© Journal of the Indian Academy of Applied Psychology
July 2021, Vol. 47, No. 2, 175 - 183

Construct Validation of the Perceptions of Assessment Tasks Inventory

Hussain Ali Alkharusi
Sultan Qaboos University, College of Education, Oman

Students' perceptions of the assessment tasks can be a basis for optimizing the classroom assessment environment. In an attempt to quantify students' perceptions of assessment, Dorman and Knightley (2006) developed Perceptions of Assessment Tasks Inventory (PATI). This study sought to provide evidence for the construct validity of the PATI. A descriptive research design was employed in this study. Using a cluster sampling process, the participants were 330 Omani students (175 females and 155 males) enrolled in grade 11 science classes. Three instruments were used in the study concerning self-efficacy, perceptions of classroom assessment environment, and perceptions of assessment tasks. Data were analyzed using Confirmatory factor analysis, cluster analysis, multivariate analysis of variance, one-way analysis of variance, and Pearson product-moment correlation analysis. Results indicated that a five-factor structure of the PATI provided a good fit to the observed data. Cluster analyses suggested three clusters of students who were distinct with regard to their perceptions of the assessment tasks. Scores of the PATI subscales correlated positively with perceived learning assessment environment and negatively with perceived performance assessment environment. The PATI scores had acceptable levels of internal consistency reliability. In conclusion, the results support the utility of the PATI for research and instructional purposes..

Keywords: Educational assessment; learning; validity; reliability; confirmatory factor analysis; cluster analysis.

Classroom assessment is one of the important components of the educational process. It affects student motivational beliefs, academic achievement, and learning (Black & Wiliam, 1998; Brookhart, 2013; Harlen & Crick, 2003). As such, its quality has consistently been a concern for educators (Black & Wiliam, 1998; Stiggins, 1992). One strategy to assess the quality of the classroom assessment is through students' perceptions of the assessment, which refer to the meaning that students make out of the assessment practices in the classroom (Brookhart & DeVoge, 1999).

Students' perceptions of the assessment can be a basis for optimizing the classroom assessment environment (Struyven, Dochy, & Janssens, 2002). Alkharusi (2011) described how classroom assessment environment can be differentiated into learning-oriented and performance-oriented environments based on students' perceptions. The perceived learning-oriented assessment environment involves asking students meaningful assessment tasks with moderate difficulty, giving them opportunities to improve their performance, and providing them informative assessment feedback. The perceived performance-oriented assessment environment involves asking students difficult and less meaningful assessment tasks with harsh assessment standards and criteria, emphasizing the importance of grades more than learning, and comparing students' performances normatively. Several research studies have addressed how perceived classroom assessment environment influence academic achievement-related outcomes (e.g., Alkharusi, 2011; Alkharusi, Aldhafri, Alnabhani, & Alkalbani, 2013; Brookhart, Walsh, & Zientarski, 2004; Watering, Gijbels, Docy, & Rijt, 2008). As such, measurement of students' perceptions of the assessment should deserve more recognition and investigation.

In an attempt to quantify students' perceptions of assessment, Dorman and

Knightley (2006) developed Perceptions of Assessment Tasks Inventory (PATI) consisting of 40 items to measure students' perceptions of the assessment tasks along five dimensions: congruence with planned learning, authenticity, student consultation, transparency, and diversity. Congruence with planned learning refers to the extent to which students perceive that the assessment tasks align with the subject's learning objectives and activities. Authenticity refers to the extent to which students perceive that the assessment tasks are relevant to their real-life situations. Student consultation refers to the extent to which students are involved and consulted in the assessment process. Transparency refers to the extent to which students are clearly informed about the purposes and forms of the assessment. Diversity refers to the extent to which students perceive that they can complete the assessment tasks at their own speed.

