

Teachers' Classroom Assessment Skills: Influence Of Gender, Subject Area, Grade Level, Teaching Experience and In-service Assessment Training

Hussain ALKHARUSI¹

¹ Asist.Prof.Dr., Sultan Qaboos University, College of Education, Department of Psychology, Sultanate of Oman.

Received: 11.07.2010

Revised: 04.01.2011

Accepted: 05.03.2011

The original language of the article is English (v.8, n.2, June 2011, pp.39-48)

ABSTRACT

A significant amount of teachers' professional time is devoted to classroom assessment-related activities. Suboptimal assessment practices might result in undesirable educational outcomes. As such, teachers' assessment skills should deserve recognition and investigation. This study aimed at investigating teachers' self-perceived assessment skills as a function of gender, subject area, grade level, teaching experience, and in-service assessment training. Participants were 213 Omani teachers from Muscat public schools. A 25-item Self-Perceived Assessment Skills Scale was developed and used in the study. Results indicated statistically significant differences on the self-perceived assessment skills with respect to teachers' gender, subject area, grade level, teaching experience, and in-service assessment training. Implications for professional preparation in classroom assessment as well as recommendations for future research are discussed.

Keywords: Science Teachers; In-service Teachers; Teachers' Skills; Teacher Education; Classroom Assessment.

INTRODUCTION

Classroom assessment 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 wide range of assessment methods, to determine the extent to which students are achieving the target instructional outcomes (Gallagher, 1998; Gronlund, 1998). 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 (Zhang & Burry-Stock, 2003). The main purpose of these activities is to improve student learning and motivation to learn (Gronlund, 2006; Harlen & Crick, 2003; Stipek, 2002). Since a substantial proportion of classroom time is devoted to the assessment for and of student learning (Mertler, 2003), suboptimal assessment practices might hinder desirable student learning and motivation. As such, it seems reasonable to argue that careful consideration of the teachers' classroom assessment skills is certainly warranted. In an attempt to guide the professional preparation of teachers in classroom assessment, the American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association (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.

Unfortunately, findings from past and recent studies of classroom assessment have consistently expressed a concern about the adequacy of teachers' assessment skills. 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' behaviour, 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:

1. Teachers try hard to be fair when assigning grades.
2. Teachers inform students about the components of the grades.
3. Achievement tests are the main contributors to grades.
4. Teachers take into account student's effort and ability in grading.
5. Elementary teachers depend on informal evidence and observations, whereas secondary teachers use paper-pencil achievement measures and other written activities in grading.
6. Teachers differ in their perceptions of the meaning and purpose of grades, and consideration of achievement and non-achievement factors in grading.
7. Teachers' grading practices deviate from the recommendations of educational measurement and assessment experts.

In a survey of 893 teachers in 34 schools, Bol, Stephenson, O'Connell, and Nunnery (1998) investigated teachers' frequent uses of traditional and alternative assessment methods in relation to teaching experience, grade level, and subject area. The traditional methods of assessment included close-ended examinations, quizzes, and other written assignments. The alternative methods of assessment included performance and observation-based assessment methods. Results showed that the most experienced teachers indicated the use of alternative assessment more often than the least experienced teachers. Also, elementary school teachers reported using alternative assessment more often than did middle and high school teachers. Mathematics teachers reported using traditional assessment methods much less frequently than did teachers in other subject areas. In a related study, Snow-Renner (1998) examined teachers' assessment practices in Colorado classrooms relative to students' opportunities to learn. Survey responses were received from 737 mathematics and science teachers in grades 4, 8, and 10 as well as from 116 elementary school and 223 secondary school students. Elementary school teachers reported a greater emphasis on alternative assessments than did secondary school teachers. Students in different classrooms reported experiencing differential opportunities to learn relative to reform-oriented assessments, suggesting that teachers' assessment practices may affect students' perceptions of the classroom environment and that this effect may vary across classrooms. Snow-Renner (1998) attributed such results to fluctuations in teacher's capacity and knowledge about assessment and to ambiguous policy definitions of assessment reforms in Colorado.

