Development and Validation of Assessing Quality Teaching Rubrics

Weiyun Chen[1], Steve Mason[1], Austin Hammond-Bennett[1], & Sandy Zlamout[1]

[1]School of Kinesiology, University of Michigan, Michigan, USA

Correspondence: Weiyun Chen, School of Kinesiology, University of Michigan, Ann Arbor, MI 48109, USA

Abstract

Purpose: This study aimed at examining the psychometric properties of the Assessing Quality Teaching Rubric (AQTR) that was designed to assess in-service teachers' quality levels of teaching practices in daily lessons.

Methods: 45 physical education lessons taught by nine physical education teachers to students in grades K-5 were videotaped.   They were all Caucasians (5 females and 4 males) with their teaching experience ranging from 6 years to over 20 years. Four investigators coded the taped lessons using AQTR assessment sheet.

Results: The total scale and the four sub-scales of the AQTR had satisfactory Cronbach's alpha coefficients. The results of t-test and the MANOVA revealed that the AQTR was a valid instrument to differentiate quality levels of the Overall Quality Teaching, Task Design, Task Presentation, Management, and Instructional Response among each teacher, and between the above-average and the below-average groups. The confirmatory factor analysis further confirmed the establishment of the construct validity of the AQTR.

Conclusions: The AQTR was a reliable and valid measure that can be used to assess in-service teachers' quality levels of teaching practices

Keywords: Quality Teaching Assessment, Essential Dimensions of Quality Teaching

1. Introduction

1.1 Theoretical Framework

Standards-based educational reform has increasingly called for improvement of quality teaching for more than two decades. Empirical studies confirmed that quality teaching is pivotal to affecting what K-12 students learn, are able to do, and value. Failure or success of educational reform rests directly on the extent to which teachers provide students with quality teaching practices in classrooms (Ball & Forzani, 2009; Fenstermacher & Richardson, 2005; Grossman & McDonald, 2008; Rink, 2006). Quality teaching implies that content should be developmentally appropriate and academically challenging, task presentation should be relevant and meaningful to students, class organization should be productive and supportive, and instructional guidance of students' learning should be engaging and interactive in daily lessons (Cohen, Raudenbush, & Ball, 2003; Fenstermacher & Richardson, 2005; Hill, Blunk, Charalambous, Lewis, Phelps, Sleep, & Ball, 2008).

Teaching practices reflect how the teacher interacts with students centering on subject matter in situated learning environments. Teaching practices are the "showcase" of displaying what the teacher values, knows, and is able to do in daily lessons (Fenstermacher & Richardson, 2005; Hill et al., 2008). Four core essential dimensions of teaching practices in daily lessons have been identified across subject areas and grade levels. They consist of what to teach, how to present information, how to organize the class, and how to guide students' learning. These four essential dimensions of teaching practices are intertwined in daily classrooms and collectively contribute to quality of teaching (Ball & Forzani, 2009; Cohen, et al., 2003; Fenstermacher & Richardson, 2005; Lampter & Graziani, 2009; Rink, 2003, 2006; Shulman, 2004).

Task design is one essential dimension of teaching practices. Task design refers to what learning tasks the teacher plans and organizes for their students to do (Ball & Forzani, 2009; Gore, 2001; Lampert, 2010; Rink, 2006; Shulman, 2004). To help students understand essential ideas of content, make connections among concepts, and accomplish intended lesson objectives, it is a teacher's responsibility to use knowledge of content and students to design and organize learning tasks that are sequentially progressive, developmentally appropriate, and maximally engaging. These critical

elements contribute to the quality of task design and organization (Ball & Forzani, 2009; Cohen et al., 2003; Hill et al., 2008; Rink, 2006).

Task presentation is another essential dimension of teaching practices. It deals with how to present and explain learning tasks to students (Ball et al., 2009; Charalambos, 2010; Hill et al., 2008; Lampert, 2010; Rink, 2006; Shulman, 2004). The way a teacher presents the learning task directly influences how students learn, understand, and interpret content. Scholars contend that to help students gain accurate information, the teacher should use terminology correctly and present concepts precisely and clearly. The teacher should use appropriate examples, analogies, metaphors, and students' familiar language to explain new concepts and topics in order to help students make connections and find information relevant to them. To help students understand the key ideas of the learning task, the teacher should model small steps and/or critical features of a learning task while presenting focused and relevant learning cues. These are critical elements of quality task presentation (Ball et al., 2009; Chen, 2001; Chen & Rovegno, 2000; Cohen et al., 2003; Hill et al., 2008; Lampert, 2010).

