European Journal of Education Studies
ISSN: 2501 - 1111
ISSN-L: 2501 - 1111
Available on-line at: www.oapub.org/edu

doi: 10.5281/zenodo.3727259 Volume 7 | Issue 2 | 2020

A STRUCTURAL EQUATION MODELING OF THE ASSESSMENT PRACTICES INVENTORY MODIFIED (APIM) SCALE: AN ASSESSMENT MODEL FOR HIGHER EDUCATION INSTITUTIONS

Musa Matovu[i]
PhD, Islamic University in Uganda,
Department of Educational Psychology,
Faculty of Education,
P. O. Box 2555,
Mbale, Uganda

Abstract:
This research analysed the psychometric properties of the Assessment Practices Inventory Modified (APIM) using structural equation modelling. The APIM scale was distributed to a sample of 350 academic staff selected from both private and public universities in Uganda. A purely quantitative approach with a stratified random sampling technique to select the participants was adopted for this study. The objective of this study was to; test the psychometric properties of the APIM scale. The results of the psychometric properties revealed that the APIM scale is an adequate instrument in measuring assessment practices among university academic staff. The hypothesised model was found to be adequate in explaining assessment practices among academic staff in universities; $\chi^2 = 902.619$, $\chi^2/df = 1.842$, $df = 490$, $p = .000$, CFI = .906, RMSEA = .052, and SRMR = .050. The findings revealed that the APIM hypothesised model was fit and adequate to explain the factors assessment practices of staff in higher education institutions. This has led to a recommendation that formal assessment training programmes should be made mandatory to all academic staff in universities in Uganda to understand assessment design, administration, interpretation and application.

Keywords: structural equation modelling, APIM Scale, assessment model, higher education institutions

1. Introduction

Assessment is derived from a Latin word 'assidere' meaning to sit beside, or by someone (Fatmawati, 2011; MOE, 2010; Orzolek, 2006; NCCA, 2005; Weddel & Van Duzer, 1997).

---

[i] Correspondence: email matovumousa@yahoo.com

This means that during the assessment process a teacher and students get into an interaction to accomplish the learning goals (Alkharusi, 2011; Gronlund, 2006; Orzolek, 2006; Weddel & Van Duzer, 1997; Satterly, 1989). Assessment is a fundamental institutional structure that offers students' academic justice by either locking or unlocking their academic potentials (MOE, 2010; Shohamy, 2004; Arter, 2003; Black & Wiliam, 1998b; Gipps, 1998; Madaus, Raczek, & Clark, 1997). Assessment is a significant component in the teaching and learning structures whose proper integration improves the learning process (Alkharusi, 2012; NCCA, 2005; Benson, 2003; Biggs, 2003; Roeber, 2002), and the effectiveness of teachers' ways of instruction (Ong, 2010; Orzolek, 2006).

The assessment structures and practices of academic instructors have become an area of research interest to many scholars in recent times (Frazier, 2007; Zhang & Burry-Stock, 2003), with its studies taking a central position in the activities of learning and teaching (McMillan, 2003). Assessment structures and practices are the ways in which academic staff award grades, analyse them, and how they use the assessment results demonstrated by apprentices to enhance the learning process (Ainsworth & Viegut, 2006). Assessment practices to be adequate should be based on an adequate model that can enhance the assessment process uninterruptedly. According to Brown (2003), assessment structures and practices of academic instructors help them to collect, interpret, and apply the information about student learning characteristics. Assessment information helps the academic instructors in student testing, performance analysis, planning, discussion, and improving student learning based on the learning objectives (Ainsworth & Viegut, 2006; Martell & Calderon, 2005; Popper, 2005; Orlich, et al., 2001; Hindi & Miller, 2000).

2. Literature

According to various studies student assessment in higher education institutions has four major components which include; (a) designing of assessments, (b) administration of assessments, (c) interpretation of assessments, and (d) interpretation of assessments.

a. Designing of Assessments

Assessment has a great influence on student learning and if not well designed, it might fail to bring out the positive qualities of a targeted approach during the assessment process (Boud et al., 1999). The design of an assessment reveals what an assessment will achieve. In assessments designed with an aim of improving the learning of the students, they are premeditated to guide students' concentration on what requires improvement in order to use the feedback (Boud & Associates, 2010). Appropriately designed assessments exploit the methods used in an assessment which can change students' approaches to learning if projected to the curriculum objectives (Boud & Associates, 2010; Kirkwood & Price, 2008).

The main considerations when designing an assessment include; the choices of assessment, preparing examiners and students, setting standards, how the assessment will be marked as well as the interpretation of the assessment out comes (O'Grady, 2006). A well-designed assessment should be able to motivate students to learn, relate to the real-life situation, show whether students have achieved their learning goals, and should affect the students' acquired skills and competencies (Warren-Piper et al., 1996). At the designing stage of an assessment the focus should be on allowing the students to express what they were taught (O'Grady, 2006; Ebersole, 2009). Designing of assessments should be done in contemplation of how the assessment will be administered, scored, interpreted, and how different stakeholders will use the feedback. It is expected that during the assessment design, there should be increased support from the different academic staff in the subject area (Ebersole, 2009; Gore et al., 2009). Standards of assessment should be met at the assessment designing stage in order to enable the learning objectives of the course to be reflected at the end of the learning session (HEA, 2007).

b. Administration of Assessments

Good administration of assessments in higher education institutions is not just indispensable, but a prerequisite in education assessment (QAAHE, 2007). Through proper administration and management of assessments, higher education institutions become certain of their standards and mechanisms. This helps them to ensure that the standards are followed and unalleviated in the various courses or programmes offered. Administration of assessments in higher education institutions guarantees whether assessment measures and procedures are just, impartial, and reliable to the students (QAAHE, 2007). It has been highlighted that today there is reasonably less deliberation higher education institutions have done which can broaden the practices of efficient administrative approaches used in the assessment of students (Peterson et al., 1999b).

