

Educational Assessment Profile of Teachers in the Sultanate of Oman

Hussain Alkharusi¹, Said Aldhafri¹, Hilal Alnabhani¹ & Muna Alkalbani²

¹ College of Education, Sultan Qaboos University, Muscat, Sultanate of Oman

² Ministry of Education, Muscat, Sultanate of Oman

Correspondence: Hussain Alkharusi, College of Education, Sultan Qaboos University, P.O. BOX 32, P.C. 123 Alkhod, Sultanate of Oman. Tel: 968-9622-2535. E-mail: hussein5@squ.edu.com

Received: February 14, 2014 Accepted: March 18, 2014 Online Published: April 29, 2014

doi:10.5539/ies.v7n5p116

URL: <http://dx.doi.org/10.5539/ies.v7n5p116>

Abstract

This study builds on a previous pilot study conducted by Alkharusi, Aldhafri, Alnabhani, and Alkalbani (2012) to explore educational assessment attitudes, competence, knowledge, and practices of in-service teachers in the Sultanate of Oman. The present study extends the previous pilot study by surveying a larger sample of in-service teachers teaching grades 5 to 12 in all educational governorates in the Sultanate of Oman as opposed to 165 in-service teachers teaching grades 5 to 10 in one educational governorate. Specifically, the study aimed at developing a profile of educational assessment attitudes, competence, knowledge, and practices for teachers in the Sultanate of Oman. The profile was developed as a function of teachers' gender, nationality, educational governorate, teaching grade, qualification, teaching subject, pre-service assessment training, in-service assessment training, teaching load, and teaching experience. The study employed a descriptive survey research design. Participants were 3557 in-service teachers teaching various subject areas in grades 5 to 12 randomly selected from all educational governorates in the Sultanate of Oman. Confirming Alkharusi et al. (2012) study, findings of the current study showed that the teachers tended to have a positive attitude towards educational assessment. Despite their perception as being competent in educational assessment, they demonstrated a low level of the educational assessment knowledge. Further, the teachers indicated using different classroom assessments mainly for grading and increasing students' desire for learning. Teaching load and teaching experience explained some of the differences in the teachers' educational assessment profile. Also, the educational assessment profile varied as a function of the selected demographic and background variables. The findings pointed to a conclusion that professional educational assessment programs for teachers should be continued and tailored to the needs and nature of the teachers' classroom realities. Future research is needed to judge the validity of the teachers' self-report surveys concerning educational assessment.

Keywords: teachers' attitudes, teachers' competence, teachers' knowledge, teachers' practices, educational assessment

1. Introduction

1.1 Educational Assessment System in the Sultanate of Oman

In the Omani educational system, assessment has traditionally been linked with formal exams; particularly high stakes, promotion, and school-leaving end of year exams. Recently, the Ministry of Education has made several educational reforms including the introduction of continuous assessment system. This system aims at offering teachers the opportunity to make stronger links between teaching, learning, and assessment. Based on this assessment system, teachers are expected to use a variety of assessment methods such as short written or oral tests, quizzes, performance assessment tasks, projects, and student self-assessment. The continuous assessment occurs in all grades in public schools. Students in grades 1–4 are assessed with written tests prepared by their teachers at the end of each textbook unit in each subject. In addition, they are assessed using classroom activities such as oral presentations, written activities, and practical exercises; and non-classroom activities such as research projects and portfolios. Students are promoted to the next grade automatically. However, if a student does not achieve 50 percent on the total subject score, he or she will be enrolled in a remedial program at the end of the school year. If a student still fails, he or she will be enrolled in another remedial program at the beginning of the following year to support his or her learning in the next grade. Students in grades 5–10 are assessed using the same system in place for students in grades 1–4. These students also take short written tests. Students need a

total score of 50 percent in each subject to pass and be promoted to the next grade. However, if a student fails an examination in any given subject, up to a maximum of three subjects, he or she will be allowed to retake the exam at the end of the school year. If the student fails the exam again, he or she must repeat the grade.

Further, the assessment system in Oman has two types of examinations: school and central examinations. The school examinations are made by the teachers in the school. The students in all grade levels take at least two school exams per semester. The central examinations are made and administered by The Tests and Examinations Administration Department in the Ministry of Education for students in grade 12 in order to obtain a General Education Diploma. At the end of each semester, students in Grades 1–11 receive a report card, which includes the scores obtained in each subject and the overall level of performance as well as any promotional comments or observations related to remedial programs from the teachers of all subjects. These report cards give parents official feedback regarding the performance of their kids in the school.

Despite the current developments in the continuous assessment system in the Sultanate of Oman, the classroom assessment practices tend to be connected to the exams. As such, most of the training in educational assessment in the Sultanate of Oman has been devoted to increasing teachers' knowledge and competence for developing, administering, and scoring exams. Accordingly, it is not uncommon finding teachers' deficiencies on other areas of the educational assessment such as developing and using performance assessment and analyzing and using assessments for making educational decisions (Alkharusi, Aldhafri, Alnabhani, & Alkalbani, 2012). In addition, there has been a concern in the country about the low performance of students in standardized international tests such as TIMSS. Given that educational assessment is part of the educational accountability system, it seems reasonable to argue that this situation activates a need to gather information regarding the current educational assessment profile of the teachers in the Sultanate of Oman. The present study aimed at developing a profile of educational assessment attitudes, competence, knowledge, and practices for teachers in the Sultanate of Oman. The profile would be developed as a function of teachers' gender, nationality, educational governorate, teaching grade, qualification, teaching subject, pre-service assessment training, in-service assessment training, teaching load, and teaching experience.

1.2 Teachers' Assessment Competence and Knowledge

Educational assessment is an important aspect of the teaching profession. It refers to the process used in the classroom by the teacher to obtain information about students' performances on assessment tasks, either as a group or individually, using a variety of assessment methods, to determine the extent to which students are achieving the target instructional outcomes (Gronlund, 2006). Information gathered from the educational assessment are used for making various educational decisions including planning classroom instruction, placing students into learning sequences, monitoring students' progress, diagnosing students' learning difficulties, providing students and parents with feedback about achievements, evaluating effectiveness of teaching, and assigning grades (Nitko, 2001). A variety of methods are used by the teachers in their daily classroom assessment including traditional assessments such as multiple-choice, true-false, matching, completion, and short-answer; and alternative assessments such as portfolios, student self-assessment, observations, and other performance-based assessments (Gronlund, 2006). The quality of these assessments and their consequences on teaching and learning depends on teachers' competence and knowledge in the educational assessment (Alkharusi, 2011b, 2011d; Alkharusi, Kazem, & Al-Musawai, 2011).

Along this line, Gronlund (2006) proposes that a well-grounded educational assessment requires a clear articulation of all planned learning outcomes of the instruction and diverse assessment methods that are related to the instruction, adequate to sample student performance, and fair to everyone. Further, the American Federation of Teachers (AFT), the National Council on Measurement in Education (NCME), and the National Education Association (NEA) (1990) have jointly delineated seven *Standards for Teacher Competence in Educational Assessment of Students*. The standards stated 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 grading procedures; communicate assessment results to various audiences; and recognize unethical, illegal, and inappropriate methods and uses of assessment.

In addition, Brookhart (2011) contends that the *Standards for Teacher Competence in Educational Assessment of Students* do not consider up-to date definitions of formative assessment and teacher's assessment knowledge and skills expected for the standards-based assessment systems. Accordingly, she suggested a number of educational assessment knowledge and skills for teachers in relation to the formative assessment and standards-based assessment systems. Based on Brookhart's framework, teachers should understand learning in the content area

they teach, be able to set and apply learning intentions congruent with both the content and depth of the standards and curriculum goals, have strategies for communicating the expectation of the learning intentions to students, understand the purposes and uses of the various types of assessment and be able to use them, be skillful in analyzing assessment methods, be skillful in providing effective meaningful feedback on student work, have the ability to develop scoring schemes to quantify student performance for making informed educational decisions, be skillful in administering external assessments and interpreting their results for decisions making, be able to apply educational decisions made out from classroom assessments, be able to communicate assessment information to students to motivate them to learn, understand the legal and ethical issues in the classroom assessment practices.

1.3 International Studies on Teachers' Assessment

Several studies around the world have examined teachers' knowledge, attitudes, and practices about educational assessment. For instance, in a survey of 555 in-service teachers in the United States, Plake and Impara (1992) developed an instrument titled the "Teacher Assessment Literacy Questionnaire (TALQ)" consisting of 35 items to measure teachers' knowledge in educational assessment based on the *Standards for Teacher Competence in the Educational Assessment* (AFT, NCME, & NEA, 1990). The findings showed that the teachers were not well prepared to assess student learning as indicated by the mean score of 23 out of 35 items correct, and hence teachers' assessment literacy requires more examinations.

In a survey of 200 pre-service and in-service teachers from the southern region of Mexico, Arce-Ferrer, Cab, and Cisneros-Cohernour (2001) investigated teachers' perspectives about the familiarity and importance of assessment practices. Results indicated that the teachers perceived that the most important assessment practices for professional development of teachers in the educational assessment are skills related to choosing assessment methods for instructional decisions, increasing reliability of tests for grading, and communicating assessment results to students. Arce-Ferrer et al. (2001) called for designing assessment training programs addressing teachers' needs of assessment knowledge and skills.