Class management is one essential dimension of teaching practices. It implies how a teacher organizes learning resources and teaching materials, groups students, organizes spaces, arranges activity formations, and distributes and collects equipment to create a productive and cooperative learning environment for students (Ball et al., 2009; Cohen et al., 2003; Reynolds, 1992; Rink, 2006). To help students work together cooperatively and productively, the teacher should use effective class organization strategies to appropriately pair students up, put students into groups, and group students into teams to provide students with opportunities to work with others. To help maximize learning time, the teacher should smoothly move students from class organization to students' task engagement. In other words, a teacher's efficiency in grouping students, organizing physical learning materials /equipment, arranging physical layouts, and locating students into working areas are critical features of the effective class organization and management (Ball et al., 2009; Ball & Forzani, 2009; Cohen et al., 2003)

Instructional response is one essential dimension of teaching practices. Students' task engagement takes place in the contextual interaction of the content, students, the teacher, and the learning resources (Ball et al., 2009; Lampert & Graziani, 2009; Rink, 2006; Shulman, 2004). Researchers stress that the teacher should closely monitor the task engagement of both individual students and the class as a whole to help the teacher know what's going on in classrooms. To help students become active and productive learners, the teacher should constantly observe and analyze the quality of students' performance and their approaches to tackling problems. Based on what is observed and analyzed, the teacher decides when to guide students to elaborate on and refine their task performance and when to allow more time for students to solve their own problems. When finding a majority of students having difficulty understanding and performing the task, the teacher should quickly adjust conditions and complexities of the task, re-explain and demonstrate a correct way or divergent ways to perform the task if necessary. Teacher should decide when to provide general feedback and when to provide specific feedback to help students maintain the quality of task engagement. These critical elements of quality instructional responses help students successfully and productively accomplish the task (Chen, 2001; Chen & Rovegno, 2000; Cohen et al., 2003; Hill et al., 2008; Lampert & Graziani, 2009; Rink, 2006; Shulman, 2004).

1.2 Purpose of the Research

The four essential dimensions of teaching practices with critical components provide a core framework for assessing the quality of teaching practices in situated classrooms (Chen, 2001; Chen & Rovegno, 2000; Cohen et al., 2003; Hill et al., 2008; Lampert & Graziani, 2009; Rink, 2006; Shulman, 2004). With an increased focus on the improvement of quality teaching, researchers stress that there is an urgent need to develop and design a classroom-based observational measure that focuses on assessing how the teacher actually teaches specific content to students in daily lessons. To this end, this study was to extend the study by (Chen, Hendricks, & Archibald, 2011) who designed and validated the AQTR with a sample of pre-service teachers' teaching. The purpose of this study was to examine the psychometric properties of the AQTR in order to determine if it can be used to assess how well the in-service teacher implemented quality teaching practices in situated classrooms. Validating the AQTR with in-service teachers aims at providing common defined criteria in relation to essential practices of quality teaching. The significance of this study lies in providing policy makers, administrators, teachers, and researchers with classroom-based assessment tool for them to conduct an administrator's evaluation, a peer-assessment, and/or a teacher's self-assessment using common language, although teaching is content and context relational and specific. As a result, teachers, administrators, and researchers would use discernible information to inform and improve teaching and learning.

2. Methods

2.1 Research Participants and Settings

Nine elementary physical education teachers and 983 students in K-5 at nine different elementary schools in the same

school district located in a suburban area in the Midwest of the United States voluntarily participated in this study. The nine teachers were five females and four males with teaching experience ranging from 6 years to 26 years.

The reason for recruiting the teachers and their students into this study was because they were voluntarily involved in a three-year Physical Education Program (PEP) grant project funded by US Department of Education. The teachers underwent same curricula, instructions, and assessment training for implementation of the PEP grant project. The teachers were informed to videotape their teaching four physical education lessons per semester based on their own teaching schedule during the second and third PEP grant project year for the purpose of assessing their teaching practices. This study was a part of the PEP grant project. The teachers and the students' parents/guardians signed the informed consent form to indicate their voluntary participation in this study. The students who did not turned in the signed consent form or whose parents/guardians did not grant their approval of this study were assigned to a specific working area by their teachers for not being videotaped. The university institutional review board granted the permission for conducting this study, as well as the school district administrator.

The student population was pre-dominantly White (92%). Students in grades K-2 had a 30-minute PE class and a 30-minute wellness class per week, while students in grades 3-5 had a 60-minute PE class per week. All names in this study are pseudonyms.

2.2 Assessing Quality Teaching Rubrics (AQTR)

Chen et al. (2011) designed the AQTR as an observational rubric to assess pre-service teachers' teaching practices that are associated with quality teaching practices in physical education contexts. The four essential dimensions of teaching practices grounded in research on teaching within different research paradigms were used as the essential dimensions of the AQTR including Task Design, Task Presentation, Management, and Instructional Response with 17 subsumed teaching components. In Task Design, there are three teaching components: Developmental Appropriateness, Maximum Participation, and Progression. Within Task Presentation, there are five teaching components: Clarity and Accuracy, Linking Prior Knowledge, Demonstration, Learning Cues, and Checking for Understanding. For Management, four teaching components included: Gaining Attention, Equipment Distribution, Grouping Students, and Transition. Regarding Instructional Response, there are five critical components listed: Monitoring, Adjusting/Re-emphasizing the Task, Reflections, General Feedback, and Specific Feedback.

