based on its pattern of factor loadings, this unrotated factor model was theoretically less meaningful and difficult to interpret. Therefore, the analysis proceeded to rotate the factor matrix orthogonally with Varimax rotation to achieve a simple and theoretically more meaningful solution. During this analysis, three items were deleted because they loaded highly on multiple factors.

As suggested by the eigenvalue rule and scree plot, the analysis yielded a five-factor structure accounting for 56.25% of the total variance. These factors were Analyzing test results (10 items), Constructing and Administering Tests (10 items), Communicating Assessment Results (9 items), Using Performance Assessment (9 items), and Grading (9 items). Table 1 presents the items, the factor loadings, the percent of the variance explained, and Cronbach alpha reliability for each factor. Based on the principal components analyses, five scales were formed for the self-perceived assessment skills. Each scale score was created by averaging the individual scores of all items loading on a factor. A high scale score reflected a high level of self-perceived assessment skill described by the scale.

TABLE 1. Factor structure of the self-perceived assessment skills

	Items	Factor loadings				
		F1	F2	F3	F4	F5
1.	Calculating and interpreting central tendency measures of the test scores.	.86				
2.	Calculating and interpreting variability measures of the test scores.	.80				
3.	Verifying content validity of the test items.	.78				
4.	Establishing reliability of the test scores.	.75				
5.	Conducting item analysis (i.e., difficulty and discrimination) for the test.	.69				
6.	Revising a test based on item analysis.	.67				
7.	Developing a frequency distribution of test scores.	.66				
8.	Using a software program (e.g., SPSS) to analyze test results.	.66				
9.	Evaluating the effectiveness of the distructors in multiple-choice items.	.59				
10.	Calculating a correlation coefficient between two sets of scores.	.55				
11.	Writing multiple-choice questions.		.73			
12.	Writing matching questions.		.72			
13.	Writing true-false questions.		.69			
14.	Writing fill-in-the-blank and short-answer questions.		.67			
15.	Writing essay questions.		.65			

continued

TABLE 1. (*continued*)

Items	Factor loadings				
	F1	F2	F3	F4	F5
16. Using a table of specifications when constructing a test.		.62			
17. Writing test items to assess higher cognitive levels.		.60			
18. Constructing a model answer for scoring essay questions.		.59			
19. Following required procedures (e.g., time limit, no hints…etc.) when administering a test.		.58			
20. Matching test items to purposes of instruction.		.52			
21. Providing oral and written feedback to students.			.78		
22. Communicating assessment results to students.			.68		
23. Communicating assessment results to parents.			.68		
24. Communicating assessment results to other educators.			.67		
25. Avoiding teaching to the test when preparing students for tests.			.64		
26. Protecting students' confidentiality with regard to assessment results.			.61		
27. Prepare for a parent-teacher meeting regarding a student's academic progress.			.60		
28. Using assessment results when making decisions (e.g., placement, promotion) about individual students.			.58		
29. Using assessment results when evaluating class improvement.			.53		
30. Assessing students through observations.				.66	
31. Evaluating oral questions from students.				.65	
32. Defining a rating scale for performance assessment criteria in advance.				.63	
33. Assessing individual and group hands-on activities.				.63	
34. Matching performance tasks to course instruction and objectives.				.56	
35. Communicating assessment criteria to students in advance.				.49	
36. Using portfolios to assess student progress.				.49	
37. Developing performance assessment tasks based on clearly defined course objectives.				.47	
38. Recording assessment result on the rating scale/checklist while observing a student's performance.				.40	
39. Informing students in advance how grades are to be assigned.					.69
40. Weighing differently projects, exams, homework, etc. when assigning semester grades.					.64
41. Using systemic grading procedures to determine borderline grades.					.58
42. Incorporating non-achievement factors (e.g., effort, classroom behavior, attendance, etc.) in the calculation of grades.					.55
43. Developing a systematic grading procedure.					.48
44. Using the grading policy regulated by Ministry of Education.					.43
45. Incorporating extra credit activities in the calculation of grades.					.38
46. Using a norm-referenced grading model.					.36
47. Using a criterion-referenced grading model.					.35
% of variance explained	14.83	12.65	12.11	10.04	6.63
Reliability	.91	.87	.89	.87	.84