The PATI has been used in several research studies. For example, Dhindsa, Omar, and Waldrip (2007) found that secondary students in Brunei tended to hold weak positive perceptions of the assessment tasks as measured by PATI. Also, Dhindsa et al. (2007) found statistically significant gender differences in perceived congruence with planned learning favoring females and perceived student consultation in assessment favoring males. They concluded that students' perceptions of assessment could be improved by enhancing teachers' practices. Gao (2012) found that classroom assessments in mathematics as perceived by high school students tended to have a strong congruence with instruction, adequate transparency, inadequate authenticity, little student consultation, and diversity. Also, statistically significant gender differences were found in perceived assessment authenticity and transparency favoring females. Gao (2012) called for more research on students' perceptions of classroom assessment in other subject areas and grade levels.

Alkharusi et al. (2013b) found that middle school students in Oman tended to hold positive perceptions of their classroom assessment tasks as measured by PATI. In addition, congruence with planned learning, student consultation, and transparency were significantly associated with students' academic achievement whereas authenticity and diversity did not contribute significantly to the variance explained in academic achievement. In another study, Alkharusi et al. (2013a) found that students' perceptions of assessment tasks had direct and indirect effects on academic self-efficacy and perceptions of task value. Alkharusi et al. (2013a) concluded that teachers should develop assessment tasks in ways that activate students' sense of efficacy to accomplish the tasks and enhance their value of engaging in the tasks. Likewise, van Dinther, Dochy, Segers, and Braeken (2014) asserted that students' perceptions of assessment can be enhanced by requiring students to do real-life tasks and giving them informative feedback, which are likely to foster their self-efficacy.

Furthermore, Alkharusi (2013) found that high degrees of perceived authenticity, transparency, and diversity in assessment tasks tended to be associated with deep learning strategies adopted by the students. Also, high degrees of congruence with planned learning and low degrees of task authenticity tended to be associated with surface learning strategies. In addition, Alkharusi et al. (2014) found that in comparison to a performance-oriented assessment environment, a learning-oriented assessment environment tended to be associated with high degrees of perceived congruence with instruction, authenticity, student consultation, and diversity.

Although the aforementioned studies suggest promise for the PATI in the educational assessment literature, validation of its scores is needed for accurate interpretations of students' perceptions of the assessment tasks. Construct validation of the PATI is crucial especially when it is used in a different educational context in terms of language, subject, and grade level from which it was originally developed. Thus, the current study aimed at examining the construct validity of the PATI based on data collected from secondary school students in the Sultanate of Oman. Validation is carried out through evidence derived from internal structure and relationships to other variables.

Purposes of the Study

If students' perceptions of the assessment tasks are important for a sound classroom assessment, their measure should yield valid and reliable scores. The PATI was widely used to measure students' perceptions of the assessment tasks. None of the studies utilizing the PATI have statistically examined its psychometric properties. Therefore, the purposes of the current study were to:

1. Present evidence regarding the construct validity of the PATI's scores by investigating the confirmation of its dimensionality using confirmatory factor analysis (CFA) and by establishing clusters of students based on their perceptions of the assessment tasks.
2. Provide evidence of convergent and discriminant validity for the PATI's scores in terms of their correlations with other constructs.
3. Provide data regarding the internal consistency of the PATI's scores.

Method

Sample

The participants in this study were 330 Omani students (175 females and 155 males) enrolled in grade 11 science classes. They were selected using a cluster sampling process from public schools across all governorates in the Sultanate of Oman. Their ages ranged from 15 to 19 years with an average of 17 and a standard deviation of .67.

Instruments

The instrument used was a self-report questionnaire with four main sections: Basic information, self-efficacy, perceptions of classroom assessment environment, and perceptions of assessment tasks. The basic information of the questionnaire covered gender and age. The questionnaire items were translated into Arabic language and subjected to a content validation process done by a panel of five 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 appropriateness of each item for the use with the targeted participants and its relevance to the construct being measured. Their feedback was used for refinement of the items. Internal consistency reliability was established using Cronbach's alpha. Following is a description of the three sections.

Self-efficacy: This section of the questionnaire included eight items measuring students' self-efficacy from the Motivated Strategies for Learning Questionnaire (MSLQ) developed by Pintrich, Smith, Garcia, and Mckeachie (1993). Responses were obtained on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). In this study, internal consistency coefficient was .76 as measured by Cronbach's alpha. An individual student's self-efficacy level was represented by an average rating score across the items.