The performance indicator of each teaching component was defined on a 3-point rating scale to identify a graduation of the quality of teaching practices. For example, a rating of "3" indicated that the teacher fully demonstrated the criteria of quality teaching practices in each teaching component. A rating of "2" indicated the teacher in some degree demonstrated the criteria of quality teaching practices. A rating of "1" indicated that the teacher did not demonstrate the criteria of quality teaching practices. Also an "n/a" indicated that the specific teaching component was not applicable to a given teaching episode.

To help evaluators assess the teacher's teaching practices in a live lesson or a videotaped lesson objectively, Chen et al. (2011) designed the AQTR Assessment Sheet. The teaching components of the four essential teaching dimensions on the AQTR Assessment Sheet were organized task by task. Chen et al. (2011) used the AQTR to assess 21 videotaped lessons taught by pre-service teachers. The results indicated that the AQTR established ecological and construct validity and had a high degree of inter-rater reliability and internal consistency.

2.3 Data Collection

2.3.1 Videotaping Lessons

Forty-five physical education lessons taught by the nine teachers to 983 students in K-5 were videotaped by the first author throughout three academic semesters. Prior to the videotaping, the investigator asked the teachers to choose their preferred lessons and date to be videotaped on the doodle meeting calendar during each of the three semesters in order to follow the teachers' regular physical education curricula. During the first semester, the investigators videotaped 14 lessons taught by the teachers to their students in grades 1-5. Throughout the second semester, the investigators videotaped 16 lessons taught by the teachers to their students in grades 1-5. In the third semester, the investigators videotaped 15 lessons taught by the teachers to their students in grades 1-5.

During the videotaping of a lesson, a camcorder was placed in an unobtrusive corner of the gymnasium to avoid interfering with the teaching. The teacher wore a wireless microphone throughout the lesson. The voice transmitter was attached to the digital camcorder in order to capture the teacher's and the students' voices. The camcorder's angles were constantly adjusted and zoomed in and out to make sure the teacher and their students were in view. The lesson was videotaped when the teacher started his/her teaching and the videotaping was stopped when the teacher dismissed the class.

2.3.2 Coding Taped Lessons

Prior to officially coding the 45 videotaped lessons, four investigators spent about 15 hours studying the AQTR and its

coding protocols, and practicing the observing and coding of four videotaped lessons which were randomly selected from the pool of the 45 videotaped lessons. After being familiar with the performance indicators of each teaching component, the rating scales, and the coding protocols, four investigators who were paired-up began to code three randomly selected taped lessons. While watching each taped lesson together, each pair independently coded each taped lesson with the AQTR Assessment Sheet to check the inter-rater reliability (IR). The IR of the coded lessons was examined by checking each investigator's coding results using the formula: % IR = [numbers of agreement ÷ (numbers of agreement + numbers of disagreement)] * 100 (Van der Mars, 1989). According to the formula, the IR of the first coded lessons was 82.4%; the IR of the second coded lesson was 84.5%; and the IR of the third coded lessons was 94%. The IR of the three coded lessons was all above 80% (Van der Mars, 1989).

Subsequently, four investigators began to officially code the 45 videotaped lessons with the AQTR assessment sheet using the coding protocols described by Chen et al. (2011). They watched each taped lesson together, but each pair independently coded each taped lesson. They coded a total of 128 learning tasks (teaching episodes) and 2,176 teaching behaviors using the AQTR across the 45 taped lessons.

2.4 Data Analysis

To determine the reliability of the AQTR, Cronbach alpha correlation coefficients and Pearson correlation coefficients were used to analyze the coding data. To analyze the construct validity of the AQTR, a confirmatory factor analysis was conducted using LISREL 8.8 to examine if the proposed four-factor measurement model of the AQTR fit the sample data.

Next, descriptive statistics, independent t-test, and MANOVA were utilized to examine if the AQTR can be used to distinguish the quality of teaching practices between the below-average group and the above-average groups. The two groups were classified based on the mean score of the total scale of the AQTR. The standardized-difference effect size (Cohen's d) (Trusty, Thompson, & Petrocelli, 2004) was used to report the mean differences of the dependent variables between the two groups.

3. Results

3.1 Reliability of the AQTR

Table 1 presents the mean scores, standard deviations, and Cronbach alpha correlation coefficients of the four essential dimensions (four sub-scales) and the Overall Quality Teaching (the total scale).

Table 1. Descriptive Statistics and alpha coefficients of the Four Essential Dimensions and Overall Quality Teaching

	Min.	*Max.*	*M*	*SD*	*α*
Task Design	1.00	3.00	2.86	.382	.75
Presentation	1.00	3.00	2.65	.497	.78
Management	1.50	3.00	2.85	.306	.70
Responses	1.00	3.00	2.31	.596	.79
Total scale	1.58	3.00	2.67	.339	.89

The total scale of the AQTR was labeled as Overall Quality Teaching because it represented the sum of the four sub-scales (four essential dimensions) and provided a comprehensive view of the quality of overall teaching practices. The Cronbach alpha correlation coefficient of the total scale was .89 and the alpha correlation coefficients of the four sub-scales ranged from .70 to .79. The results indicated that the AQTR had a high degree of internal consistency (Stevens, 2002).