Using a qualitative study, Susuwele-Banda (2005) examined perceptions and practices of six teachers from Malawi about classroom assessment. Results revealed that the teachers perceived classroom assessment as mainly for testing and as such they showed a limited ability to use different assessment methods. Also, teaching experience and teacher education program did not make a difference in teachers' perceptions of classroom assessment. Susuwele-Banda (2005) contended that teacher colleges and Ministry of Education should consider classroom assessment issues more in training programs and that collaboration between teacher colleges and Ministry of Education should be increased to better understand the challenges and reality of the classroom assessment experienced by the teachers.

Ogan-Bekiroglu (2009) employed a paralld mixed-methodology to investigate educational assessment attitudes and competence of 46 teachers in Turkey after completing a course in educational assessment. The author found that the teachers had constructivist views and a high sense of competence about educational assessment. However, they indicated some obstacles related to school policy and facilities negatively affecting their classroom assessment practices with regard to the use of alternative assessments. Ogan-Bekiroglu (2009) argued that teachers' knowledge and attitudes in educational assessment should be considered when making reforms in the educational systems. Results of both studies by Susuwele-Banda (2005) and Ogan-Bekiroglu (2009) implied that teachers' assessment practices might be a combination of many factors including teachers' personal perceptions and characteristics of the school context.

In a study of educational assessment literacy, DeLuca and Klinger (2010) surveyed 288 teacher candidates enrolled in a teacher education program in Canada and found that teacher candidates who were enrolled in an educational assessment course had higher levels of confidence in educational assessment literacy than those who did not have formal instruction in assessment. These findings suggest the importance of pre-service teacher training in educational assessment for teachers' confidence in performing their classroom assessment responsibilities. Koloi-Keaititse (2012) surveyed 691 primary and secondary school teachers in Botswana about their classroom assessment practices. Results indicated factors related to teachers' educational level, teaching experience, and assessment training contributed positively to beliefs, skills, and uses of desirable classroom assessment practices. Koloi-Keaititse recommended increasing assessment training programs for both pre-service and in-service teachers.

1.4 Omani Studies on Teachers' Assessment

A number of studies have been conducted to examine educational assessment attitudes, competence, knowledge, and practices in the Sultanate of Oman. For example, in an investigation of classroom assessment practices of

246 third preparatory science teachers from 112 schools in Oman, Alsarimi (2000) found that teachers indicated using short answer, completion, oral exams, extended answer, and multiple-choice item formats with no significant differences based on teacher's gender and years of teaching experience. Likewise, in a study of assessment knowledge, skills, and attitudes of 217 in-service teachers in Oman, Alkharusi et al. (2011) found that teachers who had a pre-service course in educational assessment demonstrated on average a higher level of educational assessment knowledge than those who did not have a pre-service assessment course.

Alkharusi (2011c) examined self-perceived assessment skills of 213 Omani teachers. He found that female teachers perceived themselves more skillful than male teachers in writing test items and communicating assessment results. Also, science teachers perceived themselves more skillful than English language teachers and fine arts teachers in developing performance assessment and analyzing assessment results. Further, sixth grade teachers indicated higher levels of self-perceived skills in developing performance assessment than eighth and tenth grade teachers. Furthermore, teaching experience correlated positively with self-perceived assessment skills, and that teachers with in-service assessment training showed a higher level of assessment skills than those without in-service assessment training. Moreover, in an investigation of 516 in-service teachers, Alkharusi (2011a) found that in-service assessment training and teaching experience correlated positively with educational assessment knowledge. Similarly, when examining educational assessment knowledge of 259 pre-service teachers who completed an educational assessment course, Alkharusi (2011b) found that male teachers tended to have on average a higher level of educational assessment knowledge than female teachers. Recently, Alkharusi et al. (2012) surveyed 165 in-service teachers from Muscat governorate about their attitudes, competence, knowledge, and practices in educational assessment. They found that although teachers held a favorable attitude towards and perceived themselves as being competent in educational assessment, they demonstrated a low level of knowledge in educational assessment. Teachers used a variety of assessments in the classroom primarily for assigning grades and motivating students to learn, with some variations by gender, grade level, and subject area. Teaching load and teaching experience accounted for some of the variations in teachers' educational assessment practices.

It seems that results of the aforementioned studies in the Sultanate of Oman did not differ to some extent from those around the world. They generally point to a conclusion that classroom assessment might be unique from teacher to another depending on gender, teaching experience, teaching grade, qualification, and assessment training. Despite the availability of studies about educational assessment of teachers in the Sultanate of Oman, findings from these studies might be limited in their generalizability to all teachers in the country in terms of educational governorate, teaching subject, and teaching grade. There is a need to have a base line of the educational assessment profile of the teachers in the Sultanate of Oman. The present study aimed at addressing this need.

1.5 Purpose and Research Questions

This study aimed at developing a profile of educational assessment attitudes, competence, knowledge, and practices for teachers in the Sultanate of Oman. The study employed a descriptive survey research design. It was guided by the following general research questions:

- 1) What is the current state of educational assessment attitudes, competence, knowledge, and practices of teachers in the Sultanate of Oman?
- 2) How do teachers' gender, nationality, educational governorate, teaching grade, qualification, teaching subject, pre-service assessment training, in-service assessment training, teaching load, and teaching experience relate to their educational assessment attitudes, competence, knowledge, and practices?

2. Methods

2.1 Participants

The participants in this study were 3557 teachers teaching grades (5-12) randomly selected from all educational governorates in the Sultanate of Oman. Table 1 shows the distribution of the participants by gender, nationality, educational governorates, teaching grade level, educational qualification, and teaching subject area.

Table 1. Distribution of the participants by gender, nationality, educational governorates, teaching grade level, educational qualification, and teaching subject area (N = 3557)

Variable		<i>f</i>	%
Gender	Male	1797	50.5
	Female	1760	49.5
Nationality	Omani	3279	92.2
	Non-Omani	278	7.8
Governorate	Muscat	416	11.7
	Al-Batinah North	667	18.8
	Al-Batinah South	546	15.4
	Al-Sharqiyah North	464	13.0
	Al-Sharqiyah South	340	9.6
	Al-Dakhiliyah	417	11.7
	Dhofar	242	6.8
	Al-Wosta	39	1.1
	Al-Dhahira	263	7.4
	Musandam	66	1.9
	Al-Buraimi	97	2.7
	Grade level	Five	191
Six		225	6.3
Seven		516	14.5
Eight		533	15.0
Nine		493	13.9
Ten		695	19.5
Eleven		704	19.8
Twelve		200	5.6
Qualification	Bachelor degree	3146	88.4
	Above bachelor degree	411	11.6
Subject area	Islamic education	433	12.2
	Arabic language	600	16.9
	English language	440	12.4
	Mathematics	595	16.7
	Science	601	16.9
	Social Studies	415	11.7
	Practical-based subjects	473	13.3

The teaching experience of the teachers ranged from 1 to 29 years with an average of 9.69 and a standard deviation of 5.39. The self-reported teaching load of the participants ranged from 4 to 22 classes per week with an average of 15.05 and a standard deviation of 3.85. The majority of the teachers (N = 3140, 88.3%) indicated that they have taken at least one course in educational assessment during their pre-service preparation whereas 417 (11.7%) teachers indicated that they did not take any course in educational assessment during their pre-service preparation. Also, the majority of the teachers (N = 2624, 73.8%) indicated that they did not take any in-service training workshop in the educational assessment whereas 933 (26.2%) teachers indicated that they have taken at least one in-service training workshop in the educational assessment.

2.2 Instrumentation

A self-report questionnaire of seven parts was used in this study. The first part was about background and demographic data of the participants including gender, nationality, governorate, current teaching grade, teaching subject, teaching load, teaching experience, and pre-service and in-service training in the educational assessment. The other six parts were about attitude towards educational assessment, self-perceived competence in educational assessment, knowledge in educational assessment, educational assessment practices, uses of classroom tests, and attitude towards classroom tests. To establish content validity, the questionnaire was given to a group of seven experts in the areas of educational measurement and psychology from Sultan Qaboos University and Ministry of Education. 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 questionnaire.

2.2.1 Attitude towards Educational Assessment

This part of the questionnaire contained 25 items from the Arabic version of the Bryant and Barnes's (1997) Attitude Toward Educational Measurement Inventory (Alkharusi, 2011d). Responses were obtained on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Scoring of the negative items was reversed so that a high score reflected a strong positive attitude towards educational assessment. An individual's attitude towards educational assessment was represented by an average rating score across all the items. Internal consistency reliability coefficient was .89 as measured by Cronbach's alpha.

2.2.2 Self-Perceived Competence in Educational Assessment

This part of the questionnaire contained 44 items from Alkharusi's (2009) Self-Confidence Scale in Educational Measurement designed to assess teachers' perceptions of confidence in their abilities to perform certain educational assessment tasks related to developing and administering assessment methods (11 items); analyzing assessment results (9 items); developing and scoring performance assessment (9 items); developing grading procedures (7 items); and communicating assessment results to various audiences (8 items). Additional seven items related to recognizing ethics of assessment were added by the authors to the questionnaire. Responses were obtained on a 5-point Likert scale ranging from 1 (*very low competence*) to 5 (*very high competence*) with high scores reflecting a high level of competence in educational assessment. An individual's self-perceived competence in each area of the educational assessment was represented by an average rating score across all the items in that area. Internal consistency reliability coefficients for the subscale scores were .78 for developing and administering assessment methods; .82 for analyzing assessment results; .73 for developing and scoring performance assessment; .73 for developing grading procedures; .70 for communicating assessment results to various audiences; and .54 for recognizing ethics of assessment.