Properly designed assessment should comprise administration that caters for the whole assessment process. The administration should be able create the most favourable environment that would maintain and uphold changes in academic the staff's assessment practices and measures. This is because it helps to understand that managing assessment practices among the academic staff contributes to students' progress and performance (QAAHE, 2007). In higher education institutions teams responsible for administering and managing assessment include; academic staff, course teams, assessment boards and committees, academic managers, and the various teaching-learning staff (QAAHE, 2007). Assessment in learning comprises of several types of assessments which are administered in different ways. The major types of assessments comprise of assessment of learning which summarises a learning session and assessment for learning which is administered during the learning progress. Assessment of learning is known for the function of grading or categorising and giving feedback about student learning at the end of a study programme, or, course unit with conventional measures. Assessment for learning is used in between the course or study unit for the purposes of helping students to improve their learning, provide effective feedback, and to fully help in understanding student learning (QAAHE, 2007; Elwood & Klenowski, 2002). There should always be equilibrium in between assessment for and assessment of learning (QAAHE, 2007). With proper administration and management of assessment, assessment for learning should be inclined in the direction of assessment of learning in the assessment process (Gibbs, 2006; Black & Wiliam, 2003; Gardner, 1999; Harlen & James, 1997).

Assessment for learning is a crucial stage in the assessment process and in the pursuit for high quality performance standards which students are expected to attain in an assessment (Black & Wiliam, 1998a). Quality assurance measures encourage a chain of procedures during the assessment administration process; to either attain maximum results from assessment for learning, or, assessment of learning. It has been noted that formalisation of assessments into a progression may obscure learners' uneasiness, restrain their expression of the uneasiness, and can generate distressing situations in their studies (Crook, Gross, & Dymott, 2006). In regard to administration of assessments five elements of administrative support for assessment are identified in literature. The literature highlights institutional support strategies, assessment leadership patterns, and assessment policies and practices that have an influence on the assessment. Literature also mentions that assessment culture in institutions and evaluation of the student assessment process are part of assessment administration (Peterson et al., 1999a). It has been articulated that generally the rationale, composition, and function of student assessment endeavours in form of institutional support are vital in the administration of assessments in higher education institutions (Peterson et al., 1999a). These are planned alternatives which characterise an institution's involvement in the establishment of a robust link between its exterior and interior assessment atmospheres (Peterson et al., 1986). The overall purpose includes the reflection of the institutions' efforts in laying a robust link connecting its exterior (outside) and interior (inside) assessment atmospheres (Peterson et al., 1986). These can also be referred to as both the external and internal assessment support strategy forces reflective of the student assessment.

c. Interpretation of Assessments

After an assessment has been administered to the students, the next step is the interpretation of the assessments. The interpretation of assessment has three levels; namely (a) scoring, (b) grading, (c) and analysing of the results.

A. Scoring Assessments

In scoring assessments, there is need to know the intended learning outcomes and criteria for assessment by the assessors (QAHE, 2006). In higher education institutions where good assessment practices prevail, assessment administrators should develop policies and procedures on scoring and moderation of the assessments. According to QAHE (2006) during the scoring exercise of an assessment there should be rules governing the marking exercise. Such policies/ rules can be at the departmental level, faculty level, or for the entire university. Before and during the scoring process of an assessment, students to undertake the assessment and assessors should all know the assessment methods and guidelines that are being used to evaluate each of the assessment tasks (QAHE, 2006). The use of clear assessment tools such as marking schemes should be a key factor in ensuring that scoring assessments is consistent and fair as possible (QAHE, 2006; Quality Matters, 2006).

During the scoring process, students should be marked anonymously. Also, if the relevant expertise is available, the assessment should be marked by someone else other than the lecturer who taught the course (QAHE, 2006). As a good procedure of scoring an assessment, all candidates should be marked on one task at a time. In scoring assessments, there should be a group of examiners to mark the work for all students, and should be able to adopt a blind double marking standard for all the assessments. The use of marking trends to facilitate comparisons and evidence on standards of the assessment should also be clearly outlined (QAHE, 2006). When the academic staff are scoring assessments, flexibility at subject level may be appropriate because different courses in a university might call for different marking methods (QAHE, 2006; Ebersole, 2009). To promote fairness and flexibility in scoring assessments in different departments, faculties can be allowed to decide on the carry mark for their course programmes, but, should be within the institutions' examinations rules and regulations (QAHE, 2006).

B. Grading Assessments

Grading of assessments is a determining step in the process of assessing students' learning and determining what they have learnt (Dunca & Noonan, 2007). After scoring an assessment grading of the assessments should follow. A grade is referred to as a single number or letter which carries several functions in indicating what the students learnt. Grades are instructors' measure of what students learnt and hold students accountable for their learning (O'Connor, 2009). Grades awarded to the students after an assessment should be able to help them understand learning, track their adeptness in learning, and should avail information on the advancement in their learning towards the targeted academic goal (O'Connor, 2009). As far as assessments are concerned marks awarded are an arbitrary measure of what students have learnt and not what they have not learnt (QAHE, 2006). Assessment grades should reflect what the students learnt, and they should also have an accurate and fair representation of the actual students' achievements in that course (Luth, 2010). To make decisions on the final grades awarded to students in an assessment it should not be a one man's event, but, a group discussion of the academic staff; both those who marked and those who did not mark the assessments (QAHE, 2006; Ebersole, 2009). At the same time this can be for comparison purposes with the same subjects or other subjects that have been marked at that level or across levels.