Note. F1 = analyzing test results. F2 = constructing and administering tests. F3 = communicating assessment results. F4 = using performance assessment. F5 = grading

STATISTICAL ANALYSIS

A 2 × 2 between-subjects multivariate analysis of variance (MANOVA) was conducted on the self-perceived skillfulness in analyzing test results, constructing and administering tests, communicating assessment results, using performance assessment, and grading. The independent variables were teaching service (pre-service versus in-service) and gender (male versus female). Prior to the analysis, all variables were examined for accuracy of data entry, missing values, normality, outliers, multicollinearity, and homogeneity of variance-covariance matrices. The data screening process revealed no concern about normality, outliers, multicollinearity, and homogeneity of variance-covariance matrices.

RESULTS AND DISCUSSION

A 2 × 2 multivariate analysis of variance (MANOVA) was performed to compare male and female pre-service and in-service teachers with regard to their self-perceived assessment skills in analyzing test results, constructing and administering tests, communicating assessment

results, using performance assessment, and grading. Table 2 presents means and standard deviations for measures of self-perceived assessment skills as a function of teaching service and gender. Table 3 presents results of the multivariate and univariate analyses of variance for the effects of teaching service and gender on the self-perceived assessment skills. As shown in Table 3, there was a statistically significant multivariate interaction effect between teaching service and gender on the self-perceived assessment skills. Approximately 10% (η^2 = .099) of the variability in the participants' self-perceived assessment skills could be accounted for by the interaction of teaching service and gender. The univariate analyses indicated statistically significant effects for teaching service on the self-perceived assessment skills in using performance assessment (η^2 = .169) and grading (η^2 = .083). As shown in Table 2, in-service teachers reported higher levels of self-perceived assessment skills in using performance assessment and grading than pre-service teachers.

In addition, the univariate analyses indicated statistically significant interaction effects between teaching service and gender on the self-perceived assessment skills in analyzing test results (η^2 = .015), constructing and administering tests (η^2 = .022), and communicating assessment results (η^2 = .076). Figures 1 through 3 display the means for measures of self-perceived assessment skills in analyzing test results, constructing and administering tests, and communicating assessment results as a function of teaching service and gender. As shown in Figure 1, male in-service teachers perceived themselves to be more skillful in analyzing test results than female in-service teachers whereas female pre-service teachers perceived themselves to be more skillful in analyzing test results than male pre-service teachers. According to Figure 2, although males reported a higher level of self-perceived assessment skills in constructing and administering tests than females for both pre-service and in-service teachers, the difference between male and female in-service teachers was smaller than the difference between male and female pre-service teachers. As shown in Figure 3, female in-service teachers perceived themselves to be more skillful in communicating assessment results than male in-service teachers whereas male pre-service teachers perceived themselves to be more skillful in communicating assessment results than female pre-service teachers.

When taking an overall level of the self-perceived skills in assessment, pre-service teachers tended to report a higher level of skills than in-service teachers in analyzing test results, whereas in-service teachers tended to report a higher level of skills in constructing and administering tests, using performance assessments, and grading. These results are in partial agreement

TABLE 2. Means and standard deviations for measures of self-perceived assessment skills as a function of teaching service and gender

Group	N	Measures of self-perceived assessment skills									
		F1		F2		F3		F4		F5	
		M	SD	M	SD	M	SD	M	SD	M	SD
Pre-service teachers											
Males	100	3.20	.98	3.80	.66	4.02	.60	4.01	.64	3.79	.77
Females	80	3.73	.96	3.33	.78	3.43	.86	4.01	.64	3.62	.75
In-service teachers											
Males	60	3.26	.98	4.19	.45	3.88	.47	4.57	.53	4.18	.72
Females	90	2.95	.99	4.08	.31	3.99	.36	4.54	.56	4.15	.76