2.2.3 Knowledge in Educational Assessment

This part of the questionnaire consisted of 28 items from the Arabic version of the Plake and Impara's (1992) Teacher Assessment Literacy Questionnaire (Alkharusi et al., 2011). It assesses teachers' knowledge and understanding of the basic principles of the educational assessment practices, terminology, development, and use of various classroom assessment methods. All items followed a multiple-choice format with four options, one being the correct answer. The KR20 reliability coefficient for the scores was .45. The average item difficulty was .43 and the average item discrimination as measured by item-total correlation was .11.

2.2.4 Educational Assessment Practices

This part of the questionnaire contained 39 items from Alkharusi's (2010) Teachers' Assessment Practices Questionnaire designed to assess teachers' frequent use of various assessment practices related to traditional assessment methods (6 items); alternative assessment methods (5 items); analysis of assessment results (6 items); assessment communication (7 items); assessment standards and criteria (5 items); student-involved assessment (4 items); and non-achievement grading factors (6 items). Responses were obtained on a 5-point Likert scale ranging from 1 (*never*) to 5 (*all of the time*) with high scores reflecting more frequent use of the assessment described in the item. An individual's frequent use of the assessment practice in a particular area was represented by an average rating score across all the items in that area. Internal consistency reliability coefficients as measured by Cronbach's alpha were .67 for traditional assessment methods; .43 for alternative assessment methods; .77 for analysis of assessment results; .66 for assessment communication; .42 for assessment standards and criteria; .55 for student-involved assessment; and .69 for non-achievement grading factors.

2.2.5 Uses of Classroom Tests

Informed by the educational assessment literature (Gallagher, 1998; Gronlund, 2006; Nitko, 2001), the teachers

were asked to indicate the extent to which they use results obtained from classroom tests in addressing 10 different areas of instructional decisions: diagnose student weakness, group students for instructional purposes, plan for instruction, assign grades, evaluate instruction, control student behavior, motivate students for learning, evaluate academic achievement, compare student performances with others, upgrade students from one grade to another. Responses were obtained on a 5-point Likert scale ranging from 1 (*never*) to 5 (*all of the time*). Internal consistency reliability coefficient was .77 as measured by Cronbach's alpha.

2.2.6 Attitude towards Classroom Tests

Informed by the literature (Green, 1992; Green & Stager, 1987), six positively worded items and four negatively worded items were used to measure teachers' attitude towards classroom tests. Responses were obtained on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Scoring of the negative items was reversed so that a high score reflected a strong positive attitude towards classroom tests. An individual's attitude towards classroom tests was represented by an average rating score across all the items. Internal consistency reliability coefficient was .79 as measured by Cronbach's alpha.

2.3 Procedures

Permission was requested from Ministry of Education and school principals to collect data from the teachers. The participants were informed that a study is being conducted to investigate teachers' assessment attitudes, competence, knowledge, and practices. 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 the 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. The participants took on average one hour to complete the questionnaire.

2.4 Data Analysis

The data analysis included descriptive statistics using frequencies, percentages, means, and standard deviations. Factorial analyses of variance (Factorial ANOVA) were also used to examine differences in teachers' attitude towards and knowledge in educational assessment as well as their attitude towards classroom tests with respect to teachers' gender, nationality, governorate, qualification, current teaching grade, teaching subject, and pre-service and in-service training in the educational assessment. Multivariate analyses of variance (MANOVA) were used to examine differences in teachers' competence and practices in educational assessment as well as their uses of classroom tests with respect to teachers' gender, nationality, governorate, qualification, current teaching grade, teaching subject, and pre-service and in-service training in the educational assessment. Post-hoc comparisons were conducted using a Least Squares Difference (LSD) test. With regard to the results of the LSD, only the largest statistically significant mean difference will be reported. Readers are invited to contact the corresponding author for details of the results concerning the LSD. Pearson product-moment correlation coefficients were computed to examine relationships of teachers' teaching load and teaching experience to their assessment attitudes, competence, knowledge, and practices.

3. Results

3.1 Attitude towards Educational Assessment

An analysis of teachers' attitude towards educational assessment is presented in Table 2. Overall, the teachers tended to have a positive attitude towards educational assessment ($M = 3.94$, $SD = .47$). The majority of the teachers (89%) reported having positive or strongly positive attitude towards educational assessment. About 10% reported being neutral in their attitude towards educational assessment and less than 1% perceived themselves to have negative or strongly negative attitude towards educational assessment.

Table 2. Frequencies for teachers' attitude towards educational assessment (N = 3557)

Scale value		<i>f</i>	%
1.00 – 1.79	Strongly negative attitude	4	.1
1.80 – 2.59	Negative attitude	17	.5
2.60 – 3.39	Neutral	368	10.3
3.40 – 4.19	Positive	2085	58.6
4.20 – 5.00	Strongly positive attitude	1083	30.4

Further analysis of teachers' attitude towards educational assessment was conducted to examine differences with respect to teachers' gender, nationality, governorate, qualification, current teaching grade, teaching subject, and pre-service and in-service training in the educational assessment using factorial ANOVA. Table 3 summarizes results of the factorial ANOVA. As shown in Table 3, there were no statistically significant differences in the attitude towards educational assessment between the teachers with respect to their educational qualification and teaching grade. However, there were statistically significant mean differences in the attitude towards educational assessment between the teachers with respect to their gender (partial $\eta^2 = .003$), nationality (partial $\eta^2 = .003$), governorate (partial $\eta^2 = .007$), teaching subject (partial $\eta^2 = .009$), pre-service assessment training (partial $\eta^2 = .010$), and in-service assessment training (partial $\eta^2 = .005$). On average, male teachers tended to have a stronger positive attitude towards educational assessment than female teachers; non-Omani teachers tended to have a stronger positive attitude towards educational assessment than Omani teachers; teachers having at least one pre-service course in the educational assessment tended to have a stronger positive attitude towards educational assessment than teachers having no pre-service course in the educational assessment; and teachers having at least one in-service training workshop in the educational assessment tended to have a stronger positive attitude towards educational assessment than teachers having no in-service training workshop in the educational assessment. The LSD test indicated that the largest statistically significant mean difference among governorates in the attitude towards educational assessment was between Al-Wosta teachers and Al-Buraimi teachers favoring Al-Wosta teachers. Also, the LSD test indicated that the largest statistically significant mean difference among teaching subjects in the attitude towards educational assessment was between science teachers and social studies teachers favoring science teachers. Pearson product-moment correlation coefficients indicated that teachers' attitude towards educational assessment correlated positively with teaching experience ($r = .08$) and negatively with teaching load ($r = -.08$), $ps < .001$.

Table 3. Factorial ANOVA for the attitude towards educational assessment

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i> -value
Gender	2.26	1	2.26	10.76	.001
Nationality	2.49	1	2.49	11.88	.001
Governorate	5.16	10	.52	2.46	.006
Qualification	.48	1	.48	2.29	.13
Teaching grade	1.12	7	.16	.76	.619
Teaching subject	6.64	6	1.11	5.28	.000
Pre-service training	7.13	1	7.13	34.02	.000
In-service training	4.04	1	4.04	19.29	.000
Error	739.25	3528	.21		

3.2 Self-Perceived Competence in Educational Assessment

Table 4 presents descriptive statistics for teachers' competencies in educational assessment. As shown in Table 4, on average, the teachers tended to perceive themselves as being moderately competent in analyzing assessment results and highly competent in developing assessment methods, developing performance assessment, developing valid grading procedures, communicating assessment results, and recognizing ethics of assessment.

More than 80% of the teachers perceived themselves as being highly or very highly competent in developing assessment methods, developing performance assessment, developing valid grading procedures, and recognizing ethics of assessment. About 75% of the teachers perceived themselves as being highly or very highly competent in communicating assessments results. Less than half of the teachers (45%) perceived themselves as being highly or very highly competent in analyzing assessments results.

Table 4. Descriptive statistics for teachers' competencies in educational assessment (N = 3557)

Assessment competencies	Scale value					<i>M</i>	<i>SD</i>
	1.00-1.79	1.80-2.59	2.60-3.39	3.40-4.19	4.20-5.00		
	Very low	Low	Moderate	High	Very high		
	<i>f</i> (%)	<i>f</i> (%)	<i>f</i> (%)	<i>f</i> (%)	<i>f</i> (%)		
1. Developing assessment methods	4 (0.1)	35 (1.0)	498 (14.0)	2005 (56.4)	1015 (28.5)	3.92	.52
2. Analyzing assessment results	51 (1.4)	525 (14.8)	1381 (38.8)	1195 (33.6)	405 (11.4)	3.00	.70
3. Developing performance assessment	6 (0.2)	50 (1.4)	627 (17.6)	1982 (55.7)	892 (25.1)	3.82	.52
4. Developing valid grading procedures	13 (0.4)	122 (3.4)	735 (20.7)	2032 (57.1)	655 (18.4)	3.71	.58
5. Communicating assessment results	9 (0.3)	52 (1.5)	841 (23.6)	1806 (50.8)	849 (23.9)	3.79	.56
6. Recognizing ethics of assessment	1 (0.0)	15 (0.4)	282 (7.9)	1775 (49.9)	1484 (41.7)	4.08	.49

Further analysis of teachers' competencies in educational assessment was conducted to examine differences with respect to teachers' gender, nationality, governorate, qualification, teaching grade, teaching subject, pre-service training in assessment, and in-service training in assessment using MANOVA. Table 5 summarizes results of the MANOVA on the teachers' competencies in educational assessment. As shown in Table 5, there were statistically significant multivariate effects on the teachers' assessment competencies with respect to gender (partial $\eta^2 = .085$), nationality (partial $\eta^2 = .015$), governorate (partial $\eta^2 = .009$), teaching grade (partial $\eta^2 = .004$), teaching subject (partial $\eta^2 = .011$), and in-service training in assessment (partial $\eta^2 = .013$). There were no statistically significant multivariate effects for qualification and pre-service assessment training on the teachers' assessment competencies.