Perceptions of classroom assessment environment: This section of the questionnaire included 18 items of the Perceived Classroom Assessment Environment Scale developed by Alkharusi (2011). This scale measures students' perceptions of the classroom assessment environment on two dimensions. The first dimension was learning-oriented assessment environment comprising 9 items (e.g., "In this class, students are given a chance to correct their mistakes"). The second dimension was performance-oriented assessment environment comprising 9 items (e.g., "In this class, the teacher compares students' performances to each other"). Responses were obtained on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). In this study, internal consistency coefficients as measured by Cronbach's alpha were .75 for the first dimension and .65 for the second dimension. Each dimension was constructed by averaging the scores of its corresponding items.

Perceptions of assessment tasks: This section of the questionnaire included 35 items of the Dorman and Knightley's (2006) Perceptions of Assessment Tasks Inventory (PATI). The items measure students' perceptions of assessment tasks in terms of congruence with planned learning, authenticity, student consultation, transparency, and diversity. Each dimension consisted of 7 items. Responses

were obtained on a 5-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). Internal consistency coefficients will be reported in the next section after validating the structure of the items. Each dimension was constructed by averaging the scores of its corresponding items.

Procedure

Permission was requested from Ministry of Education and school principals to collect data from the students during regular scheduled science classes. The students were informed that a study about their perceptions of the classroom assessment tasks is being conducted. They were informed that they were not obligated to participate in the study, and if they wished to participate, their responses would remain confidential. They were also told that participation in the study would not influence their grades or relations with the teacher in any way. Students who wished to participate were asked to respond to a self-report questionnaire, which will be described in the next section of this study. The administration of the questionnaire took about one class period, and was preceded by a brief set of instructions about how to complete the questionnaire.

Statistical Analyses

In relation to the aforementioned purposes of the study, the following statistical procedures were employed:

1. The means, standard deviations, skewness, and kurtosis values for the 35 items on the PATI were reported for descriptive purposes. In addition, alpha-if-item deleted index was computed to determine whether the score reliability of the scale would increase or decrease if the item was removed from the analysis. Also, item-total correlation coefficients were calculated to examine how well each item contributes to the measurement of the construct represented by the scale (Crocker & Algina, 1986).

2. Construct validity of the PATI was investigated by examining the confirmation of the dimensionality of the PATI using confirmatory factor analysis (CAF) with maximum likelihood estimation in EQS 6.1. The analysis was conducted using the covariance matrix. Each item was constrained to load only on its hypothesized factor. One item on each factor was constrained to equal one in order to set a metric for the factors. Factor covariances were left free to be estimated, but the measurement errors were not allowed to covary. During the analysis, items that resulted in large standardized residuals and having a small amount of proportion of explained variance ($R^2 < .10$) were identified for possible deletion. Items were deleted one at a time in order to study changes in the factor structure solution. The purpose of this item deletion was to develop the best fitted and the most parsimonious model to the data. In evaluating the fit of the model, recommendations by Schermelleh-Engel, Moosbrugger, and Müller (2003) were followed. These recommendations state that for an acceptable model fit, the ratio χ^2/df should be less than or equal to 3, the Root Mean Square Error of Approximation (RMSEA) should be less than or equal to .08, the Nonnormed Fit Index (NNFI) that is also called the Tuker-Lewis Index (TLI) should be greater than or equal to .95, and the Comparative Fit Index (CFI) should be greater than or equal to .95 (Schermelleh-Engel et al., 2003).

3. In an attempt to provide an additional evidence of construct validity, differentiation between groups was established through Ward's hierarchical clustering technique. This technique was performed on the data to establish clusters of students based on their perceptions of the assessment tasks. Dendrograms were used to determine the appropriate number of meaningful clusters represented in the data. To verify that the selected cluster solution separated the cluster groups, a multivariate analysis of variance (MANOVA) was performed on the data using the PATI subscales as dependent variables and the cluster membership as the independent variable. Two additional analyses were used to ensure the validity of the emergent clusters. First, a one-way analysis of variance (ANOVA) was conducted to examine how the clusters differed on an external criterion (i.e., self-efficacy). Second, a discriminant function analysis was used to test whether each member of a particular cluster had the right configuration of the perceptions of assessment tasks expected of that cluster.