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. Also, Alsarimi (2000) found that the teachers indicated using four main sources of information when assigning grades to students: final exams, midterm exams, class participation, and oral questioning. Also, these same teachers tended to incorporate some non-achievement factors such as student's effort in grading. The teachers commented that the grades reflect student improvement, effort, and knowledge of the subject matter. Recently, Zhang and Burry-Stock (2003) surveyed 297 teachers across grade levels and content areas about their classroom assessment practices. They found that mathematics and science teachers reported grading on non-achievement factors more frequently than did teachers in social studies and non-academic subjects.

The aforementioned studies tend to confirm that classroom assessment practices may be unique from one grade level, teaching experience, and subject area to another. It is also evident from the classroom assessment literature that there seems to be some contradictions between teachers' practices and recommendations of educational assessment experts. Therefore, teachers' assessment skills need a considerable scrutiny.

Due to the increasing importance to adequately prepare teachers for the task of classroom assessment, the present study was guided by the following primary research question: How do teachers' self-perceived assessment skills vary with teachers' gender, subject area, grade level, teaching experience, and in-service assessment training?

METHODOLOGY

a- Participants and Data Collection Process

The participants in this study were 213 teachers teaching grades six, eight, and ten in Muscat public schools in Oman. Permission was requested from Ministry of Education and

school principals to collect data from the teachers. The participants were informed that a study was being conducted to investigate teachers' classroom assessment skills. 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 instrument along with brief instructions about the information that was requested in the instrument, how to respond to the items, and where to find directions that were also included both on the cover letter and the instrument. Table 1 presents characteristics of the participants.

Table 1. *Characteristics of the participants*

Variable	<i>n</i>	%
Gender		
Male	100	46.9
Female	113	53.1
Subject area		
English language	86	40.4
Fine arts	55	25.8
Science	72	33.8
Grade level		
Grade 6	49	23
Grade 8	89	41.8
Grade 10	75	35.2
Teaching experience		
1 – 5 years	71	33.3
6 – 10 years	69	32.4
> 10 years	73	34.3
Have taken in-service assessment training?		
No	133	62.4
Yes	80	37.6

b- Instrument

The participants were asked to indicate their gender, subject area, grade level, teaching experience, and whether or not they have an in-service assessment training. In addition, informed by the literature (e.g., Alkharusi, 2009; Zhang & Burry-Stock, 2003), a 25-item Self-Perceived Assessment Skills Scale was developed and used in this study. 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 factor-analyzed with a principal-components method of extraction and a varimax orthogonal rotation. The analysis yielded a five-factor structure: (Analyzing Test Items, Communicating Assessment Results, Writing Achievement Test Items, Using Performance Assessment, and Grading) accounting for 64% of the total variance. Table 2 presents the items, the factor loadings, the percent of the variance explained, and Cronbach alpha reliability for each factor.

Table 2. 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.	.87				
2. Calculating and interpreting variability measures of the test scores.	.87				
3. Verifying content validity of the test items.	.80				
4. Establishing reliability of the test scores.	.79				
5. Conducting item analysis (i.e., difficulty and discrimination) for the test.	.50				
6. Providing oral and written feedback to students.		.77			
7. Communicating assessment results to students.		.71			
8. Communicating assessment results to parents.		.71			
9. Communicating assessment results to other educators.		.71			
10. Writing multiple-choice questions.			.77		
11. Writing matching questions.			.76		
12. Writing true-false questions.			.73		
13. Writing fill-in-the-blank and short-answer questions.			.73		
14. Writing essay questions.			.60		
15. Assessing students through observations.				.74	
16. Evaluating oral questions from students.				.73	
17. Defining a rating scale for performance assessment criteria in advance.				.66	
18. Assessing individual and group hands-on activities.				.55	
19. Matching performance tasks to course instruction and objectives.				.54	
20. Communicating performance assessment criteria to students in advance.				.49	
21. Informing students in advance how grades are to be assigned.					.75
22. Weighing differently projects, exams, homework, etc. when assigning semester grades.					.72
23. Using systemic grading procedures to determine borderline grades.					.64
24. Incorporating non-achievement factors (e.g., effort, classroom behavior, attendance, etc.) in the calculation of grades.					.56
25. Developing a systematic grading procedure.					.51
% of variance explained	15%	13%	13%	12%	11%
Reliability	.88	.80	.84	.82	.80

Note. F1 = analyzing test items. F2 = communicating assessment results. F3 = writing test items. F4 = using performance assessment. F5 = grading.

c- Data Analysis

Prior to the analysis, all variables were examined for accuracy of data entry and missing values. Independent t-tests were performed to explore differences in self-perceived assessment skills with respect to gender and in-service assessment training. One-way analyses of variance (ANOVA) were employed to investigate differences in the self-perceived assessment skills with respect to subject area, grade level, and teaching experience. Once it was determined statistically significant differences exist among the groups' means, Scheffe test was used to determine which means differ.