During the grading process, the grading practices should be quite variable, ranging from pass-fail or, norm-referenced to criterion-based systems (Luth, 2010). Norm-referenced grading is when students are graded on a curve randomly or by chance with the dimensions of a normal distribution (Luth, 2010). According to Luth (2010) if the teaching instructions are effective the distribution of students' achievement should not be on a normal curve. This means that if the students' achievements approximate normal distributions we may conclude that our educational efforts were just average, that is, as some students would pass while others fail. Grading on the curve would not be appropriate in a normal learning situation because education has a purpose, and all students should learn what they have been taught which should be reflected in the grades they attain in an assessment (Guskey & Bailey, 2001; Bloom et al., 1981). According to O'Connor (2009) awarding of grades should reflect what students have attained academically with their behaviour assessed separately. Using grades as punishment to students because of their prior awful behaviours is a bad assessment practice which can discourage students from studies. This is because students can develop perceptions that they are perpetual low achievers (O'Connor, 2009).

One of the persistent debates in higher education teaching today is whether to grade "on the curve" which is known as norm-referenced grading or grade against a supreme or set standard which is referred to as criterion-referenced grading. Luth (2010) highlighted that grading on the curve is educationally dysfunctional in higher education institutions. According to different studies, it has been revealed that grading and grades in higher education institutions should be according to the students' understanding and based on a particular standard than on relative standing of the class (McKeachie & Svinicki, 2006; Luth, 2010). Using norm-referenced assessment or grading on the curve just makes students to struggle amongst themselves for high marks or positions, but, not for the sake of learning (Luth, 2010). In higher education institutions the greatest benefit of criterion-referenced assessment is that it assesses students according to their competencies. The grading or schemes of criterion-referenced assessment compel course instructors to make their assessment standards explicit before the assessment (Knight, 2004). In cases where students are assigned grades, they are assigned to them on the basis of performance standards they could have achieved on the given criteria (Connoley, 2004). Thus, student performance is compared to previously specified criteria to the assessment and not to a relative norm or other fellow students. Using criterion-referenced assessment lecturers and students can locate where students are succeeding in their learning and where they failing as well.

C. Analysing Assessment Results

Different studies suggest that before academic instructors analyse or use information obtained from assessments, they should have the requisite of assessment competencies (Paris & Hoffman, 2004; Earl, *2003*; Partridge et al., 2003; Black & Wiliam, 1998b). Such competencies can be acquired through short trainings, experience, professional development courses among others (Black & Wiliam, 1998a; Paris & Hoffman, 2004). Students and academic staff should be aware of the methods which would be used to interpret the results before the assessment exercise. Students and academic staff should know how the assessment results would affect the students' progression within a given study programme and their efforts in the learning outcomes (Black & Wiliam, 1998b). At the beginning of a given course the pass mark should be mentioned and explained to students, and they should understand the significance of each mark in their effort to complete a given course (QAHE, 2006). The interpretation of marks needs to be fair and known to students and academic staff. The pass marks in courses need to have been evaluated prior to the assessments by the institutions to establish whether they qualify to decide whether students would pass or fail, or to have attained the minimum knowledge required for that particular course (QAHE, 2006).

4. Application of Assessment Results

Results from an assessment cannot be relevant unless they are applied for the reason an assessment was undertaken. According to Peterson et al. (1999a) results from assessments in education institutions are utilised in the following ways. Assessment results can be used for educational decisions making, teaching/ learning improvement, and policy design. When the results of an assessment are applied or used, they should revise educational goals, design academic programmes, and reorganise educational programmes and departments (Peterson et al., 1999b). Assessment results can also be used to assist in modifying academic staff's assessment practices, assessment rules and regulations, improve on the education curriculum, increase field learning improve teaching methods, and to improve on other educational services and many more (Peterson et al., 1999b). Assessment results can be a condition of determining faculty promotion and faculty awards to university academic staff. Assessment results can also be used to find out whether the faculties have an impact on students' learning. The faculties' impact on students' learning can be through knowing whether results from assessments have stimulated campus discussions among students, contributed to faculty's goals, and have substance on the methods of instruction used in the learning process (Peterson et al., 1999b).

4.1 Objective of the Study

The objective of the study was;

- To test the psychometric properties of the Assessment Practices Inventory Modified (APIM) scale.

4.2 Research Hypothesis

- The hypothesised model does not explain the factors influencing assessment practices among academic staff in universities in Uganda.

5. Methods

5.1 Population, Sample and Sample Selection

The ideal population of the study comprised of all universities in Uganda and their academic staff. The universities in Uganda are clustered into public and private universities, which have academic staff at different academic levels (teaching assistants, assistant lecturers, lecturers, associate professors, and professors). The sample for this study was selected from 3 public and 2 private universities were randomly selected in Uganda. Simple random sampling was used to select the universities from their clusters on the basis of either being public or privately funded, or owned. Stratified sampling method was used to select participants from their strata which were which were designed in accordance to their specialisations and academic levels in their respective universities. The Assessment Practices Inventory Modified (APIM) was distributed to 310 academic staff in different academic levels and specialisations of the different universities in Uganda. Out of the 310 questionnaires distributed, 299 questionnaires were returned by the participants from whom 287 questionnaires were used for data analysis. 287 questionnaires were used for data analysis because they were fully filled by the respondents.

5.2 Instrumentation

The Assessment Practices Inventory Modified (APIM) scale which was tested for adequacy and also used as an instrument of data collection for this study is based primarily on Assessment Practices Inventory Revised (API_R) scale which was developed by Burry-Stock & Frazier (2008). In the APIM scale some of the items were adopted entirely, others were adapted to suit the purpose of the study, while several other items were adopted from the original Assessment Practices Inventory (API) developed by Zhang and Burry-Sock (1994) and from the Inventory of Institutional Support for Student Assessment by Peterson (1997). The Burry-Stock & Frazier (2008) API_R scale was adopted because it was the most appropriate and most recently developed assessment practices inventory which was used to collect information on teachers' perceived application of assessment practices.