4. Convergent and discriminant validity were

examined by computing the Pearson product-moment correlation coefficients between scores of PATI's subscales and scores of perceived learning- and performance-oriented assessment environments.

5. Reliability of the PATI's scores was evaluated using Cronbach's alpha.

Results

Descriptive Statistics

The means, standard deviations, skewness, kurtosis, alpha-if-item deleted, and item-total correlation values for the 35 items on the PATI were calculated to examine quality of the items. Item means generally indicated a moderate ceiling effect. The skewness and kurtosis values for all PATI items were within acceptable range of ± 1.96 (George & Mallery, 2001), suggesting no concern about deviation from normality. As for the inventory's internal consistency, the alpha-if-item deleted indices showed improvement in the total score reliability only by less than .10. The item-total correlation coefficients were positive and statistically significant ranging from .20 to .60, suggesting that each item contributes positively to the measurement of students' perceptions of assessment tasks. These results lead to retain all items at this point of the analysis.

Construct Validity

Dimensionality of the PATI: In order to validate the dimensional construct of the PATI, a structural model consisting of five correlated factors was tested: Congruence with planned learning, authenticity, student consultation, transparency, and diversity. Confirmatory factor analysis was performed to test the theoretical model. Table 1 shows goodness-of-fit indices. Improvements were made to the model as a function of these results.

According to Table 1, the χ^2/df falls within the acceptable values. The NNFI and CFI values approached the minimum criterion for good fit (.90). The RMSEA values were acceptable by being less than .08. In order to improve the model, items with $R^2 < .10$ in the standardized solution were removed. These items were item 4 "My assignments are related to what I am learning in science" from congruence with planned learning, item 17 "I am asked about the types of assessment I would like to have in science" from student consultation, item 23 "I am told in advance when I am being assessed" from transparency, and item 35 "When I am faster than others, I move on to new assessment tasks" from diversity.

As can be seen in Table 1, the NNFI and CFI indices improved, although they failed to reach .90. The AIC index indicates improvement in terms of model parsimony. Thus, the factorial structure of the PATI consisted of 31 items divided into five correlated factors. The first factor was congruence with planned learning and it consisted of 6 items with factor loadings ranging between .40 and .59. The second factor was authenticity and it consisted of 7 items with factor loadings ranging between .38 and .64. The third factor was student consultation and it consisted of 6 items with factor loadings ranging between .35 and .59. The fourth factor was transparency and it consisted of 6 items with factor loadings ranging between .42 and .70. The fifth factor was diversity and it consisted of 6 items with factor loadings ranging between .41 and .57. The correlation coefficients between the factors were positive and significant ranging from .52 to .64.

Cluster membership. In addition, Ward's hierarchical clustering technique was performed on the data to establish clusters of students based on their perceptions of the assessment tasks: Congruence with planned learning, authenticity, student consultation, transparency, and diversity. Dendrograms based on the hierarchical cluster analysis suggested a three-cluster solution. The means and standard deviations for the three cluster groups on the clustering variables are displayed in Table 2.

Table 1. Confirmatory Factor Analysis of the Theoretical Model of the PATI

Model	χ^2	χ^2/df	*NNFI*	*CFI*	*RMSEA*	*Model AIC*
Five correlated factors	1127.65	2.05	.78	.80	.06, 90%CI = [.05,.06]	27.65
Five correlated factors without items 4, 17, 23, & 35	776.19	1.83	.85	.86	.05, 90%CI = [.05,.06]	-71.81

Table 2. Means and Standard Deviations for the Three Cluster Groups on the Clustering Variables