Note: F1 = analyzing test results. F2 = constructing and administering tests. F3 = communicating assessment results. F4 = using performance assessment. F5 = grading

TABLE 3. Multivariate and univariate analyses of variance for the effects of teaching service and gender on the self-perceived assessment skills

Variable	Multivariate	Univariate				
		F1	F2	F3	F4	F5
	PF(5, 322)	F(1, 326)	F(1, 326)	F(1, 326)	F(1, 326)	F(1, 326)
Teaching service (TS)	22.09***	2.69	74.62***	10.27**	66.08***	29.47***
Gender (G)	5.41***	.36	19.40***	12.08**	1.06	1.38
TS × G	7.07***	5.00*	7.50**	28.85**	1.01	.71

Note: F1 = analyzing test results. F2 = constructing and administering tests. F3 = communicating assessment results. F4 = using performance assessment. F5 = grading

Multivariate F ratios were genserated from Wilks's statistic

*p < .05. **p < .01. ***p < .001

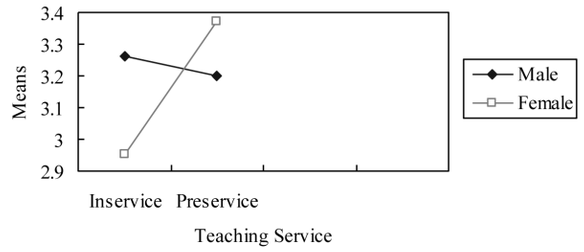

FIGURE 1. Means for self-perceived assessment skills in analyzing test results as a function of teaching service and gender

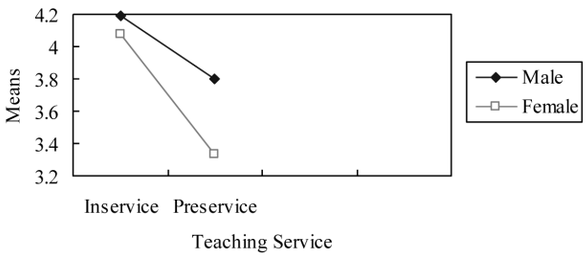

FIGURE 2. Means for self-perceived assessment skills in constructing and administering tests as a function of teaching service and gender

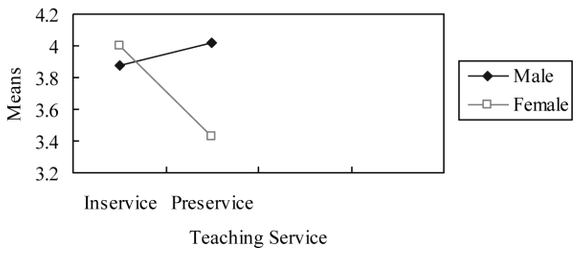

FIGURE 3. Means for self-perceived assessment skills in communicating assessment results as a function of teaching service and gender

with previous research (Merter 2004; Volante & Fazio 2007). Specifically, Mertler (2004) found that in-service teachers were more skillful than pre-service teachers in developing, administering, and scoring assessments as well as in interpreting assessment results. Likewise, Volante and Fazio (2007) found that levels of assessment skills remained relatively low for teacher candidates across the years of the teacher education program.

On one hand, the recent completion of a course in educational assessment could contribute the high level of assessment skills in analyzing test results reported by pre-service teachers. This suggests the need to conduct continuous in-service training programs to compensate for the in-service teachers' low levels of skills in the statistical analysis of test results. On the other hand, the high levels of assessment skills reported by in-service teachers in constructing and administering tests, using performance assessments, and grading highlight the importance of classroom experience. Pre-service teachers might need to practice applying the principles of educational assessment in actual classroom settings (Bangert & Kelting-Gibson 2006). Effective teaching and learning in educational assessment may result from connecting the content materials of the course to the daily and ongoing assessment activities of the classroom and providing the pre-service teachers with real-life opportunities for application of the educational assessment concepts and principles (Criswell & Criswell 1995; Stiggins 1995). In the past research, teachers have indicated that they are more concerned with the daily issues related to their classroom assessment practices and less with the technical principles of assessment (Airasian 1991; Arter 1999; Stiggins 1991). The present study findings imply that it may be important that teacher educators better consider the reality of the classrooms making the pre-service educational assessment course more applicable to the day-to-day work of the classroom teacher.