Table 2 presents the Pearson correlations among the four essential dimensions (sub-scales) and Overall Quality Teaching (the total scale) of the AQTR.

Table 2. Pearson Correlations between the Sub-Scales and the Total-Scale

	Task Design	Presentation	Management	Response	Total-Scale
Task Design	1				
Presentation	.42**	1			
Management	.35**	.34**	1		
Response	.47**	.56**	.30**	1	
Total-Scale	.72**	.81**	.58**	.84**	1

Note: ** represents $p < .01$

Results of the bivariate correlation coefficients between the four sub-scales indicated that associations between Task Design and Task Presentation, Task Design and Instructional Response, Task Presentation and Instructional Response were strongly linked to each other at $p < .01$ level. Management was moderately correlated with Task Design, Task Presentation, and Instructional Response at $p < .01$ level. Each sub-scale was strongly correlated with Overall Quality Teaching, the total scale at $p < .01$ level.

3.2 Construct Validity of the AQTR

3.2.1 Confirmatory Factor Analysis

To examine if the proposed 4-factor measurement model fit the observed sample, this study used the multiple goodness-of-fit indices including (a) Comparative Fit Index (CFI), (b) Non-Normed Fit Index (NNFI), and (c), Incremental Fit Index (IFI), (d) Root Mean Square Error of Approximation (RMSEA), and (e) Standardized Root Mean Square Residual (SRMR) (Meyer, Gamst, & Guarino, 2006). Values of NNFI, CFI, and IFI greater than .95 are considered an excellent fit to the data (Meyers et al., 2006). In this study, the goodness fit indices were .96 of NNFI, .97 of CFI, and .97 of IFI. They were greater than .95 and close to 1, indicating an excellent model fit to the observed data. The RMSEA and SRMR are the "badness-of-fit" indices. The value of RMSEA and SRMR is less than .05 indicating an excellent model fit, and the value between .05 and .08 reflects an acceptably good fit (Kline, 2005). In this study, value of RMSEA was .07 and of SRMR was .06, indicating a good model fit to the observed data. Overall, the results of this study indicated good model fit to the observed data (Kline, 2005) and supported the construct validity of the AQTR.

3.2.2 Differences of Quality Teaching Practices among Each Teacher's Teaching

To determine if the AQTR can be used to differentiate the quality of each teacher's teaching practices, a one-way multivariate analysis of variance (MANOVA) was conducted on the mean scores of the total scale and the four sub-scales in the AQTR. Descriptive statistics of the four essential dimensions and the Overall Quality Teaching of each teacher's teaching are presented in Table 3.

Table 3. Descriptive Statistics of Four Sub-Scales and the Total Scale Among the Nine Teachers

Teachers	Mean	SD	# of Tasks
Sub-Scale 1: Task Design			
Rebecca	2.99	.07	23
Craig	2.84	.44	17
Ron	2.98	.08	18
Linda	2.77	.37	11
Sheryl	3.00	.00	10
Betty	2.57	.48	11
John	3.00	.00	10
Dan	2.66	.67	17
Emily	2.88	.40	11
Sub-Scale 2: Task Presentation			
Rebecca	2.83	.34	23
Craig	2.62	.44	17
Ron	2.20	.40	18
Linda	2.62	.39	11
Sheryl	2.89	.24	10
Betty	2.13	.85	11
John	3.00	.00	10
Dan	2.28	.54	17
Emily	2.65	.50	11
Sub-Scale 3: Management			
Rebecca	3.00	.00	23
Craig	2.68	.42	17
Ron	2.96	.18	18
Linda	2.91	.17	11
Sheryl	3.00	.00	10
Betty	2.60	.47	11
John	2.83	.31	10
Dan	2.81	.31	17

Emily	2.75	.35	11
Sub-Scale 4: Instructional Response			
Rebecca	2.53	.40	23
Craig	2.00	.73	17
Ron	2.24	.59	18
Linda	2.24	.49	11
Sheryl	2.79	.20	10
Betty	2.11	.69	11
John	2.80	.23	10
Dan	1.90	.45	17
Emily	2.31	.60	11
Total - Scale: Overall Quality Teaching			
Rebecca	2.84	.17	23
Craig	2.53	.43	17
Ron	2.70	.19	18
Linda	2.63	.27	11
Sheryl	2.92	.09	10
Betty	2.35	.54	11
John	2.91	.11	10
Dan	2.41	.31	17
Emily	2.76	.25	11

The results of the MANOVA analysis yielded a significant main effect of Overall Quality Teaching among each teacher's teaching practices ($Wilk's$ $\lambda = .48$, $F_{1, 8} = 2.93$, $p < .01$). Subsequently, the results of MANOVA revealed significant differences on Task Design, Task Presentation, Management, and Instructional Response among each teacher's teaching practices ($F = 2.59$, $df = 8$, $p < .01$; $F_8 = 5.89$, $df = 8$, $p < .01$; $F_8 = 3.67$, $df = 8$, $p < .01$; and $F_8 = 4.91$, $df = 8$, $p < .01$). The results indicated that the AQTR was a valid instrument to differentiate the quality of each teacher's teaching practices on the four essential dimensions including Task Design, Task Presentation, Management, Instructional Response, and Overall Quality Teaching.