Table 5. MANOVA for teachers' competencies in educational assessment

Variable	<i>Wilks' Lambda</i>	<i>F</i>	<i>Hypothesis df</i>	<i>Error df</i>	<i>p-value</i>
Gender	.915	54.22	6	3523	.000
Nationality	.985	8.91	6	3523	.000
Governorate	.949	3.08	60	18463.15	.000
Qualification	.998	1.45	6	3523	.191
Teaching grade	.979	1.78	42	16527.79	.001
Subject	.935	6.62	36	15473.34	.000
Pre-service training	.997	1.65	6	3523	.130
In-service training	.987	7.68	6	3523	.000

The univariate analyses showed statistically significant gender differences favoring females on the perceived competence in developing assessment methods; $F(1, 3528) = 69.83, p = .000$, partial $\eta^2 = .019$; developing performance assessment; $F(1, 3528) = 68.61, p = .017$, partial $\eta^2 = .019$; developing valid grading procedures; $F(1, 3528) = 41.94, p = .021$, partial $\eta^2 = .012$; communicating assessment results; $F(1, 3528) = 45.17, p = .001$, partial $\eta^2 = .013$; and recognizing ethics of assessment; $F(1, 3528) = 259.49, p = .001$, partial $\eta^2 = .069$. Also, the analyses showed statistically significant nationality differences favoring non-Omani teachers on the perceived competence in developing assessment methods; $F(1, 3528) = 27.33, p = .000$, partial $\eta^2 = .008$; analyzing assessment results; $F(1, 3528) = 32.61, p = .000$, partial $\eta^2 = .009$; developing performance assessment; $F(1, 3528) = 16.06, p = .017$, partial $\eta^2 = .005$; developing valid grading procedures; $F(1, 3528) = 5.95, p = .015$, partial $\eta^2 = .002$; communicating assessment results; $F(1, 3528) = 6.06, p = .014$, partial $\eta^2 = .002$; and recognizing ethics of assessment; $F(1, 3528) = 5.34, p = .021$, partial $\eta^2 = .002$. Furthermore, the univariate analyses showed statistically significant differences with respect to in-service assessment training favoring teachers having at least one in-service training workshop or course in assessment on the perceived competence in developing assessment methods; $F(1, 3528) = 35.62, p = .000$, partial $\eta^2 = .010$; analyzing assessment results; $F(1, 3528) = 34.06, p = .000$, partial $\eta^2 = .010$; developing performance assessment; $F(1, 3528) = 32.79, p = .000$, partial $\eta^2 = .009$; developing valid grading procedures; $F(1, 3528) = 32.17, p = .000$, partial $\eta^2 = .009$; communicating assessment results; $F(1, 3528) = 20.99, p = .000$, partial $\eta^2 = .006$; and recognizing ethics of assessment; $F(1, 3528) = 8.51, p = .004$, partial $\eta^2 = .002$. Likewise, the univariate analyses showed statistically significant differences with respect to educational qualification favoring teachers having above bachelor degree on the perceived competence in developing performance assessment; $F(1, 3528) = 6.00, p = .014$, partial $\eta^2 = .002$; and communicating assessment results; $F(1, 3528) = 4.79, p = .029$, partial $\eta^2 = .001$.

In addition, there were statistically significant univariate effects for governorate on the perceived competence in developing assessment methods; $F(10, 3528) = 3.20, p = .000$, partial $\eta^2 = .009$; analyzing assessment results; $F(10, 3528) = 5.37, p = .000$, partial $\eta^2 = .015$; developing performance assessment; $F(10, 3528) = 1.90, p = .041$, partial $\eta^2 = .005$; developing valid grading procedures; $F(10, 3528) = 2.61, p = .004$, partial $\eta^2 = .007$; communicating assessment results; $F(10, 3528) = 3.35, p = .000$, partial $\eta^2 = .009$; and recognizing ethics of assessment; $F(10, 3528) = 2.07, p = .023$, partial $\eta^2 = .006$. According to the LSD test, the largest statistically significant mean difference on the perceived competence in developing assessment methods, analyzing assessment results, and communicating assessment results were between Musandam teachers and Al-Sharqiyah North teachers favoring Musandam teachers. Also, the largest statistically significant mean difference on the perceived competence in developing performance assessment was between Musandam teachers and Al-Dakhiliyah teachers favoring Musandam teachers. Likewise, the largest statistically significant mean difference on the perceived competence in developing valid grading procedures was between Musandam teachers and Al-Batinah South teachers favoring Musandam teachers. Finally, the largest statistically significant mean difference on the perceived competence in recognizing ethics of assessment was between Musandam teachers and Al-Buraimi teachers favoring Musandam teachers.

Further, there were statistically significant univariate effects for teaching grade on the perceived competence in recognizing ethics in assessment; $F(7, 3528) = 2.89, p = .005$, partial $\eta^2 = .006$. According to the LSD test, the largest statistically significant mean difference on the perceived competence in recognizing ethics of assessment was between grade 12 teachers and grade 5 teachers favoring grade 12 teachers.

Moreover, there were statistically significant univariate effects for teaching subject on the perceived competence in analyzing assessment results; $F(6, 3528) = 6.58, p = .000$, partial $\eta^2 = .011$; developing performance assessment; $F(6, 3528) = 4.11, p = .000$, partial $\eta^2 = .007$; and communicating assessment results; $F(6, 3528) = 4.81, p = .000$, partial $\eta^2 = .008$. According to the LSD test, the largest statistically significant mean differences on the perceived competence in analyzing assessment results and communicating assessment results were between Arabic language teachers and science teachers favoring Arabic language teachers. Also, the largest statistically significant mean difference on the perceived competence in developing performance assessment was between Islamic education teachers and mathematics teachers favoring Islamic education teachers.

Table 6 displays Pearson product-moment correlation coefficients of teaching load per week and teaching experience with teacher's competence in the educational assessment. As shown in Table 6, weekly teaching load correlated negatively with teacher's self-perceived competence in developing assessment methods and recognizing ethics of assessment. According to Table 6, there were statistically significant positive relationships between teaching experience and teacher's self-perceived competence in developing assessment methods, analyzing assessment results, developing performance assessment, and communicating assessment results to

various audiences.

Table 6. Pearson product-moment correlation coefficients of teaching load and teaching experience with teacher's competence in the educational assessment (N = 3557)

Variable	Teaching load	Teaching experience
1. Developing assessment methods	-.040*	.099**
2. Analyzing assessment results	.000	.107**
3. Developing performance assessment	-.007	.039*
4. Developing valid grading procedures	.007	.019
5. Communicating assessment results	-.016	.058**
6. Recognizing ethics of assessment	-.036*	.008

* $p < .05$, ** $p < .01$.

3.3 Knowledge in Educational Assessment

The scores of the participating teachers on the TALQ ranged from 1 to 21 with an average of 12.01 and a standard deviation of 3.15. Twenty-five percent of the teachers answered 10 items or less correctly out of 28 items of the TALQ. Half of the teachers answered 12 items or less correctly out of 28 items of the TALQ. Three quarters of the teachers answered 14 items or less correctly out of 28 items of the TALQ.

Further analysis of teachers' knowledge in educational assessment was conducted to examine differences with respect to teachers' gender, nationality, governorate, qualification, teaching grade, teaching subject, pre-service training in assessment, and in-service training in assessment using factorial ANOVA. Table 7 summarizes results of the factorial ANOVA. As shown in Table 7, there were no statistically significant differences in the educational assessment knowledge as measured by TALQ's scores between the teachers with respect to their nationality, qualification, and in-service training in assessment. However, there was a statistically significant mean difference in the educational assessment knowledge with respect to teacher's gender (partial $\eta^2 = .032$) favoring female teachers. Also, there was a statistically significant mean difference in the educational assessment knowledge with respect to pre-service assessment training (partial $\eta^2 = .012$) favoring teachers having at least one pre-service course in educational assessment. Further, there was a statistically significant mean difference in the educational assessment knowledge with respect to governorate (partial $\eta^2 = .009$). The LSD test showed that the largest statistically significant mean difference in the educational assessment among governorates was between Musandam teachers and Al-Buraimi teachers favoring Musandam teachers. In addition, there was a statistically significant mean difference in the educational assessment knowledge with respect to the teaching grade (partial $\eta^2 = .005$). The LSD test showed that the largest statistically significant mean difference in the educational assessment among teaching grades was between the 12th grade teachers and the 6th grade teachers favoring the 12th grade teachers. Moreover, there was a statistically significant mean difference in the educational assessment knowledge with respect to the teaching subject (partial $\eta^2 = .027$). The LSD test showed that the largest statistically significant mean difference in the educational assessment among teaching subjects was between Arabic language teachers and mathematics teachers favoring mathematics teachers. Pearson product-moment correlation coefficients indicated that teachers' knowledge in educational assessment correlated positively with teaching experience ($r = .07$) and negatively with teaching load ($r = -.07$), $ps < .001$.