Variable	Cluster 1 (n = 52)		Cluster 2 (n = 181)		Cluster 3 (n = 97)	
	M	SD	M	SD	M	SD
Congruence	2.72	.61	3.53	.52	4.21	.35
Authenticity	2.72	.56	3.59	.39	4.30	.34
Student consultation	2.54	.61	3.22	.55	3.92	.43
Transparency	2.71	.58	3.66	.46	4.43	.38
Diversity	2.42	.57	3.15	.62	3.88	.53

As shown in Table 2, students in Cluster 1 had on average the least positive perceptions of the assessment tasks. By contrast, students in Cluster 3 had on average the most positive perceptions of the assessment tasks. Students in Cluster 2 scored between Clusters 1 and 2 on all variables.

To understand the characteristics of the three cluster groups, a χ^2-test analysis was conducted to compare the distributions of the three cluster groups in terms of gender (male/female). Results revealed a statistically significant difference between the three cluster groups on the distributions of gender, $\chi^2(2) = 17.76$, $p = .000$. More males than females were in cluster 1 whereas cluster 2 consisted of more females than males. There were no significant differences between the number of males and females in cluster 3. This suggests that females tended to have more positive perceptions about assessment tasks than males.

To verify that the selected cluster solution separated the cluster groups, a multivariate analysis of variance (MANOVA) was performed on the data using the PATI subscales as dependent variables and the cluster membership as the independent variable. The results indicated a statistically significant multivariate effect, $F(10, 646) = 71.06$, $p < .001$, $\eta^2 = .52$. The univariate effects indicated statistically significant differences between the cluster groups in the congruence with planned learning, $F(2, 327) = 78.33$, $p < .001$, $\eta^2 = .50$; authenticity, $F(2, 327) = 87.45$, $p < .001$, $\eta^2 = .62$; student consultation, $F(2, 327) = 68.07$, $p < .001$, $\eta^2 = .43$; transparency, $F(2, 327) = 102.52$, $p < .001$, $\eta^2 = .60$; and diversity, $F(2, 327) = 76.95$, $p < .001$, $\eta^2 = .41$. Post hoc analyses of the between-group differences using Scheffe's test elaborated the differences among the groups. Specifically, students of Cluster 3 had significantly higher scores on all subscales of the PATI than did all other clusters, ps < .001. Further, those in Cluster 2 scored significantly higher on all subscales than did those in Cluster 1, ps < .001.

To ensure the validity of the emergent clusters, a one-way analysis of variance (ANOVA) was conducted to examine how the clusters differed on an external criterion, which was self-efficacy. Results indicated that the clusters differed significantly on this external criterion, $F(2, 327) = 53.25$, $p < .001$, $\eta^2 = .25$. Post hoc analyses of the between-group differences using Scheffe's test elaborated the differences among the groups. Specifically, students of Cluster 3 (M = 4.20, SD = .47) scored significantly higher than those in Cluster 2 (M = 3.70, SD = .59) and those in Cluster 1 (M = 3.24, SD = .63), ps < .001. Further, those in Cluster 2 scored significantly higher than those in Cluster 1, $p < .001$.

Furthermore, a discriminant function analysis was used to test whether each member of a particular cluster had the right configuration of the perceptions of the assessment tasks expected of that cluster. Using students' perceptions of the assessment tasks in terms of congruence with planned learning, authenticity, student consultation, transparency, and diversity; it was predicted correct cluster membership for 92.4% of the cases. Predicted membership in Cluster 3 was 94.8% accurate, followed by a 91.7% accuracy rate for Cluster 2 and a 90.4% accuracy rate for Cluster 1. Calculation of Cohen's kappa revealed that the accuracy of the predicted classification was an 87%

improvement over what would be expected by chance. These findings provide evidence that the three emergent clusters are unique.

Convergent and discriminant validity. To provide evidence for convergent validity of the PATI's scores, Pearson product-moment correlation coefficients relating scores of the perceived learning-oriented assessment environment and the PATI's subscales scores were calculated. Results indicated that the scores of the perceived learning-oriented assessment environment correlated positively and significantly (ps < .001) with all PATI scales: Congruence with planned learning, r = .64; authenticity, r = .70; student consultation, r = .67; transparency, r = .65; and diversity, r = .66.