3.2.3 Differences of Quality Teaching Practices between Two Groups

To further examine the construct validity of the AQTR, the mean score of Overall Quality Teaching was used to divide into two groups. The mean score of Overall Quality Teaching was 2.67 with a standard deviation of .34 for the total sample of 128 teaching episodes. The mean scores of Overall Quality Teaching greater than 2.67 were classified into group 1 (above-average group), while the mean scores of Overall Quality Teaching lower than 2.67 was categorized group 2 (below-average group). Table 4 shows the descriptive statistics and effect sizes of the four sub-scales and the total scale of the AQTR between the two groups.

Table 4. Descriptive Statistics of the Four Sub-Scales and the Total Scale between Two Groups

	Above-Average Group	Below-Average Group	
	M (SD)	M (SD)	Cohen's d
Task Design	2.97 (.17)	2.72 (.51)	.66
Task Presentation	2.82 (.32)	2.42 (.59)	.84
Management	2.93 (.22)	2.73 (.37)	.66
Instructional Response	2.52 (.50)	2.04 (.60)	.87
Overall Quality Teaching	2.81 (.19)	2.48 (.40)	1.1

The t-test was used to determine if the AQTR could be used to differentiate the teachers' overall quality teaching practices between the two groups. The t-test yielded a significant difference on the mean scores of Overall Quality Teaching between the two groups (Mean $_{above-average}$ = 2.81 vs. Mean $_{below-average}$ = 2.48, $t = 6.225$, $df = 126$, $p < .001$, Cohen's $d = 1.05$). The results of t-test indicated that the AQTR was a valid instrument to distinguish levels of the overall quality teaching between the two groups.

The MANOVA was used to examine if there were significant differences on four dependent variables including Task Design, Task Presentation, Management, and Instructional Response between the two groups. The results of MANOVA revealed a significant main effect of Overall Quality Teaching between the two groups ($\lambda = .76$, $F = 9.515$, $p < .001$). Subsequently, the MANOVA revealed significant differences on Task Design, Task Presentation, Management, and Instructional Response between the two groups ($F_1 = 15.61$, $p < .01$; $F_1 = 24.59$, $p < .01$; $F_1 = 11.77$, $p < .01$; $F_1 = $

24.43, $p < .01$). The results indicated that the AQTR could be used to distinguish the quality of the four essential dimensions and the overall quality teaching between the above-average group and the below-average group.

4. Discussion and Implications

This study aimed at examining the psychometric properties of the AQTR that could provide administrators, teachers, and researchers with a validated classroom-based observation rubric to assess the extent to which teachers demonstrated the quality of teaching practices during a live lesson and/or a videotaped lesson. The discussion of the results is organized into two parts: (a) validation of the AQTR, and (b) implications of the AQTR.

4.1 Validation of the AQTR

4.1.1 Internal Consistency

Internal consistency "measures whether several items that propose to measure the same general construct produce similar scores" (Wikipedia, 2012). In line with the previous findings grounded in research on teaching (e.g. Fenstermacher & Richardson, 2005, Lampter & Graziani, 2009; Reynolds, 1992, Shulman, 2004), .89 alpha correlation coefficient of the total scale in the AQTR indicated that the 17 teaching components within the four essential dimensions measured the construct of Overall Quality Teaching. Consistent with the previous research findings (Ball & Forzani, 2009; Rink, 2003, 2006; Shulman, 2004) and NASPE's (2009) appropriate instructional practices guidelines, .75 alpha correlation coefficient of Task Design sub-scale indicated that three critical teaching components including Developmentally Appropriate and Challenging Tasks, Maximally Engaging Tasks, and Progressively Sequential Tasks were all essential items used to measure the latent variable of Task Design. Similarly, .78 alpha correlation coefficient of Task Presentation sub-scale illustrated that five teaching components consisting of Clarity and Accuracy of Task Presentation, Linking to Prior Knowledge, Demonstration, Learning Cues, and Checking for Understanding were all teaching elements proposed to measure the construct of Task Presentation. Likewise, .70 alpha correlation coefficient of Management sub-scale indicated that four teaching components comprising of Keeping Attention, Equipment Distribution/Returning, Grouping Students, and Transition were key items used to measure the construct of Management. Lastly, .79 alpha correlation coefficient of Instructional Response revealed that five teaching components including Monitoring the Class, Adjusting/Re-emphasize the Task, Positive/General Feedback, Specific Performance-related Feedback, and Reflections were key items utilized to measure the latent variable of Instructional Response. The results of this study indicated that the 17 teaching components represented critical features and essential characteristics of quality teaching practices. How well the teacher implemented the quality of Task Design, Task Presentation, Management, and Instructional Responses depends on how well the teacher demonstrated desirable features of each teaching component in classroom situations (Lampter & Graziani, 2009; Reynolds, 1992; Rink, 2003, 2006; Shulman, 2004).