RESULTS and DISCUSSION

a- Assessment Skills by Gender

Table 3 summarizes results of the independent samples t-tests on gender differences in the self-perceived assessment skills. As shown in Table 3, female teachers reported on average a higher level of self-perceived assessment skills in communicating assessment results and writing test items than male teachers. However, there were no statistically significant gender differences in the self-perceived assessment skills in analyzing test items,

using performance assessment, and grading. Although these results are in disagreement with studies investigating teachers' assessment practices (Alsarimi, 2000; Mertler, 1998; 1999), a qualitative approach of investigation might shed more light on gender differences in self-perceived assessment skills.

Table 3. Results of the independent samples *t*-tests on gender differences in the self-perceived assessment skills

Variable	Females (<i>n</i> = 113)		Males (<i>n</i> = 100)		<i>t</i>	Cohen's <i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Analyzing test items	3.25	.97	3.30	1.02	.40	-
Communicating assessment results	3.78	.69	3.48	.88	2.84**	.38
Writing test items	3.75	.76	3.50	.92	2.16*	.30
Using performance assessment	4.31	.63	4.15	.71	1.68	-
Grading	3.96	.78	3.94	.79	.20	-

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

b- Assessment Skills by Subject area

Table 4 displays means and standard deviations of the teachers' self-perceived assessment skills by subject area. Results revealed that there were statistically significant differences across subject areas in analyzing test items, $F(2, 210) = 9.51, p < .001, \eta^2 = .08$; communicating assessment results, $F(2, 210) = 16.21, p < .001, \eta^2 = .13$; writing test items, $F(2, 210) = 6.27, p < .01, \eta^2 = .06$; using performance assessment, $F(2, 210) = 17.74, p < .001, \eta^2 = .15$; and grading, $F(2, 210) = 6.65, p < .01, \eta^2 = .06$. Scheffe's test indicated that science teachers reported on average higher levels of self-perceived skilfulness than English language teachers and fine arts teachers in analyzing test items, writing test items, using performance assessment, and grading. Also, on average, English language teachers reported a lower level of self-perceived assessment skills than fine arts teachers and science teachers in communicating assessment results. However, there were no statistically significant differences between English language teachers and fine arts teachers in the self-perceived assessment skills in analyzing test items, writing test items, using performance assessment, and grading. Also, there were no statistically significant differences between science teachers and fine arts teachers in the self-perceived assessment skills in communicating assessment results. These results lend support to the previous research findings that teachers' assessment practices differ across subject areas (Stiggins & Conklin, 1992; Zhang & Burry-Stock, 2003). The results imply a need to tailor classroom assessment training to suit the different needs of teachers based on the content area they teach.

Table 4. Means and standard deviations of the self-perceived assessment skills by subject area

Variable	English (<i>n</i> = 86)		Fine arts (<i>n</i> = 55)		Science (<i>n</i> = 72)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Analyzing test items	3.08	1.01	3.05	.92	3.67	.92
Communicating assessment results	3.29	.93	3.77	.57	3.95	.62
Writing test items	3.47	.97	3.52	.77	3.91	.68
Using performance assessment	4.01	.73	4.13	.54	4.58	.53
Grading	3.80	.82	3.82	.68	4.21	.75

c- Assessment Skills by Grade level

Table 5 displays means and standard deviations of the teachers' self-perceived assessment skills by grade level. Results revealed that there were statistically significant differences across grade levels in analyzing test items, $F(2, 210) = 4.27, p < .05, \eta^2 = .04$; communicating assessment results, $F(2, 210) = 5.60, p < .01, \eta^2 = .05$; using performance assessment, $F(2, 210) = 9.07, p < .001, \eta^2 = .08$; and grading, $F(2, 210) = 3.11, p < .05, \eta^2 = .03$. Scheffe's test indicated that grade six teachers reported on average higher levels of self-perceived skilfulness than grade 10 teachers in analyzing test items and grading. Also, on average, grade six teachers reported a higher level of self-perceived assessment skills in using performance assessment than both grade eight teachers and grade 10 years. Moreover, grade 10 teachers reported on average a higher level of self-perceived skilfulness than grade eight teachers in communicating assessment results. These results are consistent with studies examining teachers' assessment practices (Bol et al., 1998; Zhang & Burry-Stock, 2003). The results imply that classroom assessment courses in teacher education programs might need to be matched with what teachers need to know for classroom practice delineated by the specific grade level.