The API_R which was an adopted instrument in this study originates from the API designed by Zhang and Burry-Stock (1994). The API scale was revised twice by Zhang and Burry-Stock in 1995 and 2003. The API was also revised by Burry-Stock & Frazier in 2008. The original API by Zhang & Burry-Stock (1994 - 2003) was made up of 67 items. This was developed based on the "Standard Teacher Competence in Education Assessment of Students (1990)" which outlined seven key assessment practices used in student assessment (Frazier, 2007). The seven practices included;

1) Teachers should be skilled in choosing assessment methods appropriate for instructional decisions.
2) Teachers should be skilled in developing assessment methods appropriate for instructional decisions.
3) Teachers should be skilled in administering, scoring, and interpreting the results of test set externally and those tests a teacher sets him/ herself.
4) Teachers ought to have the expertise of interpreting results of assessment in the learning decision making considering students individually, in the preparation of teaching, during the curriculum developing exercise, and in the improving of all school activities.
5) Teachers must have the skills in building suitable grading system for the students which is largely based on how the students perform.

6) Teachers should have adequate skills in reporting or giving feedback of what has been observed in an assessment to the different education stakeholders such as the students themselves, caretakers (parents), other teachers and society.
7) Teacher in the assessment process ought to have skills in scrutinising unscrupulous, dishonest, and unsuitable methods of assessment and application of the information generated from an assessment (Frazier, 2007; American Federation of Teacher et al., 1990).

When the original API was used to collect data, which was analysed on factor analysis a set of seven skills on how teachers perceived their practices and abilities in using assessments were generated. The seven assessment practices generated accounted for 61.87% variability in the assessment practices model. Among the seven new practices which were generated, five practices were similar to the assessment practices in the "Standard Teacher Competence in Education Assessment of Students (1990)", while also two new practices were generated; (a) using paper-pencil tests and (b) non-achievement-based grading. The seven practices generated using the original API included;

1) Using paper-pencil tests.
2) Standardised testing, test revision and instructional improvement.
3) Using performance assessment.
4) Communicating assessment results.
5) Non-achievement-based grading.
6) Grading and test validity.
7) Addressing ethical concerns (Frazier, 2007; Zhang & Burry-Stock, 2003).

The API scale was revised by Burry-Stock and Frazier (2008) to come up with the Assessment Practice Inventory Revised (API_R) scale. The API_R scale was used to collect data which was again analysed using factor analysis. This study generated six assessment practices instead of the seven practices which were generated using the API scale. The practices generated from the data collected by the API_R accounted for 61.91% variability in the assessment practices among teacher's secondary school (Burry-Stock and Frazier, 2008). In the six generated practices, four practices were similar to the practices generated using the API scale, while the other two practices were newly generated; (a) externally required assessment skills and (b) student-oriented grading. The assessment practices generated using the API_R scale included;

1) Teachers' Assessment Development and Application (TADA);
2) Formative Assessment (FA);
3) Teacher Oriented Grading Related Activities (TOGRA);
4) Externally Required Assessment Skill (ERAS);
5) Student Oriented Grading Related Activities (SOGRA);
6) Ethics (ETHICS).

The APIM which was an instrument used for data collection in this research consisted of 50 statements which described design, administration, interpretation, and application of assessment practices among academic staff in Ugandan universities. The items of the APIM were on a five point likert-scale, rated from not at all skilled to highly skilled (1 = *Not at all skilled*, 2 = A *little skilled*, 3 = *Some-what skilled*, 4 =*Skilled*, and 5 = *Highly skilled*). This is similar to the five-point rating likert-scale that was adopted in the API_R. The APIM questionnaire comprised of two sections. Section I asked for demographic information from the academic staff such as academic levels, types of universities, specialisations, and assessment courses attended among others. Section II of the APIM questionnaire comprised of 50 items on a scale to which participants in the study rated their abilities in designing, administering, interpreting, and applying assessment results in a learning situation.

The APIM scale adopted thirty-eight (38) items entirely from the API_R scale without modification. 12 items from the API_R were either modified (re-phrased), or dropped from the APIM scale based on the results of the pilot study which examined their suitability for use in this study. Five (5) of the affected items (Items 3, 6, 10, 17, and 27) were modified by rephrasing them to suit the context of this study, which is the university assessment. The other seven (7) items were dropped from the APIM scale due to similarity to other items, or, not being applicable in assessment at the university level. Out of the seven (7) items dropped from the APIM scale, three (3) items (Items 5, 25, and 26) were replaced with items from the Zhang and Burry-Stock (1994) original API scale, while the other four (4) items (Items 2, 31, 42, and 44) were replaced with the items from Institutional Support for Student Assessment (Peterson et al., 1999a) as in Table 1.