Table 7. Factorial ANOVA for the knowledge in educational assessment

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i> -value
Gender	1053.25	1	1053.25	115.57	.000
Nationality	8.16	1	8.16	.895	.344
Governorate	285.90	10	28.59	3.14	.001
Qualification	29.12	1	29.12	3.20	.074
Teaching grade	163.25	7	23.32	2.56	.013
Teaching subject	885.65	6	147.61	16.20	.000
Pre-service training	378.74	1	378.74	41.56	.000
In-service training	18.15	1	18.15	1.99	.158
Error	32151.52	3528	9.11		

3.4 Educational Assessment Practices

Table 8 presents descriptive statistics for teachers' assessment practices. As shown in Table 8, the teachers reported on average involving students in the assessment process, analyzing assessment results, using alternative assessment methods, and using non-achievement factors in grading some of the time. More than one third of the teachers (37%) indicated involving students in the assessment process most to all of the time. Less than one third of the teachers (31.9%) reported analyzing assessment results most to all of the time. Less than half of the teachers (48%) mentioned using alternative assessments most to all of the time. More than two third of the teachers (68.4%) indicated using non-achievement factors in grading never to some of the time. Also, the teachers reported on average using traditional assessment methods, developing scoring criteria and standards for performance assessments, and communicating assessment results to various audiences most of the time. Nearly half of the teachers (48.2%) indicated using traditional assessment methods most to all of the time. About 88% of the teachers reported communicating assessment results to various audiences most to all of the time. Approximately three quarters of the teachers (75.5%) indicated developing scoring criteria and standards for performance assessments most to all of the time.

Table 8. Descriptive statistics for teachers' assessment practices (N = 3557)

Assessment practices	Scale value					<i>M</i>	<i>SD</i>
	1.00-1.79	1.80-2.59	2.60-3.39	3.40-4.19	4.20-5.00		
	Never	Seldom	Some of the time	Most of the time	All of the time		
	<i>f</i> (%)	<i>f</i> (%)	<i>f</i> (%)	<i>f</i> (%)	<i>f</i> (%)		
1. Traditional assessment methods	29 (0.8)	380 (10.7)	1435 (40.3)	1304 (36.7)	409 (11.5)	3.40	.68
2. Alternative assessment methods	15 (0.4)	270 (7.6)	1563 (43.9)	1405 (39.5)	304 (8.5)	3.29	.59
3. Analysis of assessment results	168 (4.7)	838 (23.6)	1419 (39.9)	884 (24.9)	248 (7.0)	3.07	.77
4. Assessment communication	2 (0.1)	30 (0.8)	382 (10.7)	1756 (49.4)	1387 (39.0)	4.02	.54
5. Assessment standards and criteria	5 (0.1)	101 (2.8)	763 (21.5)	1854 (52.1)	834 (23.4)	3.68	.58
6. Student-involved assessment	155 (4.4)	706 (19.8)	1382 (38.9)	977 (27.5)	337 (9.5)	3.15	.74
7. Non-achievement grading factors	224 (6.3)	838 (23.6)	1369 (38.5)	929 (26.1)	197 (5.5)	3.01	.78

Further analysis of teachers' assessment practices was conducted to examine differences with respect to teachers'

gender, nationality, governorate, qualification, teaching grade, teaching subject, pre-service training in assessment, and in-service training in assessment using MANOVA. Table 9 summarizes results of the MANOVA on the teachers' assessment practices. As shown in Table 8, there were statistically significant multivariate effects on the teachers' assessment practices with respect to gender (partial $\eta^2 = .175$), nationality (partial $\eta^2 = .022$), governorate (partial $\eta^2 = .009$), teaching grade (partial $\eta^2 = .006$), teaching subject (partial $\eta^2 = .049$), pre-service assessment training (partial $\eta^2 = .010$), and in-service training in assessment (partial $\eta^2 = .013$). There were no statistically significant multivariate effects for qualification on the teachers' assessment practices.

Table 9. MANOVA for teachers' assessment practices

Variable	Wilks' Lambda	F	Hypothesis df	Error df	p-value
Gender	.825	106.48	7	3522	.000
Nationality	.978	11.09	7	3522	.000
Governorate	.936	3.34	70	20543.43	.000
Qualification	.998	1.23	7	3522	.282
Teaching grade	.957	3.16	49	17885.02	.000
Subject	.741	25.96	42	16523.10	.000
Pre-service training	.990	5.02	7	3522	.000
In-service training	.987	6.49	7	3522	.000

The univariate analyses showed statistically significant gender differences favoring males on analyzing assessment results; $F(1, 3528) = 34.73, p = .000$, partial $\eta^2 = .010$; involving students in assessment; $F(1, 3528) = 20.01, p = .000$, partial $\eta^2 = .006$; and using non-achievement grading factors; $F(1, 3528) = 279.87, p = .000$, partial $\eta^2 = .073$. Further, the analyses showed statistically significant gender differences favoring females on communicating assessment results; $F(1, 3528) = 197.65, p = .000$, partial $\eta^2 = .053$; and using assessment standards and criteria; $F(1, 3528) = 10.63, p = .001$, partial $\eta^2 = .003$. Also, the analyses showed statistically significant nationality differences favoring non-Omani teachers on using traditional assessment methods; $F(1, 3528) = 19.02, p = .000$, partial $\eta^2 = .005$; using alternative assessment methods; $F(1, 3528) = 15.77, p = .000$, partial $\eta^2 = .004$; and analyzing assessment results; $F(1, 3528) = 24.42, p = .000$, partial $\eta^2 = .007$. Further, the analyses showed statistically significant nationality differences favoring Omani on using assessment standards and criteria; $F(1, 3528) = 7.57, p = .006$, partial $\eta^2 = .002$. Furthermore, the univariate analyses showed statistically significant differences with respect to pre-service assessment training on using assessment standards and criteria favoring teachers having at least one pre-service course in assessment; $F(1, 3528) = 4.03, p = .045$, partial $\eta^2 = .001$; and using non-achievement grading factors favoring teachers having no pre-service course in assessment; $F(1, 3528) = 18.45, p = .045$, partial $\eta^2 = .005$. Likewise, the univariate analyses showed statistically significant differences with respect to in-service assessment training on using traditional assessment methods favoring teachers having no in-service course in assessment; $F(1, 3528) = 8.53, p = .004$, partial $\eta^2 = .002$. Also, the univariate analyses showed statistically significant differences with respect to in-service assessment training favoring teachers having at least one in-service course in assessment on using alternative assessment methods; $F(1, 3528) = 25.07, p = .000$, partial $\eta^2 = .007$; analyzing assessment results; $F(1, 3528) = 30.12, p = .000$, partial $\eta^2 = .008$; communicating assessment results to various audiences; $F(1, 3528) = 18.26, p = .000$, partial $\eta^2 = .005$; using assessment standards and criteria; $F(1, 3528) = 14.69, p = .000$, partial $\eta^2 = .004$; and using student involved-assessment; $F(1, 3528) = 8.03, p = .005$, partial $\eta^2 = .002$.

In addition, there were statistically significant univariate effects for governorate on using alternative assessment methods; $F(10, 3528) = 4.27, p = .000$, partial $\eta^2 = .012$; analyzing assessment results; $F(10, 3528) = 6.76, p = .000$, partial $\eta^2 = .019$; communicating assessment results; $F(10, 3528) = 2.41, p = .007$, partial $\eta^2 = .007$; using assessment standards and criteria; $F(10, 3528) = 3.60, p = .000$, partial $\eta^2 = .010$; using student-involved assessment; $F(10, 3528) = 2.82, p = .002$, partial $\eta^2 = .008$; and using non-achievement grading factors; $F(10, 3528) = 3.99, p = .000$, partial $\eta^2 = .011$. According to the LSD test, the largest statistically significant mean difference on using alternative assessments was between Musandam teachers and Al-Sharqiyah South teachers favoring Musandam teachers. Also, the largest statistically significant mean difference on analyzing assessment results and involving students in assessment was between Al-Wosta teachers and Al-Sharqiyah North teachers

favoring Al-Wosta teachers. Likewise, the largest statistically significant mean difference on communicating assessment results was between Musandam teachers and Al-Buraimi teachers favoring Musandam teachers. Also, the largest statistically significant mean difference on using assessment standards and criteria was between Al-Batinah North teachers and Al-Wosta teachers favoring Al-Batinah North teachers. Finally, the largest statistically significant mean difference on using non-achievement grading factors was between Al-Wosta teachers and Al-Dakhiliyah teachers favoring Al-Wosta teachers.

Further, there were statistically significant univariate effects for teaching grade on using traditional assessment methods; $F(7, 3528) = 5.06, p = .000$, partial $\eta^2 = .010$; using alternative assessment methods; $F(7, 3528) = 7.01, p = .000$, partial $\eta^2 = .014$; analyzing assessment results; $F(7, 3528) = 2.02, p = .049$, partial $\eta^2 = .004$; using assessment standards and criteria; $F(7, 3528) = 2.10, p = .040$, partial $\eta^2 = .004$; and using non-achievement grading factors; $F(7, 3528) = 4.01, p = .000$, partial $\eta^2 = .008$. According to the LSD test, the largest statistically significant mean difference on using traditional assessment methods was between grade 7 teachers and grade 11 teachers favoring grade 7 teachers. Also, the largest statistically significant mean difference on using alternative assessment methods was between grade 6 teachers and grade 9 teachers favoring grade 9 teachers. Further, the largest statistically significant mean difference on analyzing assessment results was between grade 11 teachers and grade 12 teachers favoring grade 12 teachers. The largest statistically significant mean difference on using assessment standards and criteria was between grade 5 teachers and grade 12 teachers favoring grade 12 teachers. Finally, the largest statistically significant mean difference on using non-achievement grading factors was between grade 5 teachers and grade 11 teachers favoring grade 5 teachers.