4.1.2 Correlations

Consistent with the previous findings of research on teaching (Reynolds, 1992; Shulman, 2004), the Pearson correlation coefficients of the four sub-scales of the AQTR yielded positively and moderately strong relationship between each of the sub-scale. The results indicated the four essential dimensions are interrelated to each other though each represents a unique dimension of quality teaching practices. This study was congruent with the previous empirical findings that the four essential dimensions of teaching practices independently but collectively influenced the extent to which the overall quality teaching could be enacted in classrooms (Ball & Forzani, 2009; Cohen, et al., 2003; Fenstermacher & Richardson, 2005; Gore, 2001; Grant & Gillette, 2006; Lampter & Graziani, 2009; Reynolds, 1992; Rink, 2003, 2006; Shulman, 2004).

Examination of Pearson correlation coefficients between each sub-scale and the total scale of the AQTR revealed a unique relationship between the four essential dimensions and the Overall Quality Teaching. The results indicated that Instructional Responses was most strongly associated with the Overall Quality Teaching total scale, followed by Task Presentation, and Task Design. In contrast, Management exhibited moderately strong relationship with the Overall Quality Teaching total scale. Interestingly, Chen, Mason, Staniszewski, Upton and Valley (2012) examined levels of each teacher's demonstration of quality teaching in terms of the four essential dimensions and the overall quality teaching. They found that when the teacher did not demonstrate the quality of Instructional Response but fully demonstrated the quality of Task Design, and mostly demonstrated the quality of Task Presentation and Management, his overall quality teaching was categorized into the level of "Partially Demonstrated." The results suggest that the teacher's demonstration of quality Instructional Response contributed most to the overall quality of teaching practices though the other three dimensions were important also. In short, the results of internal consistency confirmed that 17-item AQTR captured essential features of the quality teaching practices in terms of the four essential dimensions. The bivariate correlations further confirmed that the four sub-scales of the AQTR reflected essential constructs of the quality of teaching practices and interrelated relationship among them.

4.1.3 Construct Validity

Supporting the findings above, the results of the goodness-of-fit indices indicated that four sub-scales of the AQTR confirmed a priori theoretical constructs of Task Design, Task Presentation, Management, and Instructional Response underlying the quality of teaching practices. The results of the confirmatory factor analysis provided a psychometrical support for the priori factorial structure of the AQTR that captured the essential dimensions of quality teaching practices. Each essential dimension of the AQTR represented a unique theoretical construct of and an integral part of quality teaching practices.

Landing a support for previous findings of research on teaching (Reynolds, 1992; Shulman, 2004), in this study, the Task Design essential dimension focuses on assessing the extent to which the teacher designed academically rigorous and developmentally appropriate learning tasks for students to learn and to work on, how well the teacher broke down the learning tasks into learnable pieces and organized them into a sequence so that each task was built on each other, and the degree to which the tasks designed by the teacher provided students with maximum engagement (Ball & Forzani, 2009; Hill et al., 2008; NASPE, 2009; Rink, 2006; Shulman, 2004). The Task Presentation essential dimension focuses on measuring the extent to which the teacher used key words and demonstrations, related students' prior knowledge, and provided students' familiar scenarios to accurately present new information to students, and the degree to which the teacher used various teaching strategies to check for students' understanding of the task (Hill et al., 2008; Lampter & Graziani, 2009; Reynolds, 1992; Rink, 2003, 2006; Shulman, 2004). The Management essential dimension was central to assessing how efficiently the teacher used effective methods and instructional routines to organize students, equipment/teaching materials, and space for engaging in the tasks (Ball & Forzani, 2009; Cohen, et al., 2003; Fenstermacher & Richardson, 2005; NASPE, 2009; Rink, 2006). The Instructional Response essential dimension aimed at assessing the extent to which the teacher discerned students' emerging problems and analyzed the quality of task performance, how effectively the teacher provided tailored and specific guidance, re-adjusted complexity and conditions of the learning task, and engaged students in active and responsible learning processes (Hill et al., 2008; Lampter & Graziani, 2009; Shulman, 2004). The results of this study indicated that the four essential dimensions in the AQTR captured the core theoretical constructs of quality teaching practices grounded from research on teaching. Each of the four essential dimensions played a critical role in contributing to the quality of teaching practices in classrooms (Ball & Forzani, 2009; Cohen, et al., 2003; Fenstermacher & Richardson, 2005; Gore, 2001; Grant & Gillette, 2006; Lampter & Graziani, 2009; Reynolds, 1992; Rink, 2003, 2006; Shulman, 2004).