When considering gender differences in assessment skills, Alkharusi (2009) found that males tended to have on average a higher level of assessment skills than females in a survey of 211 pre-service teachers in Oman. The findings of the present study extend previous research by underpinning the importance of examining gender differences in patterns of assessment skills and type teaching service. The significant interactions between gender and teaching service status found in this study with respect to assessment skills are intriguing and call for further research. Classroom observations and interviews might shed more light on why male pre-service teachers or female in-service teachers are more skillful in certain assessment areas than their counterparts.

Conclusion

This study examined differences between pre-service and in-service teachers' self-perceived assessment skills as a function of gender. The results point to a general conclusion that pre-service teachers perceived themselves to be more skillful in analyzing assessment results than in-service teachers; whereas in-service teachers perceived themselves to be more skillful in developing and administering assessment, using performance assessments, and grading. Although this conclusion varies as a function of gender, it suggests the need for providing pre-service teachers more opportunities to practice assessment principles in classrooms in order to foster their skills in developing and administering assessment, using performance assessments, and grading. Also, the conclusion suggests the need for offering continuous training for in-service teachers in analyzing assessment results.

The research design of this study is cross-sectional. A longitudinal study examining educational assessment skills

of the teachers over time from pre-service to in-service might be required to identify developmental progression. In addition, the gender interaction effect on the self-perceived assessment skills crossing pre-service and in-service teachers found in the present study suggests that unique combinations of teachers' characteristics and assessment practices might yield different outcomes of student learning and motivation. Uncovering these combinations may help lead to improved classroom assessment practices and student outcomes. Classroom observations and interviews might shed some light on gender differences in patterns of teachers' assessment skills and practices. These might validate the results of the self-report questionnaires. The questionnaires themselves may also need to be administrated to a representative sample selected from different geographic regions across the country. Finally, further studies might need to consider the environmental pressures of the school and classroom as they relate to the assessment practices of the teachers.

REFERENCES

Airasian, P.W. 1991. Perspective on measurement instruction. *Educational Measurement: Issues and Practice* 10: 13-16, 26.

Alkharusi, H. 2009. Correlates of Teacher Education Students' Academic Performance in an Educational Measurement Course. *International Journal of Learning* 16: 1-15.

Alkharusi, H., Kazem, A. & Al-Musawai, A. 2010. Traditional versus computer-mediated approaches of teaching educational measurement. *Journal of Instructional Psychology* 37: 99-111.

Alkharusi, H., Kazem, A. M. & Al-Musawi, A. 2011. Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education* 39: 113-123.

Alsarimi, A.M. 2000. Classroom Assessment and Grading Practices in the Sultanate of Oman. Unpublished dissertation, University of Pittsburgh, Pennsylvania.

American Foundation of Teachers, National Council on Measurement in Education & National Education Association. 1990. Standards for teacher competence in educational assessment of students. *Educational Measurement: Issues and Practice* 2: 30-32.

Arter, J. 1999. Teaching about performance assessment. *Educational Measurement: Issues and Practice* 18: 30-44.

Bangert, A. & Kelting-Gibson, L. 2006. Teaching principles of assessment literacy through teacher work sample methodology. *Teacher Education and Practice* 19: 351-364.

Brookhart, S.M. 1994. Teachers' grading: Practice and theory. *Applied Measurement in Education* 7: 279-301.

Criswell, J.R. & Criswell, S.J. 1995. Modeling alternative classroom assessment practices in teacher education coursework. *Journal of Instructional Psychology* 22: 190-193.

Feldman, A., Alibrandi, M. & Kropf, A. 1998. Grading with points: The determination of report card grades by high school science teachers. *School Science and Mathematics* 98: 140-148.