To provide evidence for discriminant validity of the PATI's scores, Pearson product-moment correlation coefficients relating scores of the perceived performance-oriented assessment environment and the PATI's subscales scores were calculated. Results indicated that the scores of the perceived performance-oriented assessment environment correlated negatively and significantly (ps < .001) with all PATI scales: Congruence with planned learning, r = -.32; authenticity, r = -.34; student consultation, r = -.23; transparency, r = -.33; and diversity, r = -.18.

Reliability

Internal consistency reliabilities for the scores of the PATI's scales were measured by Cronbach's alpha. Results showed that the internal consistency coefficients were .68 for congruence with planned learning, .72 for authenticity, .63 for student consultation, .75 for transparency, and .68 for diversity.

Discussion and Conclusion

Positive students' perceptions of the assessment tasks have become a valuable outcome for teachers' assessment practices in the classroom. This is because student positive views of the assessment tasks may contribute to an environment that is conducive for effective learning. To adequately assess these perceptions, Dorman and Knightley (2006) developed a 35-item instrument entitled Perceptions of Assessment Tasks Inventory (PATI). There has been an increased interest in utilizing the PATI to understand perceptions of the students towards the assessment tasks developed by the teacher in the classroom (e.g., Alkharusi, 2013; Alkharusi et al., 2014; Dhindsa et al., 2007). However, the data metric properties of this measure have not been explored statistically. The present study reports on the construct validity of the PATI for students enrolled in the 11th grade science classes in Oman.

The results generally supported the original five-factor structure of the PATI as designed by Dorman and Knightley (2006) thereby suggesting that the PATI reliably measures five distinct, yet related aspects of students' perceptions of assessment tasks: congruence with planned learning, authenticity, student consultation, transparency, and diversity. However, four items were statistically excluded to improve the goodness-of-fit indices of the factor model leading to the development of a modified 31-item PATI. A closer examination of the content of these items revealed that the assessment practices stated in the items might not be applicable in the educational assessment context of the 11th grade science classes in Oman. Further research is needed to cross validate the PATI in other educational assessment contexts. Also, the positive moderate correlations found among the subscales of the PATI suggest that these aspects make up the core of effective assessment practices.

Also, the results revealed three clusters of students based on their perceptions of the assessment tasks. The clusters were found to be statistically distinct with regard to the perceptions of the assessment tasks and self-efficacy. In addition, scores of the PATI subscales correlated positively with perceived learning assessment environment and negatively with perceived performance assessment environment. These results are in line with previous research confirming the notion that supportive assessment environment might result in greater student self-efficacy and learning (e.g., Alkharusi et al., 2013a; van Dinther et al., 2014). Furthermore, like past studies, the results of the current study supported gender differences with respect to the perceptions of the assessment environment favoring female students (e.g.,

Dhindsa et al., 2007; Gao, 2012). These results imply that more efforts are needed to help male students develop positive perceptions of the assessment environment. These efforts might be in terms of more alignment between features of the assessment task and student capability to do the tasks.

Overall, these results provide support for using PATI by teachers and researchers. It offers information that can be used by teachers to better orient classroom assessment toward improved student motivation and learning. Researchers might need to conduct studies linking between perceptions of assessment tasks and other characteristics of students and teachers. Another possible research area would be to cross-validate the PATI in different educational assessment settings.

References

Alkharusi, H. (2011). Development and datametric properties of a scale measuring students' perceptions of the classroom assessment environment. *International Journal of Instruction, 4,* 105 – 120.

Alkharusi, H. (2013). Canonical correlational models of students'perceptions of assessment tasks, motivational orientations and learning strategies. *International Journal of Instruction, 6*, 21 – 38.