Moreover, there were statistically significant univariate effects for teaching subject on using traditional assessment methods; $F(5, 3528) = 111.85, p = .000$, partial $\eta^2 = .016$; using alternative assessment methods; $F(5, 3528) = 6.91, p = .000$, partial $\eta^2 = .012$; analyzing assessment results; $F(5, 3528) = 10.41, p = .049$, partial $\eta^2 = .017$; using assessment standards and criteria; $F(5, 3528) = 10.29, p = .000$, partial $\eta^2 = .017$; using student involved-assessment; $F(5, 3528) = 11.29, p = .000$, partial $\eta^2 = .019$; and using non-achievement grading factors; $F(5, 3528) = 35.57, p = .000$, partial $\eta^2 = .057$. According to the LSD test, the largest statistically significant mean difference on using traditional assessment methods was between Islamic education teachers and mathematics teachers favoring Islamic education teachers. Also, the largest statistically significant mean difference on using alternative assessment methods was between mathematics teachers and teachers of practical-based subjects favoring teachers of practical-based subjects. Further, the largest statistically significant mean difference on analyzing assessment results was between Islamic education teachers and science teachers favoring Islamic education teachers. The largest statistically significant mean difference on using assessment standards and criteria was between science teachers and English language teachers favoring science teachers. The largest statistically significant mean difference on using student-involved assessment was between Arabic language teachers and science teachers favoring Arabic language teachers. Finally, the largest statistically significant mean difference on using non-achievement grading factors was between science teachers and teachers of practical-based subjects favoring teachers of practical-based subjects.

Table 10 displays Pearson product-moment correlation coefficients of teaching load per week and teaching experience with teachers' assessment practices. As shown in Table 10, the teaching load correlated positively with the teacher's use of traditional assessments, alternative assessment, student-involved assessment, and non-achievement grading factors. There were no statistically significant correlations between teaching load and teacher's use of analysis of assessment results, communication of assessment results, and assessment standards and criteria. According to Table 10, the teaching experience correlated negatively with teachers' use of alternative assessments and positively with analysis of assessment results. However, the teaching experience did not correlate significantly with the teacher's use of traditional assessments, communication of assessment results, assessment standards and criteria, student-involved assessment, and non-achievement grading factors.

Table 10. Pearson product-moment correlation coefficients of teaching load and teaching experience with teachers' assessment practices (N = 3557)

Variable	Teaching load	Teaching experience
1. Traditional assessment methods	.041*	.012
2. Alternative assessment methods	.051**	-.051**
3. Analysis of assessment results	.006	.112**
4. Assessment communication	-.032	.004
5. Assessment standards and criteria	-.025	-.015
6. Student-involved assessment	.034*	.007
7. Non-achievement grading factors	.056**	.005

* $p < .05$, ** $p < .01$.

3.5 Uses of Classroom Tests

Table 11 presents descriptive statistics for teachers' uses of classroom tests. As shown in Table 11, on average, the teachers reported using classroom tests for diagnosing students' weaknesses, assigning grades, motivating students for learning, and evaluating academic achievement all of the time. Also, on average the teachers indicated using classroom tests most of the time for other purposes such as grouping students for instruction, planning for instruction, evaluating instructional methods, controlling students' behavior, comparing students' performances with each other, and upgrading students from one class to another.

Table 11. Descriptive statistics for teachers' uses of classroom tests (N = 3557)

Uses of classroom tests	Scale value					M	SD
	1.00-1.79	1.80-2.59	2.60-3.39	3.40-4.19	4.20-5.00		
	Never	Seldom	Some of the time	Most of the time	All of the time		
	f (%)	f (%)	f (%)	f (%)	f (%)		
1. Diagnose student weaknesses	21 (0.6)	57 (1.6)	396 (11.1)	1351 (38.0)	1732 (48.7)	4.33	.78
2. Group students for instruction	103 (2.9)	280 (7.9)	1152 (32.4)	1181 (33.2)	841 (23.6)	3.67	1.01
3. Plan for instruction	71 (2.0)	230 (6.5)	1012 (28.5)	1483 (41.7)	761 (21.4)	3.74	.93
4. Assign grades	10 (0.3)	26 (0.7)	212 (6.0)	863 (24.3)	2446 (68.8)	4.61	.66
5. Evaluate instructional methods	53 (1.5)	221 (6.2)	879 (24.7)	1451 (40.8)	953 (26.8)	3.85	.94
6. Control student behavior	122 (3.4)	207 (5.8)	412 (11.6)	982 (27.6)	1834 (51.6)	4.18	1.07
7. Motivate students for learning	20 (0.6)	47 (1.3)	275 (7.7)	994 (27.9)	2221 (62.4)	4.50	.74
8. Evaluate academic achievement	10 (0.3)	35 (1.0)	238 (6.7)	1012 (28.5)	2262 (63.6)	4.54	.69
9. Compare students' performances	166 (4.7)	338 (9.5)	943 (26.5)	1183 (33.3)	927 (26.1)	3.67	1.10
10. Upgrade students to upper classes	125 (3.5)	199 (5.6)	651 (18.3)	1374 (38.6)	1208 (34.0)	3.94	1.03

Further analysis of teachers' uses of classroom tests was conducted to examine differences with respect to teachers' gender, nationality, governorate, qualification, teaching grade, teaching subject, pre-service training in assessment, and in-service training in assessment using MANOVA. Table 12 summarizes results of the MANOVA on the teachers' uses of classroom tests. As shown in Table 11, there were statistically significant multivariate effects on the teachers' uses of classroom tests with respect to gender (partial $\eta^2 = .112$), governorate (partial $\eta^2 = .007$), teaching grade (partial $\eta^2 = .006$), teaching subject (partial $\eta^2 = .011$), pre-service assessment training (partial $\eta^2 = .006$), and in-service training in assessment (partial $\eta^2 = .010$). There were no statistically significant multivariate effects for nationality and qualification on the teachers' uses of classroom tests.

Table 12. MANOVA for teachers' uses of classroom tests

Variable	Wilks' Lambda	F	Hypothesis df	Error df	p-value
Gender	.888	44.32	10	3519	.000
Nationality	.996	1.53	10	3519	.124
Governorate	.936	2.33	100	25206.90	.000
Qualification	.996	1.26	10	3519	.245
Teaching grade	.961	2.01	70	20525.94	.000
Subject	.936	3.92	60	18442.19	.000
Pre-service training	.994	1.99	10	3519	.031
In-service training	.990	3.514	10	3519	.000

The univariate analyses showed statistically significant gender differences favoring females on using classroom tests for diagnosing students' weaknesses; $F(1, 3528) = 35.23, p = .000$, partial $\eta^2 = .010$; grouping students for instruction; $F(1, 3528) = 249.37, p = .000$, partial $\eta^2 = .066$; planning for instruction; $F(1, 3528) = 29.71, p = .000$, partial $\eta^2 = .008$; assigning grades; $F(1, 3528) = 118.72, p = .000$, partial $\eta^2 = .033$; evaluating instructional methods; $F(1, 3528) = 20.58, p = .000$, partial $\eta^2 = .006$; motivating students for learning; $F(1, 3528) = 14.05, p = .000$, partial $\eta^2 = .004$; evaluating academic achievement; $F(1, 3528) = 112.31, p = .000$, partial $\eta^2 = .031$; and upgrading students to upper classes; $F(1, 3528) = 23.97, p = .000$, partial $\eta^2 = .007$. Also, the analyses showed statistically significant differences with respect to pre-service assessment training on using classroom tests for assigning grades favoring teachers having at least one pre-service course in assessment $F(1, 3528) = 4.92, p = .027$, partial $\eta^2 = .001$. Furthermore, the univariate analyses showed statistically significant differences with respect to in-service assessment training favoring teachers having at least one in-service course in assessment on using classroom tests for diagnosing students' weaknesses; $F(1, 3528) = 5.52, p = .019$, partial $\eta^2 = .002$; grouping students for instruction; $F(1, 3528) = 15.15, p = .000$, partial $\eta^2 = .004$; planning for instruction; $F(1, 3528) = 15.42, p = .000$, partial $\eta^2 = .004$; evaluating instructional methods; $F(1, 3528) = 12.63, p = .000$, partial $\eta^2 = .004$; motivating students for learning; $F(1, 3528) = 9.26, p = .002$, partial $\eta^2 = .003$; evaluating academic achievement; $F(1, 3528) = 5.34, p = .021$, partial $\eta^2 = .002$; and upgrading students to upper classes; $F(1, 3528) = 10.28, p = .001$, partial $\eta^2 = .003$.