Table 1: Original, Modified and Replaced Items in the APIM

Rephrased Items			
No.	Original Items	Rephrased/ Modified Items	Source
Q3	• Constructing rubrics to grade performance assessment (i.e speeches, pictures, projects, etc.)	• Considering rubrics when marking/ grading students in an assessment.	Modified
Q6	• Aligning my instructional objectives to required standards (such as state, district, and testing objectives) simultaneously	• Aligning tests to university assessment standards.	Modified
Q10	• Using assessment results when developing lesson plans.	• Using previous assessment results to improve course outlines.	Modified
Q17	• Inviting students to participate in assessment development	• Involving students in assessment design and development.	Modified
Q27	• Incorporating effort into the final achievement grades	• Incorporating effort into students' final achievement grades.	Modified
Removed Items			
No.	Original Items	Replaced Items	
Q5	• Using assessments to provide my students with ungraded feedback	• Protecting students' confidentiality with regard to test scores or grades.	API

Q25	• Using portfolios to display student progress	• Avoiding teaching to test when preparing students for assessments.	API
Q26	• Assessing special needs students toward the standards specified in their I.E.P.	• Calculating and interpreting student assessment scores.	API
Q1	• Communicating grading procedures to parents/ guardians	• Marking according to university assessment policy and standards.	IISSA
Q31	• Writing multiple-choice questions	• Emphasizing quality control when assessing students.	IISSA
Q42	• Recording assessment results on the rating scale/ checklists while observing a student's performance	• Controlling external stakeholders influence when assessing students.	IISSA
Q44	• Assessing groups/ individual hands-on activities.	• Controlling institutional (university) influence when assessing students.	IISSA

API = Assessment Practices Inventory, IISSA = Inventory of Institutional Support for Student Assessment.

5.3 Testing the Hypothesized Model

Testing the hypothesised model was done using the Structural Equation Modelling (SEM). Structural equation modelling helps to account for measurement errors, and also to confirm or disconfirm a model (Schumacker & Lomax, 2004). This study extended the original assessment practices model of student assessment to include administration as another variable in explaining university academic staff assessment practices. Confirmatory Factor Analysis (CFA) was used to validate the hypothesised model of assessment practices. This was done by specifying the model and testing for the measurement model adequacy. In testing the measurement model, the constructs of the instrument were validated in order to verify the factor structure of the hypothesised measurement model and also, its adequacy (Byrne, 2010).

The first order factor measurement model was tested for the construct validity of the model. Using the CFA measurement model, the factorial validity of assessment practices among university academic staff was confirmed. This also availed the scores to measure the instrument's appropriateness in measuring assessment practices among university academic staff. Then, a second order factor CFA model was used to test for the significance and practical importance of the individual constructs (design, administration, interpretation and application) onto the academic staff's assessment practices. AMOS version 18 was used to compute and test the hypothesised model.

5.4 Testing the Adequacy of the Measurement Model

A measurement model was first constructed and tested to determine whether the coefficients of the variables were significant. In interpreting results of a model, a good model would have Chi-Square (χ^2) nearer to zero (0), p-value less than .05, CFI greater than .90, RMSEA below .080 (Byrne, 2006), and SRMR below .060 (Brown, 2011). In a total analysis of an adequate model, the Chi-Square (χ^2) values, p-values, Root Mean Square Error of Approximation (RMSEA), and Standard Root Mean Squared Residual (SRMR) were all looked at to determine whether the model fit the data. The results of all the fit indices (χ^2, p-value, CFI, RMSEA, and SRMR) would be within the required range in order to pronounce that the model had achieved fit to test the assessment practices among academic staff in Ugandan universities.

6. Results

The demographic characteristics of the participants included gender, class size, academic qualification, and formal training in assessment which were generated from the data are as follows. In terms of gender, 208(64.8%) of the participants were males while 113(35.2%) were females (Table 2). As regards to formal training in assessment, only 99(30.8%) of the academic staff had undertaken a formal course in assessment while 222(69.2%) of the academic staff had never taken any course in assessment. For the highest academic qualification attained by the academic staff, 29(9.0%) had bachelor degrees, 16(5.0%) had postgraduate diplomas, 194(60.4%) had master's degree while 82(25.6%) had doctorates as their highest qualifications. On the class size assessed by academic staff, 134(41.7%) assessed small classes while 187(58.3%) assessed large classes.

Table 2: Frequency of the Academic Staff by other Demographic Characteristics

Variables	Frequency	Percentage
Gender		
Male	208	64.8
Female	113	35.2
Assessment course taken		
Yes	99	30.8
No	222	69.2
Highest academic qualification		
Bachelor's Degree	29	9.0
Post Graduate Diploma	16	5.0
Master's Degree	194	60.4
Doctorate (PhD)	82	25.6
Class size assessed		
Small	134	41.7
Large	187	58.3

6.1 Factor Analysis

After factor analysing most of the items loaded onto their factors (four factors) as hypothesised which make up assessment practices among the academic staff (Table 3). Again, the results generated for the reliability test and factor analysis provided acceptable confidence in the hypothesised factors which make up assessment practices in order to be analysed using structural equation modelling. The four hypothesised factors included; design, administration, interpretation, and application of assessment results. According to Table 3, 14 out of the 15 items loaded onto factor one (design) as they had been hypothesised with 1 item (Q3) shifting to factor 2 (interpretation). Factor two (interpretation) reserved most of its items confirming that they are contributors to academic staff's skills of assessment interpretation. Factor 2 (interpretation) lost one of its hypothesised items (Q50) to factor 3 (application).

Factor 3 (application) had been hypothesised to have 9 items that make up the academic staff's assessment practices sub-scale of application. All the hypothesised items loaded onto factor (factor 3) with two more items; one item from factor 2 (Q50) and another item from factor 4 (Q40). Factor 4 (administration) had most of its items maintained as hypothesised though it lost one item (Q40) to factor 3. Although items Q28 was dropped during the validation of the scale using Rasch measurement analysis in research question one because of its poor fit statistics, it was included in the structural equation modelling for cross validation.