Table 5. Means and standard deviations of the self-perceived assessment skills by grade level

Variable	Grade 6 (n = 49)		Grade 8 (n = 89)		Grade 10 (n = 75)	
	M	SD	M	SD	M	SD
Analyzing test items	3.60	.93	3.26	1.01	3.07	.97
Communicating assessment results	3.76	.54	3.43	.95	3.81	.69
Writing test items	3.63	.58	3.62	.99	3.64	.83
Using performance assessment	4.57	.55	4.19	.72	4.07	.61
Grading	4.18	.76	3.92	.82	3.83	.74

d- Assessment Skills by Teaching Experience

Table 6 displays means and standard deviations of the teachers' self-perceived assessment skills by teaching experience. Results revealed that there were statistically significant differences across teaching experience in analyzing test items, $F(2, 210) = 9.69, p < .001, \eta^2 = .08$; communicating assessment results, $F(2, 210) = 30.08, p < .001, \eta^2 = .22$; writing test items, $F(2, 210) = 9.80, p < .001, \eta^2 = .09$; using performance assessment, $F(2, 210) = 17.39, p < .001, \eta^2 = .14$; and grading, $F(2, 210) = 7.32, p < .001, \eta^2 = .07$. Scheffe's test indicated that teachers with more than 10 years of teaching experience reported on average higher levels of self-perceived skilfulness in analyzing test items, communicating assessment results, writing test items, using performance assessment, and grading than both teachers with 1 to 5 years of teaching experience and teachers with 6 to 10 years of teaching experience. Also, on average, teachers with 6 to 10 years of teaching experience reported a higher level of self-perceived assessment skills in communicating assessment results than with teachers with 1 to 5 years of teaching experience. However, there were no statistically significant differences between teachers with 1 to 5 years of teaching experience and teachers with 6 to 10 years of teaching experience in the self-perceived assessment skills in analyzing test items, writing test items, using performance assessment, and grading. Although these results disagree with previous studies of teachers' assessment practices (Alsarimi, 2000; Mertler, 1998, 1999), the current study results highlight the importance of teaching experience in that, assessment skills might best be mastered through practice and classroom experience.

As such, classroom assessment courses in teacher education programs might need to be integrated with teaching practicum.

Table 6. Means and standard deviations of the self-perceived assessment skills by teaching experience

Variable	1 – 5 years (n = 71)		6 – 10 years (n = 69)		> 10 years (n = 73)	
	M	SD	M	SD	M	SD
Analyzing test items	3.03	1.02	3.10	.93	3.67	.91
Communicating assessment results	3.11	.88	3.87	.60	3.94	.62
Writing test items	3.32	.95	3.66	.81	3.92	.68
Using performance assessment	4.01	.73	4.10	.59	4.58	.53
Grading	3.76	.82	3.86	.71	4.22	.74

e- Assessment Skills by In-service Assessment Training

Table 7 summarizes results of the independent samples t-tests on differences in the self-perceived assessment skills with respect to the in-service assessment training. As shown in Table 7, teachers who had received in-service assessment training perceived themselves to be more skilled than those without in-service assessment training in analyzing test items, communicating assessment results, writing test items, using performance assessment, and grading. Like Zhang and Burry-Stock's (2003) study findings, the present study findings emphasize that continuous assessment training for in-service teachers might have a significant impact on their assessment skills.