Likewise, there were statistically significant univariate effects for governorate on using classroom tests for grouping students for instruction; $F(10, 3528) = 3.24, p = .000$, partial $\eta^2 = .009$; assigning grades; $F(10, 3528) = 3.66, p = .000$, partial $\eta^2 = .010$; evaluating instructional methods; $F(10, 3528) = 2.05, p = .026$, partial $\eta^2 = .006$; controlling student behavior; $F(10, 3528) = 1.95, p = .035$, partial $\eta^2 = .005$; motivating students for learning; $F(10, 3528) = 2.67, p = .003$, partial $\eta^2 = .008$; evaluating academic achievement; $F(10, 3528) = 2.06, p = .024$, partial $\eta^2 = .006$; and comparing students' performances; $F(10, 3528) = 4.41, p = .000$, partial $\eta^2 = .012$. The LSD test showed that the largest statistically significant mean differences on using classroom tests for grouping students for instruction was between Al-Dhahira teachers and Al-Sharqiyah North teachers favoring Al-Sharqiyah North teachers. The largest statistically significant mean differences on using classroom tests for assigning grades and controlling student behavior was between Al-Wosta teachers and Musandam teachers favoring Al-Wosta teachers. The largest statistically significant mean differences on using classroom tests for evaluating instructional methods was between Musandam teachers and Al-Sharqiyah North teachers favoring Musandam teachers. The largest statistically significant mean differences on using classroom tests for motivating

students for learning was between Al-Batinah North teachers and Al-Wosta teachers favoring Al-Wosta teachers. The largest statistically significant mean differences on using classroom tests for evaluating academic achievement was between Al-Sharqiyah North teachers and Al-Batinah North teachers favoring Al-Sharqiyah North teachers. The largest statistically significant mean differences on using classroom tests for comparing students' performances was between Al-Buraimi teachers and Musandam teachers favoring Al-Buraimi teachers.

In addition, there were statistically significant univariate effects for teaching grade on using classroom tests for grouping students for instruction; $F(7, 3528) = 5.52, p = .000$, partial $\eta^2 = .011$; and upgrading students to upper grades; $F(7, 3528) = 2.70, p = .009$, partial $\eta^2 = .005$. The LSD test showed that the largest statistically significant mean differences on using classroom tests for grouping students for instruction was between grade 8 teachers and grade 5 teachers favoring grade 8 teachers. According to the LSD test, the largest statistically significant mean difference on using classroom tests for upgrading students to upper grade was between grade 9 teachers and grade 5 teachers favoring grade 9 teachers.

Moreover, there were statistically significant univariate effects for teaching subject on using classroom tests for diagnosing student weaknesses; $F(6, 3528) = 2.13, p = .047$, partial $\eta^2 = .004$; grouping students for instruction; $F(6, 3528) = 11.77, p = .000$, partial $\eta^2 = .020$; planning for instruction; $F(6, 3528) = 2.30, p = .032$, partial $\eta^2 = .004$; assigning grades; $F(6, 3528) = 4.10, p = .000$, partial $\eta^2 = .007$; evaluating instructional methods; $F(6, 3528) = 2.83, p = .009$, partial $\eta^2 = .005$; controlling student behavior; $F(6, 3528) = 5.77, p = .000$, partial $\eta^2 = .010$; and upgrading students to upper classes; $F(6, 3528) = 2.55, p = .018$, partial $\eta^2 = .004$. According to the LSD test, the largest statistically significant mean difference on using classroom tests for diagnosing student weaknesses was between Arabic language teachers and teachers of practical-based subjects favoring Arabic language teachers. Also, the largest statistically significant mean difference on using classroom tests for grouping students for instruction was between mathematics teachers and English language teachers favoring mathematics teachers. Further, the largest statistically significant mean differences on using classroom tests for planning for instruction and for controlling student behavior was between mathematics teachers and teachers of practical-based subjects favoring teachers of practical-based subjects. The largest statistically significant mean difference on using classroom tests for assigning grades was between mathematics teachers and Islamic education teachers favoring mathematics teachers. The largest statistically significant mean difference on using classroom tests for evaluating instructional methods was between science teachers and teachers of practical-based subjects favoring teachers of practical-based subjects. Finally, the largest statistically significant mean difference on using classroom tests for upgrading students to upper classes was between science teachers and Arabic language favoring Arabic language teachers.

Table 13 displays Pearson product-moment correlation coefficients of teaching load per week and teaching experience with teachers' uses of classroom tests. As shown in Table 13, weekly teaching load correlated positively with teacher's use of classroom tests for grouping students for instruction, planning for instruction, and comparing student performance with others. There were no statistically significant correlations between weekly teaching load and teacher's use of classroom tests for diagnosing student weaknesses, assigning grades, evaluating instructional methods, controlling students' behavior, motivating students, evaluating academic achievement, and upgrading students from one grade to another. According to Table 13, there were statistically significant negative relationships between teaching experience and teacher's use of classroom tests for grouping students for instruction and planning for instruction. However, teaching experience did not correlate significantly with teacher's use of classroom tests for diagnosing students' weakness, assigning grades, evaluating instructional methods, controlling students' behavior, motivating students, comparing students' performances with each other, and upgrading students from one grade to another.

Table 13. Pearson product-moment correlation coefficients of teaching load and teaching experience with teachers' uses of classroom tests (N = 3557)

Variable	Teaching load	Teaching experience
1. Diagnose student weaknesses	-.023	.024
2. Group students for instruction	.039*	-.093***
3. Plan for instruction	.051**	-.048**
4. Assign grades	-.011	.001
5. Evaluate instructional methods	.004	-.008
6. Control student behavior	.026	-.030
7. Motivate students for learning	.012	-.002
8. Evaluate academic achievement	-.032	.009
9. Compare students' performances	.056**	-.029
10. Upgrade students to upper classes	-.002	.001

* $p < .05$, ** $p < .01$, *** $p < .001$.

3.6 Attitude towards Classroom Tests

An analysis of teachers' attitude towards classroom tests is presented in Table 14. On average, the teachers tended to have a strongly positive attitude towards classroom tests ($M = 4.35$, $SD = .52$). The majority of the teachers (94.4%) reported having positive or strongly positive attitude towards classroom tests.

Table 14. Frequencies for teachers' attitude towards classroom tests (N = 3557)

Scale value		<i>f</i>	%
1.00–1.79	Strongly negative attitude	0	0
1.80–2.59	Negative attitude	6	.1
2.60–3.39	Neutral	190	5.3
3.40–4.19	Positive	873	24.5
4.20–5.00	Strongly positive attitude	2488	69.9

Further analysis of teachers' attitude towards classroom tests was conducted to examine differences with respect to teachers' gender, nationality, governorate, qualification, current teaching grade, teaching subject, and pre-service and in-service training in the educational assessment using factorial ANOVA. Table 15 summarizes results of the factorial ANOVA. As shown in Table 15, there were no statistically significant differences in the attitude towards classroom tests between the teachers with respect to their nationality, educational qualification, teaching grade, and in-service training in the educational assessment. However, there were statistically significant mean differences in the attitude towards classroom tests between the teachers with respect to their gender (partial $\eta^2 = .038$), governorate (partial $\eta^2 = .013$), teaching subject (partial $\eta^2 = .008$), and pre-service assessment training (partial $\eta^2 = .003$). On average, female teachers tended to have a stronger positive attitude towards classroom tests than male teachers and that teachers having at least one pre-service course in the educational assessment tended to have a stronger positive attitude towards classroom tests than teachers having no pre-service course in the educational assessment. The LSD test indicated that the largest statistically significant mean difference among governorates in the attitude towards classroom tests was between Musandam teachers and Al-Sharqiyah North teachers favoring Al-Sharqiyah North teachers. Also, the LSD test indicated that the largest statistically significant mean difference among teaching subjects in the attitude towards classroom tests was between mathematics teachers and English language teachers favoring mathematics teachers. Pearson product-moment correlation coefficients indicated that teachers' attitude towards classroom tests correlated positively with teaching experience ($r = .04$) and negatively with teaching load ($r = -.05$), $ps < .05$.

Table 15. Factorial ANOVA for the attitude towards classroom tests

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i> -value
Gender	35.10	1	35.10	140.21	.000
Nationality	.05	1	.05	.19	.664
Governorate	11.23	10	1.12	4.49	.000
Qualification	.90	1	.90	3.59	.058
Teaching grade	1.58	7	.23	.90	.504
Teaching subject	7.01	6	1.17	4.67	.000
Pre-service training	2.29	1	2.29	9.15	.003
In-service training	.26	1	.26	1.04	.309
Error	883.27	3528			

4. Discussion

Educational assessment is considered “a key policy lever for improving education” Koh (2011, p. 255). It plays a critical role in shaping student academic motivation and performance (Brookhart, 2004). As such, like other countries in the world, Sultanate of Oman strives to improve the quality of education by advocating for a continuous educational assessment system that could make a positive impact on the instructional and learning process. This requires an understanding of the teachers’ current state profile of the educational assessment in terms of attitudes, competence, knowledge, and practices. This is because educators have long recognized that teachers’ knowledge and beliefs might influence their classroom practices (Calderhead, 1996; Green, 1992). As might be expected, undesirable teachers’ knowledge and beliefs about educational assessment could cripple the quality of the educational outcomes (Popham, 2009). Thus, the present study sought to develop such profile as a function of teachers’ gender, nationality, educational governorate, teaching grade, qualification, teaching subject, pre-service assessment training, in-service assessment training, teaching load, and teaching experience. Overall, results revealed that the teachers held a favorable attitude towards and perceived themselves as being competent in the educational assessment. However, they demonstrated a low level of knowledge in the educational assessment. The teachers indicated using a variety of assessments in the classroom primarily for assigning grades and motivating students to learn. Teaching load and teaching experience accounted for some of the variations in teachers’ educational assessment attitudes, competence, knowledge, and practices. The educational assessment profile varied as a function of the selected demographic and background variables. These results add support to the existing literature on classroom assessment (e.g., Alkharusi 2011a, 2011b, 2011c; Alkharusi et al., 2011; Alkharusi et al., 2012; Arce-Ferrer et al., 2001; Alsarimi, 2000; DeLuca & Klinger, 2010; Lyon, 2011; Mertler, 2003; Mertler & Campbell, 2005; Ogan-Bekiroglu, 2009; Plake & Impara, 1992; Zhang & Burry-Stock, 2003).