Table 3: Principal Component Matrix with Varimax Rotation

	Component			
	Design	**Interpretation**	**Application**	**Administration**
Q20	**.762**	.132	.030	.097
Q34	**.755**	.023	.071	-.002
Q32	**.753**	.098	.082	.045
Q6	**.750**	.073	.196	.081
Q7	**.740**	.139	.153	.142
Q38	**.727**	.102	.089	.088
Q35	**.725**	.077	.165	.097
Q37	**.719**	.080	.080	.093
Q36	**.708**	.130	.131	.064
Q33	**.674**	-.027	.041	-.025
Q24	**.631**	.135	.095	.196
Q29	**.614**	-.014	.084	.042
Q17	**.576**	.289	.001	.124
Q23	**.459**	.202	.052	.062
Q43	.014	**.766**	.151	.140
Q21	.127	**.744**	.059	.088
Q26	.045	**.743**	.081	.120
Q47	.134	**.738**	.035	.001
Q27	.140	**.712**	.053	.133
Q5	.098	**.700**	.185	.145
Q45	.082	**.689**	.070	.155
Q13	.093	**.648**	.197	.085
Q16	-.004	**.607**	.093	.091
Q46	.142	**.590**	.189	.126
Q3	.130	**.432**	.185	.151
Q2	.114	**.192**	.026	-.029
Q8	.089	.072	**.759**	.176
Q10	.165	.084	**.740**	.140
Q12	.112	.090	**.729**	.114
Q4	.069	.113	**.704**	.142

Q30	.122	.131	**.701**	.054
Q39	.189	.063	**.695**	.102
Q11	.179	.161	**.652**	.061
Q48	.178	.122	**.645**	.064
Q40	.008	.122	**.574**	.081
Q49	-.054	.066	**.523**	-.011
Q50	.135	.257	**.482**	.156
Q41	.033	.083	.143	**.762**
Q31	.058	-.049	.139	**.758**
Q19	.127	.077	.157	**.738**
Q18	.139	.064	.070	**.722**
Q44	.018	.117	.012	**.662**
Q42	.094	.067	.193	**.654**
Q25	.135	.109	.012	**.631**
Q22	.027	.098	.088	**.621**
Q14	.183	.207	.168	**.531**
Q15	.123	.075	.084	**.496**
Q28	-.069	.036	-.071	**.483**
Q9	.076	.070	.187	**.446**
Q1	.019	.101	-.002	**.274**

Note: Q = Item, Extraction Method: Principal Component Analysis, Rotation Method: Varimax with Kaiser Normalization.

6.2 Structural Equation Modelling

After factor analysis was conducted to validate how items loaded onto the respective constructs Confirmatory Factor Analysis was undertaken to specify the model, and also to test for the model adequacy (Byrne, 2010; Anderson & Gerbing, 1988). Testing of the hypothesised model was undertaken to test for the fitness and adequacy of the model. Also, it was done to ensure that the factors adequately measured what they were supposed to measure (see Barry & Stewart, 1997). The use of structural equation modelling by the researcher in this study was to validate a single latent construct that would explain university academic staff assessment practices with multiple indicators (see Byrne, 2000, 2006, 2010; Sabatelli & Bartle, 1995).

In this study, a three-stage structural equation modelling was undertaken using AMOS 18 to test the hypotheses. In the preliminary stage, the study specified the model which explained assessment practice among academic staff. Later, the study estimated the measurement model to test for the model fit, and also assessed the model adequacy. In estimating the hypothesised model using covariance matrices, the estimations fulfilled the fundamental statistical distribution theory by giving appropriate approximations for the properties. This was due to the study having adopted a maximum likelihood in estimating the model. After the model had been estimated, the researcher applied different methods to estimate goodness-of-fit of the model. The researcher employed the traditionally used measures to determine what comprised of adequate model, that is; (a) reasonableness of the estimates, (b) consistence of the model that collected data, and (c) the proportions of inconsistency of the dependent variables that were explained by the exogenous variables were also observed.

The study employed a critical ratio (CR) of 1.96 ($p > .05$) and above to be a significant relation between variables. Different statistical test indices were employed to determine the model adequacy; the Comparative Fit Index (CFI), Standardised Root Mean Residual (SRMR), and the Root Mean Square Error of Approximation (RMSEA) as the best fit indices for testing models (Byrne, 2006, 2010; Brown, 2011; Chen et al., 2005). The Chi-Square values were also analysed to assess the level of fit and significance in the model (Marcoulides & Schumacker, 2001). To examine the overall model fit, the Comparative Fit Index (CFI), Standardised Root Mean Residual (SRMR), and the Root Mean Square Error of Approximation (RMSEA) were used to determine the degree of fit between the covariance matrices inferred by the model (Kline, 1998). According to model testing, a satisfactory model fit should indicate a CFI equal or greater than .900 (Byrne, 2006; Kline, 2010; Bentler, 1992), SRMR equal or less than .060 (Brown, 2011), and RMSEA values equal or less than .080 (Byrne, 2010) to signify that the model is adequately parsimonious (Byrne, 2010; James et al., 1982). To choose these goodness-of-fit estimates, it was based on literature of testing the model fit (Byrne, 2000, 2006, 2010; Brown, 2011).

6.3 Testing the Model

Prior to modelling of the relationship, the measurement model was evaluated using CFA to confirm the construct structure of the measurement model. Using AMOS 18, Confirmatory Factor Analysis was employed to examine the adequacy of the model and the psychometric properties of the factors influencing academic staff's assessment practices in Ugandan universities. The results revealed in the testing of the validity of the model using the maximum likelihood estimation specified that the hypothesised model did not adequately fit the data on the initial run with the 50 items loaded onto the four factors; design (14 items), administration (13 items), interpretation (12 items), and application (11 items). The initial measurement model using 50 indicators failed to fit the data adequately, that is, some of the model fit statistics were not substantive; χ^2 = 2239.729, χ^2/df = 1.916, df = 1169, p = .000, CFI = .847, SRMR = .059, and RMSEA = .054, suggesting a need for revision.