Gallagher, D., Jo. 1998. *Classroom Assessment for Teachers*. New Jersey, NJ: Prentice-Hall.

Green, K.E. 1992. Differing opinions on testing between pre-service and in-service teachers. *Journal of Educational Research* 86: 87-42.

Greenstein, L. 2004. Finding Balance in Classroom Assessment: High School Teachers' Knowledge and Practice. Unpublished doctoral dissertation, Johnson & Wales University, Providence, Rhode Island.

Gronlund, N.E. 2006. *Assessment of Student Achievement.* (8th ed. Boston: Pearson.

Gullickson, A.R. 1982. *The Practice of Testing in Elementary and Secondary Schools*. Vermillion, SD: University of South Dakota. (ERIC Document Reproduction Service No. ED229391)

Hills, J.R. 1991. Apathy concerning grading and testing. *Phi Delta Kappan* 72: 540-545.

Johns, J.L., & Davis, S.J. 1991. *Perceptions of Preservice and Inservice Teachers Regarding Test-taking Procedures and Test-wiseness Programs* (Literacy Research Report No. 4). DeKalb, IL: Northern Illinois University. (ERIC Document Reproduction Service No. ED335327)

Mertler, C.A. 1998. *Classroom Assessment Practices of Ohio Teachers*. Paper presented at the meeting of the Mid-Western Educational Research Association, Chicago.

Mertler, C.A. 1999. *Teachers' (Mis)conceptions of Classroom test Validity and Reliability.* Paper presented at the meeting of the Mid-Western Educational Research Association, Chicago.

Mertler, C.A. 2004. Secondary teachers' assessment literacy: Does classroom experience make a difference. *American Secondary Education* 33: 49-64.

Mertler, C.A. 2009. Teachers' assessment knowledge and their perceptions of the impact of classroom assessment professional development. *Improving Schools* 12: 101-113.

Nitko, A.J. 2001. *Educational Assessment of Students.* 3rd ed. Upper Saddle River, NJ: Prentice-Hall.

Popham, W.J. 2000. *Modern Educational Measurement: Practical Guidelines for Educational Leaders.* 3rd ed. Boston: Allyn & Bacon.

Shavelson, R.J., Young, D.B., Ayala, C.C., Brandon, P.R., Furtak, E.M., Ruiz-Primo, M.A., Tomita, M.K. & Yin, Y. 2008. On the impact of curriculum-embedded formative assessment on learning: A collaboration between curriculum and assessment developers. *Applied Measurement in Education* 21: 295-314.

Stiggins, R.J. 1991. Relevant classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice* 18: 23-27.

Stiggins, R. 1995. Assessment literacy for the 21st century. *Phi Delta Kappan* 77: 238-245.

Stiggins, R. & Chappuis, J. 2005. Using student-involved classroom assessment to close achievement gaps. *Theory Into Practice* 44: 11-18.

Stiggins, R.J. & Conklin, N.F. 1992. *In teachers' Hands: Investigating the Practices of Classroom Assessment.* Albany, NY: State University of New York Press.

Stiggins, R.J., Frisbie, D.A. & Griswold, P.A. 1989. Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice* 8: 5-14.

Tabachnick, B.G. & Fidell, L.S. 2001. *Using Multivariate Statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.

Volante, L. & Fazio, X. 2007. Exploring teacher candidates' assessment literacy: Implications for teacher education reform and professional development. *Canadian Journal of Education* 30: 749-770.

Wolfe, E.W., Viger, S.G., Jarvinen, D.W. & Linksman, J. 2007. Validation of scores from a measure of teachers' efficacy toward standards-aligned classroom assessment. *Educational and Psychological Measurement* 67: 460-474.

Zhang, Z. & Burry-Stock, J.A. 2003. Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education* 16: 323-342.

For more information please contact
Hussain Alkharusi
Department of Psychology
College of Education
Sultan Qaboos University
Sultanate of Oman