As revealed in this study, there are three areas of educational assessment process that need further attention by the Ministry of Education and teacher educators when designing training programs for teachers. These are analysis of assessment results, student-involved assessment, and non-achievement grading. Compared to the other areas of the educational assessment process, the participating teachers reported a lower practice related to the analysis of the assessment results and student involvement in the assessment process. Also, they reported using non-achievement grading factors when assigning grades for students, which in turn, do not align with those recommended by educational assessment experts which state that non-achievement factors such as effort, ability, interest, and motivation should not be incorporated into academic grades because they are complex to be operationally defined and measured (Stiggins, Frisbie, & Griswold, 1989). The results of the study indicated that teaching load and educational assessment training may play a critical role in teachers’ attitudes, competence, knowledge, and practices in educational assessment. These results were in agreement with Lyon’s (2011) study who found that teaching load and other school responsibilities could cause conflicts between teachers’ assessment beliefs and practices. As suggested by Quilter and Gallini (2000), it might be argued that when teachers are inundated with demands from school administrators and educational supervisors to do various activities in the classroom and school beside their teaching activities, it is unlikely that they will invest the time and effort necessary for performing other aspects of the educational assessment process. Thus, it is recommended that the Ministry of Education considers the teaching load of the teachers to enable them to invest the time and effort in performing all aspects of the classroom assessment activities.

Parallel to Zhang and Burry-Stock (2003), the present study provides additional evidence for the importance of pre-service and in-service training in educational assessment. Theoretically, assessment should be aligned with the objectives of the curriculum and that selection of the assessment strategies should be guided by the nature of the instructional content and the needs of the students (Tomlinson & Moon, 2013). Given the variations in the educational assessment practices with respect to teachers' teaching grade and teaching subject, it seems reasonable to argue that educational assessment training should be tailored to match the specific nature of the teaching grade and teaching subject. In addition, it is recommended that the Ministry of Education and teacher education institutes should continue offering in-service professional development programs in the educational assessment for the teachers. Analysis of assessment results, student-involved assessment, and standards-based grading are three of the programs that might need to be considered more in training teachers.

It should be noted that this study was based on the assumption that the sample and the sampling method would be an adequate representation of the population with respect to the variables of interest: gender, nationality, educational governorate, teaching grade, qualification, teaching subject, pre-service assessment training, in-service assessment training, teaching load, and teaching experience. Also, it was assumed that an equal representation of participants in all identified groups would be obtained so that a fair comparison of the groups could be conducted. Finally, the generalizability of the present study findings are limited by the use of self-report questionnaire, in which individuals might not be able to accurately report what they know about themselves. This issue might have attenuated the effect sizes for the observed differences. Future research might consider using interviews and direct observations of teachers' assessment practices to judge the validity of the teachers' responses to the questionnaire.

Acknowledgements

This research was thankfully supported by a grant (RC/EDU/PSYC/12/01) from The Research Council in Oman. This funding source had no involvement in the conduct of the research and preparation of the article.

References

- Alkharusi, H. (2009). Correlates of teacher education students' academic performance in an educational measurement course. *International Journal of Learning, 16*, 1-15.
- Alkharusi, H. (2010). A multilevel linear model of teachers' assessment practices and students' perceptions of the classroom assessment environment. *Procedia Social and Behavioral Sciences, 5*, 5-11. <http://dx.doi.org/10.1016/j.sbspro.2010.07.041>
- Alkharusi, H. (2011b). A logistic regression model predicting assessment literacy among in-service teachers. *Journal of Theory and Practice in Education, 7*, 280-291.
- Alkharusi, H. (2011c). Teachers' classroom assessment skills: Influence of gender, subject area, grade level, teaching experience, and in-service assessment training. *Journal of Turkish Science Education, 8*, 39-48.
- Alkharusi, H. (2011d). Validity and reliability of the attitude toward educational measurement inventory. *The Asia-Pacific Education Researcher, 20*, 612-620.
- Alkharusi, H. A. (2011a). An analysis of the internal and external structure of the teacher assessment literacy questionnaire. *International Journal of Learning, 18*, 515-528.
- Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2012). Educational assessment attitudes, competence, knowledge, and practices: An exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning, 1*, 217-232. <http://dx.doi.org/10.5539/jel.v1n2p217>
- Alkharusi, H., Kazem, A. M., & Al-Musawai, A. (2011). Knowledge, skills, and attitudes of preservice and inservice teachers in educational measurement. *Asia-Pacific Journal of Teacher Education, 39*, 113-123. <http://dx.doi.org/10.1080/1359866X.2011.560649>
- Alsarimi, A. M. (2000). *Classroom assessment and grading practices in the Sultanate of Oman* (Unpublished Doctoral dissertation, University of Pittsburgh, Pennsylvania).
- American Federation 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, 9*, 30-32. <http://dx.doi.org/10.1111/j.1745-3992.1990.tb00391.x>
- Arce-Ferrer, A. J., Cab, V. P., & Cisneros-Cohernour, E. J. (2001, April). *Teachers' assessment competencies*. Paper presented at the meeting of the American Educational Research Association, Seattle, OH.
- Brookhart, S. M. (2004). Classroom assessment: Tensions and intersections in theory and practice. *Teachers*

- College Record*, 106, 429-458. <http://dx.doi.org/10.1111/j.1467-9620.2004.00346.x>
- Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice*, 30, 3-12. <http://dx.doi.org/10.1111/j.1745-3992.2010.00195.x>
- Bryant, N. C., & Barnes, L. L. B. (1997). Development and validation of the attitude toward educational measurement inventory. *Educational and Psychological Measurement*, 57, 870-875. <http://dx.doi.org/10.1177/0013164497057005011>
- Calderhead, J. (1996). Teacher: Beliefs and knowledge. In D. C. Berliner, & R. C. Calfie (Eds.), *Handbook of educational psychology* (pp. 709-725). New York: Macmillan.
- DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: Identifying gaps in teacher candidates learning. *Assessment in Education: Principles, Policy & Practice*, 17, 419-438. <http://dx.doi.org/10.1080/0969594X.2010.516643>
- Gallagher, D. J. (1998). *Classroom assessment for teachers*. Upper Saddle River, NJ: Merrill.
- Green, K. E. (1992). Differing opinions on testing between preservice and inservice teachers. *Journal of Educational Research*, 86, 37-42. <http://dx.doi.org/10.1080/00220671.1992.9941825>
- Green, K. E., & Stager, S. F. (1987). Testing: Coursework, attitudes, and practice. *Educational Research Quarterly*, 11, 48-55.
- Gronlund, N. E. (2006). *Assessment of student achievement* (8th ed.). Boston: Pearson.
- Koh, K. H. (2011). Improving teachers' assessment literacy through professional development. *Teaching Education*, 22, 255-276. <http://dx.doi.org/10.1080/10476210.2011.593164>
- Koloi-Keaikitse, S. (2012). *Classroom assessment practices: A survey of Botswana primary and secondary school teachers* (Unpublished doctoral dissertation, Ball State University, Indiana).
- Lyon, E. G. (2011). Beliefs, practices and reflection: Exploring a science teacher's classroom assessment through the assessment triangle model. *Journal of Science Teacher Education*, 22, 417-435. <http://dx.doi.org/10.1007/s10972-011-9241-4>
- Mertler, C. A. (2003, October). *Preservice versus inservice teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Mertler, C. A., & Campbell, C. (2005, April). *Measuring teachers' knowledge and application of classroom assessment concepts: Development of the assessment knowledge inventory*. Paper presented at the meeting of the American Educational Research Association, Montreal, Quebec, Canada.
- Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.
- Ogan-Bekiroglu, F. (2009). Assessing assessment: Examination of pre-service physics teachers' attitudes towards assessment and factors affecting their attitudes. *International Journal of Science Education*, 31, 1-29. <http://dx.doi.org/10.1080/09500690701630448>
- Plake, B. S., & Impara, J. C. (1992). *Teacher competencies questionnaire description*. Lincoln, NE: University of Nebraska.
- Popham, W. J. (2009). Assessment literacy for teachers: Faddish or fundamental? *Theory into Practice*, 48, 4-11. <http://dx.doi.org/10.1080/00405840802577536>
- Quilter, S. M., & Gallini, J. K. (2000). Teachers' assessment literacy and attitudes. *The Teacher Educator*, 36, 115-131. <http://dx.doi.org/10.1080/08878730009555257>
- 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. <http://dx.doi.org/10.1111/j.1745-3992.1989.tb00315.x>
- Susuwele-Banda, W. J. (2005). *Classroom assessment in Malawi: Teachers' perceptions and practices in mathematics* (Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University, Virginia).
- Tomlinson, C. A., & Moon, T. R. (2013). Differentiation and classroom assessment. In J. H. McMillan (Ed.), *SAGE handbook of research on classroom assessment* (pp. 415-430). Los Angeles, CA: SAGE Publications. <http://dx.doi.org/10.4135/9781452218649.n23>

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*, 323-342. http://dx.doi.org/10.1207/S15324818AME1604_4

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).