Consistent with the study by Chen et al. (2012), the results of MANOVA in this study indicated that the Overall Quality Teaching, the total scale, in the AQTR was a valid construct to distinguish quality levels of an individual teacher's teaching practices. Furthermore, the results of MANOVA indicated that Task Design, Task Presentation, Management, and Instructional Response four sub-scales in the AQTR were valid essential dimensions to gauge different levels of quality teaching practices demonstrated by an individual teacher while teaching their actual lessons.

Extending the study by Chen et al. (2011) who examined whether the AQTR can be used to differentiate levels of the pre-service teachers' quality teaching practices between the high- and low-quality teaching groups, this study used the mean score of the Overall Quality Teaching total scale to classify teaching episodes into the above-average group and the below-average group. The results of the *t*-test yielded a significant difference of the Overall Quality Teaching between the groups. Subsequently, the results of MANOVA indicated that the quality of Task Design, Task Presentation, Management, and Instructional Response of the above-average group were significantly higher than the below-average group. The results indicated that the AQTR was a valid observation-based instrument to differentiate the levels of quality teaching practices among each in-service teacher's teaching and between the two groups. The four sub-scales of the AQTR provided a reliable and validated assessment tool for administrators, teachers, and researchers to evaluate in-service teachers' teaching in their daily lessons.

4.2 Implications for Future Teaching and Research

The ultimate goal of standards-based educational reform is to improve the quality of teaching practices. Developing and designing a reliable and validated assessment instrument that measures if the teachers implemented critical features of quality teaching in their daily lessons is a stepping stone helping to accomplish the central mission of the standards-based movement (Ball et al., 2009; Fenstermacher & Richardson, 2005; Grossman & McDonald, 2008). Ball et al. (2009) and Grossman and McDonald (2008) stress that there are critical needs to design an assessment instrument that describes core facets of quality teaching practices. Grossman and McDonald (2008) pointed out that "one direction for research on teaching would be to continue the search to identify such 'common factors' in teaching that are critical for success" (p. 187). Researchers argued that no matter what specific subject content is being taught, the teacher's core works consist of designing and presenting the tasks for their students, organizing the students for completing the tasks, and responding to what the students are doing and saying (Ball et al, 2009; Fenstermacher & Richardson, 2005).

Grossman and McDonald (2008) emphasized that the instrument with a core framework for the quality of teaching provides a "common language" for identifying various levels of quality teaching means and what desired quality teaching looks like in classrooms. Meeting this pragmatic need, one significant contribution of this study is that the four essential dimensions of the AQTR represent multiple facets of quality teaching practices that are generic to teaching across subject areas. This study suggests that the four essential dimensions of the AQTR provide a shared theoretical framework for administrators, teachers, teacher educators, and researchers to evaluate the extent to which the in-service teachers demonstrated the quality of Task Design, Task Presentation, Management, and Instructional Response in classrooms.

Researchers noted that there has been increasing call for use of observation-based assessments to evaluate the quality of teaching in classrooms (Fenstermacher & Richardson, 2005; Mangiante, 2011). This study is significant in meeting that critical need and suggests that the AQTR could be used as an observational tool to assess the quality of teachers' teaching practices both in the field of physical education and general education. The results of this study suggest that the critical teaching components of the four essential teaching dimensions in the AQTR adequately described the various levels of critical features of quality teaching practices across various subject areas. In addition, with an increasing recognition of the needs for conducting an administrator's evaluation, peer evaluation, and the teacher's self-evaluation, this study serves this need because the AQTR provides a common language for both school administrators and teachers to communicate about what a teacher is doing in classrooms reflects the characteristics of quality teaching. With the valid and reliable observation-based rubrics, school administrators may provide specific feedback about the strengths and weaknesses of the teacher's teaching practices based on what they observed. The AQTR also provides teachers with an observational instrument to conduct a peer-assessment of their peer's teaching. Teachers also may use the AQTR as the self-assessment tool to self-evaluate what aspects of teaching they fully demonstrated and/or what aspects of teaching they need to improve most after they complete their lessons during or after a school day.

Although the performance indicators of each critical teaching component in the AQTR are specific to the context of physical education teaching, the four essential dimensions in the AQTR represent essential dimensions of quality teaching practices and are generic to teaching across subject areas. This study suggests that researchers and teachers may use the four essential dimensions of the AQTR as the core framework to modify, add, and/or delete some teaching components based on specific teaching contexts. Accordingly, researchers and teachers may describe the performance indicators of the corresponding teaching components modified and added to the AQTR in order to better adapt to the contextual needs of specific subject area and teaching situations. Besides that, this study suggests that researchers could use the AQTR for future studies. For example, researchers could use the AQTR to compare and contrast the quality levels of teachers' teaching practices among different school districts to provide research-based information for policy makers. They also could use the AQTR to investigate the association between the quality levels of teachers' teaching and different levels of students' achievement.   In conjunction with qualitative research methods, researchers could use the AQTR to assess the quality levels of teachers' teaching within the context of descriptive and rich classroom environments. Researchers also could conduct further validation studies using broad samples of teachers from K-12 public schools across various subject areas. In short, AQTR with the sound psychometric properties is an observational assessment tool for researchers to use in multiple ways.