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2013a). The impact of students' perceptions of assessment tasks on self-efficacy and perception of task value: A path analysis. *Social Behavior and Personality: An international Journal, 41*, 1681-1692. https://doi.org/10.2224/sbp.2013.41.10.1681

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2013b). The relationship between Students' Perceptions of the Classroom Assessment Tasks and academic achievement. Paper presented at the 15th Annual International Conference on Education, Athens, Greece, May 20-23, 2013. ATINER'S Conference Paper Series, No: EDU2013-0404.

Alkharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2014). Modeling the relationship between perceptions of assessment tasks and classroom assessment environment as a function of gender. *The Asia-Pacific Education Researcher, 23,* 93-104. http://dx.doi.org/10.1007/s40299-013-0090-0.

Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80,* 139 – 148.

Brookhart, S. M. (2013). Classroom assessment in the context of motivation. In J. H. McMillan (Ed.), SAGE handbook of research on classroom assessment (pp. 35 – 54). Los Angeles, CA: SAGE Publications. https://doi.org/10.4135/9781452218649.n3

Brookhart, S. M., & DeVoge, J. G. (1999). Testing a theory about the role of classroom assessment in student motivation and achievement. *Applied Measurement in Education, 12*, 409–425. https://doi.org/10.1207/S15324818AME1204_5

Brookhart, S. M., Walsh, J. M., & Zientarski, W. A. (2006). The dynamics of motivation and effort for classroom assessment in middle school science and social studies. *Applied Measurement in Education, 19*, 151–184. https://doi.org/10.1207/s15324818ame1902_5

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. Belmont, CA : Wadsworth Group/Thomson Learning.

Dhindsa, H. S., Omar, K., & Waldrip, B. (2007). Upper secondary Bruneian science students' perceptions of assessment. *International Journal of Science Education, 29*, 1261–1280. https://doi.org/10.1080/09500690600991149

Dorman, J. P., & Knightley, W. M. (2006). Development and validation of an instrument to assess secondary school students'perceptions of assessment tasks. *Educational Studies, 32,* 47–58. https://doi.org/10.1080/03055690500415951

Gao, M. (2012). Classroom assessments in mathematics: High school students' perceptions. *International Journal of Business and Social Science, 3*, 63- 68.

George, D., & Mallery, P. (2001). SPSS for windows step by step: A simple guide and reference 10.0 update (3rd ed.). Boston: Allyn & Bacon.

Harlen, W., & Deakin-Crick, R. (2003). Testing and motivation for learning. Assessment in Education: Principles, *Policy & Practice, 10*, 169 – 207. http://dx.doi.org/10.1080/0969594032000121270.

Pintrich, P. R., Smith, D. A. F., Garcia, T., Mckeachie, W. (1993). Reliability and predictive validity of the motivated strategies for learning questionnaire

(MSLQ). *Educational and Psychological Measurement, 53*, 801-813. http://dx.doi.org/10.1177/0013164493053003024.

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online, 8*, 23 – 74.

Stiggins, R. J. (1992). High quality classroom assessment: What does it really mean? Educational Measurement: *Issues and Practice, 11*, 35 – 39. https://doi.org/10.1111/j.1745-3992.1992.tb00241.x

Struyven, K., Dochy, F., & Janssens, S. (2003). Students' perceptions about new modes of assessment in higher education: A review. In M. Segers, F. Dochy, & E. Cascallar (Eds.), Optimising new modes of assessment: In search of qualities and standards (pp. 171 – 224). Dordrecht: Kluwer. http://dx.doi.org/10.1007/0-306-48125-1_8.

Van Dinther, M., Dochy, F., Segers, M., & Braeken, J. (2014). Student perceptions of assessment and student self-efficacy in competence-based education. *Educational Studies, 40,* 330-351. https://doi.org/10.1080/03055698.2014.898577

Watering, G., Gijbels, D., Dochy, F., & Rijt, J. (2008). Students'assessment preferences, perceptions of assessment and their relationships to study results. *Higher Education, 56*, 645–658. https://doi.org/10.1007/s10734-008-9116-6.

Hussain Alkharusi, Professor, College of Education, Sultan Qaboos University, Sultanate of Oman, Email: hussein393500@gmail.com