It was observed that some of the loadings of the indicators did not yield good fit. This called for the model to be revised with indicators of low loading (below .50) being removed to improve the fit statistics of the overall model. To improve on the model, six indicators were removed from the factor of administration (Items 19, 44, 5, 28, 9, and 1) and five indicators from the factor of design (Items 32, 34, 35, 17, and 23). Also, both three indicators were dropped from the factors of interpretation (Items 26, 3, and 2) and application (Items 40, 49, and 50). Each time a modification was made, the fit indices were taken note of in the Chi-Square values, CFI, and RMSEA. After the revision of the initial model the results of the overall fit statistics of the improved measurement model were adequate; χ^2 = 902.619, χ^2/df = 1.842, df = 490; p = .000; CFI = .906; SRMR = .050, and RMSEA = .052 (Figure 4.13). The major fit indices; Comparative Fit Index (CFI) = .906, p-value = .000, Root Mean Square Error of Approximation (RMSEA) = .052, and Standardised Root

Mean Residual (SRMR) = .050 were all adequate to confirm that the model was parsimonious (see Figure 1).

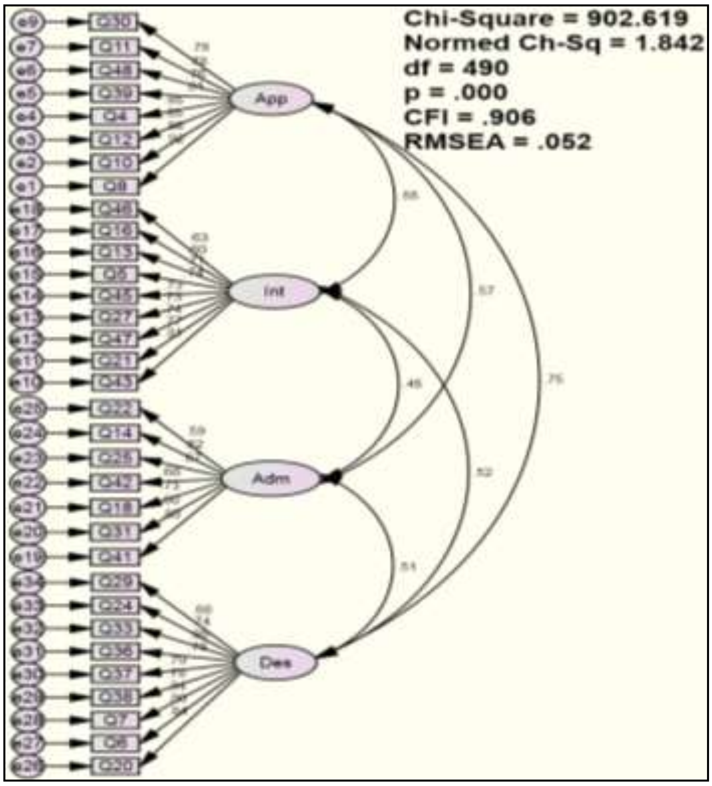

Figure 1: Revised Measurement Model
Note: Des. = Design, Adm. = Administration, Int. = Interpretation, App. = Application.

When the poorly loading indicators were removed the revised measurement model was free from offending estimates and remained with only indicators which had good factor loadings, and were all significant at $p < .001$. Also, the path coefficients were statistically significant at $p < .001$, showing both the significance and practical importance of the model. According to this there was no evidence to reject the revised measurement model. The measurement model covariance highlighted that the variables in the study yielded optimum correlations. This shows that all the variables which were included in the model contributed to explaining the latent traits in the model. In Figure 1 the path coefficients are also not highly correlated (.45 to .75) which is a confirmation of discriminant validity in the measurement model. After revising the measurement model, the data which was extracted was supported by the measurement adequacy of an AVE of .577. The AVE was greater than the squared correlation (.32) which was also evidence of discriminant validity; to support the evidence of construct validity of the model.

When the dimensions (design, administration, interpretation and application) in the construct were linked to a second order factor (assessment practices), it was found that application contributed .67, administration .56, while both interpretation and design contributed .55 (see Figure 2). The results of the path coefficients in Figure 2 were all significant and had practical importance. The measurement model with path coefficients also yielded adequate fit statistics; $\chi^2 = 769.490$, $\chi^2/df = 1.567$, $df = 491$; $p = .000$; CFI = .937; SRMR = .052, and RMSEA = .043 revealing that it was a parsimonious model (Figure 2).