References

Ball, D. L., & Forzani, F. M. (2009). The work of teaching and the challenge for teacher education. *Journal of Teacher Education, 60*, 497-511. http://dx.doi.org/10.1177/0022487109348479.

Ball, D. L., Sleep, L., Boerst, T. A., & Bass, H. (2009). Combining the development of practice and the practice of development in teacher education. *The Elementary School Journal, 109,* 458-474. http://dx.doi.org/10.1086/596996.

Charalambos, Y. C. (2010). Mathematical knowledge for teaching and task unfolding: An exploratory study. *The Elementary School Journal, 110*, 247-278. http://dx.doi.org/10.1086/648978.

Chen, W. (2001). Description of an expert teacher's constructivist-oriented teaching: Engaging students' critical thinking in learning creative dance. *Research Quarterly for Exercise and Sport, 72,* 366-375. http://dx.doi.org/10.1080/02701367.2001.10608973.

Chen, W., & Rovegno, I. (2000). Examination of expert and novice teachers' constructivist-oriented teaching practices using a movement approach to elementary physical education. *Research Quarterly for Exercise and Sport, 71*, 357-372. http://dx.doi.org/10.1080/02701367.2000.10608919

Chen, W., Hendricks, K., & Archibald, K. (2011) Assessing pre-service teachers' quality teaching practices. *Educational Research and Evaluation. 17*, 13-32. http://dx.doi.org/10.1080/13803611,.2011.578008

Chen, W., Mason, S., Staniszewski, C., Upton, A., & Valley, M. (2012). Assessing the quality of teachers' teaching practices. *Educational Assessment, Evaluation and Accountability, 24*, 25-41. http://dx.doi.org/10.1007/s11092-011-9134-2.

Cohen, D. K., Raudenbush, S. W., & Ball, D. L. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis, 25*, 119-142.

Fenstermacher, G. D., & Richardson, V. (2005). On making determinants of quality in teaching. *Teachers College Record, 107*, 186-213.

Gore, J. M. (2001). Beyond our differences: A reassembling of what matters in teacher education. *Journal of Teacher Education, 52,* 124-135. http://dx.doi.org/10.1177/0022487101052002004.

Grant, C. A., & Gillette, M. (2006). A candid talk to teacher educators about effectively preparing teachers who can teach everyone's children. *Journal of Teacher Education, 57*, 292-299. http://dx.doi.org/10.1177/0022487105285894

Grossman, P., & McDonald, M. (2008). Back to the future: Directions for research in teaching and teacher education. *American Educational Research Journal, 45,* 184-205. http://dx.doi.org/10.3102/000283120 7312906.

Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L., & Ball, D. L. (2008). Mathematical knowledge for teaching and the mathematical   quality of   instruction: An exploratory study. *Cognition and Instruction, 26*, 430-511. http://dx.doi.org/10.1080/07370000802177235.

Kline, R. B. (2005). *Principles and practice of structural equation model* (2nd ed.). New York: Guilford.

Lampert, M. (2010). Learning teaching in, from, and fro practice: What do we mean? *Journal of Teacher Education, 60*, 21-34. http://dx.doi.org/10.1177/0022487109347321.

Lampert, M., & Graziani, F. (2009). Instructional activities as a tool for teachers and teacher educators' learning. *The Elementary School Journal, 109*, 491-509. http://dx.doi.org/10.1086/596998.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research-Design and interpretation.* Thousand Oaks, CA: Sage.

National Association for Sport and Physical Education [NASPE] (2009). *Appropriate instructional practice guidelines for elementary school physical education*. Reston, VA: NASPE Publications.

Rink, J. E. (2003). Effective instruction in physical education. In S. J. Silverman and C. D. Ennis (Ed.), *Student learning in physical education: Applying research to enhance instruction* (2nd edition) (pp. 165-186). Champaign, IL: Human Kinetics.

Rink, J. E. (2006). *Teaching physical education for learning* (5th Ed.). Boston, MA, McGraw Hill.

Shulman, L. S. (2004). The wisdom of practice : Essays on teaching, learning, and learning to teach. San Francisco, CA: Jossey-Bass.

Stevens, J. (2002). *Applied multivariate statistics for the social sciences* (4th ed.).   Hillsdale, NJ: Lawrence Erlbaum.

Van der Mars, H. (1989). Observer reliability: Issues and procedures. In P. Darst, D.

Zakrajsek, & V. Mancini (Eds.), *Analyzing physical education and sport instruction* (2nd ed., pp. 53-80). Champaign, IL: Human Kinetics.