Table 7. Results of the independent samples t-tests on the differences in the self-perceived assessment skills with respect to the in-service assessment training

Variable	With training (n = 80)		Without training (n = 133)		t	Cohen's d
	M	SD	M	SD		
Analyzing test items	3.66	.95	3.04	.95	4.59***	.65
Communicating assessment results	3.95	.60	3.45	.84	4.66***	.68
Writing test items	3.95	.67	3.44	.89	4.45***	.65
Using performance assessment	4.58	.53	4.03	.66	6.24***	.92
Grading	4.22	.72	3.79	.78	4.01***	.57

*** $p < .001$.

CONCLUSION

It has been estimated that teachers spend as much as a third to a half of their professional time in classroom assessment activities ranging from designing assessment tasks to grading, communicating assessment results to their respective audiences (Stiggins & Conklin, 1988, 1992). With classroom assessment taking its place as a major component in the educational process, the quality of teaching and learning relies in part on teachers' assessment skills (Daniel & King, 1998). Due to the increasing importance of the professional preparation of teachers in classroom assessment, the present study aimed at investigating teachers' self-perceived assessment skills as a function of teachers' gender, subject area, grade level, teaching experience, and in-service assessment training.

The results point to the following conclusions: (a) female teachers perceived themselves to be more skilful in writing test items and communicating assessment results; (b) teachers self-perceived assessment skills are reflective of the nature of the subjects and grade levels they teach; (c) as teaching experience increases, teachers self-perceived assessment skills tend

to increase; and (d) teachers with in-service assessment training showed a higher level of self-perceived assessment skills than those without in-service assessment training. These results suggest that male teachers might need more attention regarding their assessment skills in writing test items and communicating assessment results. Also, the findings clearly highlight the value of teachers' training in classroom assessment, and the need to match this training with the nature of classroom assessment delineated by subject areas and grade levels. It should be noted that the results of this study are limited by the use of a self-report survey and the participating sample. Further studies may use classroom observations to analyze teachers' assessment practices. Also, the survey should be sent to a more representative sample across the country.

REFERENCES

- Alkharusi, H. (2009). Correlates of Teacher Education Students' Academic Performance in an Educational Measurement Course. *International Journal of Learning, 16*, 1-15.
- 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.
- Bol, L., Stephenson, P. L., O'Connell, A. A., & Nunnery, J. A. (1998). Influence of experience, grade level, and subject area on teachers' assessment practices. *The Journal of Educational Research, 91*, 323 – 330.
- Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education, 7*(4), 279 – 301.
- Daniel, L. G., & King, D. A. (1998). Knowledge and use of testing and measurement literacy of elementary and secondary teachers. *Journal of Educational Research, 91*, 331 – 343.
- Gallagher, D., Jo. (1998). *Classroom assessment for teachers*. Upper Saddle River, NJ: Merrill.
- Gronlund, N. E. (1998). *Assessment of student achievement* (6th ed.). Boston: Pearson.
- Gronlund, N. E. (2006). *Assessment of student achievement* (8th ed.). Boston: Pearson.
- Gullickson, A. R. (1982, November). *The practice of testing in elementary and secondary schools*. Vermillion, SD: University of South Dakota. (ERIC Document Reproduction Service No. ED229391)
- Harlen, W., & Crick, R. D. (2003). Testing and motivation for learning. *Assessment in Education: Principles, Policy & Practice, 10*, 169 – 207.
- Hills, J. R. (1991). Apathy concerning grading and testing. *Phi Delta Kappan, 72*, 540 – 545.
- Mertler, C. A. (1998, October). *Classroom assessment practices of Ohio teachers*. Paper presented at the meeting of the Mid-Western Educational Research Association, Chicago, IL.
- Mertler, C. A. (1999, October). *Teachers' (Mis)conceptions of classroom test validity and reliability*. Paper presented at the meeting of the Mid-Western Educational Research Association, Chicago, IL.
- Mertler, C. A. (2003, October). *Preservice versus in-service teachers' assessment literacy: Does classroom experience make a difference?* Paper presented at the meeting of the Mid-Western Educational Research Association, Columbus, OH.
- Snow-Renner, R. (1998, April). *Mathematics assessment practices in Colorado classrooms: Implications about variations in capacity and students' opportunities to learn*. Paper presented at the meeting of the American Educational Research Association, San Diego, CA.
- Stiggins, R. J., & Conklin, N. F. (1988, November). *Teacher training in assessment*. Portland, OR: Northwest Regional Educational Laboratory. (ERIC Document Reproduction Service No. ED303439)
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.
- Stipek, D. (2002). *Motivation to learn: Integrating theory and practice* (4th ed.). Boston: Allyn and Bacon.
- Zhang, Z., & Burry-Stock, J. A. (1994). *Assessment Practices Inventory*. Tuscaloosa, AL: The University of Alabama.