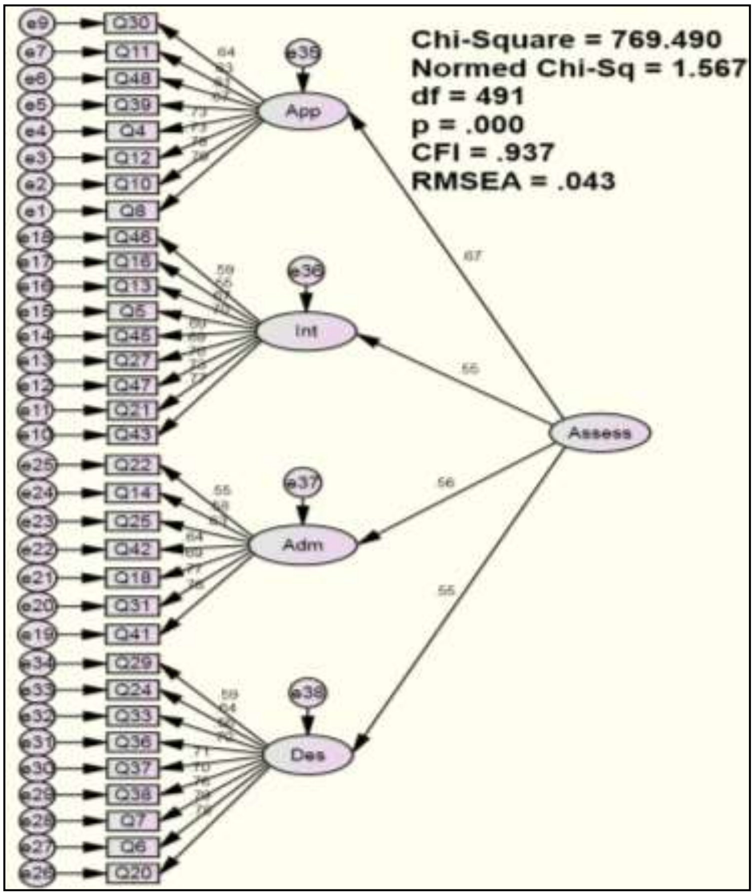

Figure 2: Revised Measurement Model with Path Coefficients
Note: Des. = Design, Adm. = Administration, Int. = Interpretation, App. = Application, Assess = Assessment Practices.

In analysis of the results from both the initial and the revised models, the revised model had better fit statistics than the initial model (see Table 4). Analysing results of the Chi-Square, p-value, robust CFI, RMSEA, and SRMR they indicated that the revised model was better in explaining the factors influencing assessment practices among the academic staff in Ugandan universities than the initial model. Table 7 highlights the results of the initial and revised measurement model fit.

Table 4: Fit statistics for the Initial and Revised Measurement Models

Model	χ^2	df	χ^2/df	CFI	RMSEA	SRMR	p
Initial	2239.729	1169	1.916	.847	.054	.059	.000
Revised	902.619	490	1.842	.906	.052	.050	.000

Note: Acceptable Levels for each Fit Criterion: Chi-Square near 0 with $p \leq .05$, CFI ≥ .900, RMSEA ≤ .080, SRMR ≤ .060 (Brown, 2011; Kline, 2006).

After identifying and testing the measurement model, it is revealed that the hypothesised model was fit explain the factors influencing assessment practices among university academic staff. According to the fit indices of the revised measurement model (Figure 2), the model was adequate in explaining assessment practices. Confirmatory factor analysis has highlighted the relationships between the different variables in the model, and their importance.

7. Discussions and Conclusions

In the findings, it was revealed that the hypothesised model for assessment practices explains the competencies in assessment practices among the academic staff in Ugandan universities. From the results of confirmatory factor analysis, the model was found to have adequate fit. The results of the measurement model suggested that there was an interaction between the variables; design, administration, interpretation, and application in the academic staff's assessment practices (Byrne, 2010). Testing the assessment practices measurement model using structural equation modelling, the measurement model did not produce good fit statistics on the first run. Some fit statistics were not in the required range, that is, CFI was below .900 which is the threshold (Byrne, 2004, 2006, 2010; Schreiber et al., 2006). But other major fit statistics like the *p*-value, RMSEA, and SRMR were not worse off. The measurement model did not produce good fit because it also had some indicators which did not have adequate loadings. Once the poorly loading indicators were removed from measurement model, the model produced reasonable data-to-model fit (Byrne, 2010).

After the initial measurement model was revised, the squared multiple correlation coefficients, t-values, factor loadings, and fit indices were all examined to determine whether the measurement model had appropriate fit (Schreiber et al., 2006; Byrne, 2010; Lei & Wu, 2007). From the results generated by the improved model, an appropriate fitting measurement model was defined by the fit indices with the required fit standards. At this stage the hypothesised model was taken to have adequate fit statistics which can define the relationships between the different latent variables of the competencies in assessment practices among academic staff in universities. The model generated regression weight and covariance matrices which were of practical importance, and were statistically significant (Byrne, 2010). In an inspection of the results of the major fit statistics of the revised measurement model like *p*-value, CFI, RMSEA, and SRMR the model was worthy explaining the assessment practices competencies among university academic staff. This is because the values of all the major fit statistics were within the required range for an adequate measurement model (Schreiber et al., 2006).

It was also observed that all the indicators which remained on the various dimensions loaded appropriately onto their respective latent variables, and even the variables had appropriate correlations (Brown, 2007). The model realising good fit statistics and factor loadings is an indication of a good measurement model (Brown, 2007; Byrne, 2006, 2010). This also means that the hypothesised model is representative in measuring assessment practices competencies among academic staff in universities in Uganda (Lei & Wu, 2007; Schreiber et al., 2006). In an overall analysis of the testing of the hypothesised model, the variables in the model were found to be related to one another meaning that they formed unity in the model (Byrne, 2010). Also, the measurement model was confirmed to be adequate in explaining competencies in assessment practices among university academic staff. The results of this study have unearthed the factors that influence assessment practices among academic staff in Ugandan universities. From the results obtained several implications of both knowledge and practical importance have been realized in this study. As mentioned earlier, this study was conceptualised from the Ainsworth and Viegut (2006) theory of assessment that had three variables; namely design, interpretation and use. On the other hand, Peterson et al. (1999a) also highlighted that administration of assessment was an important aspect in the assessment practices among academic staff. This study joined the Ainsworth and Viegut (2006) and Peterson et al. (1999a) theories to extend the assessment practices model which has shown potential to contribute to theory. This means that instead of the three variables as proposed by Ainsworth and Viegut (2006), they are now four variables in the assessment practices model, meaning that, another variable has been added to the theory of assessment practices among academic staff. It can be recommended that future researchers would use the new model in their studies and in application of assessment practices knowledge as it has been found to be of both practical and theoretical importance in this study.

The findings of this study have also supported the findings of other studies as put forward by both Ainsworth and Viegut (2006) and Peterson et al. (1999a). The research has highlighted the competencies in assessment practices. The findings of this research support findings of other studies which revealed that assessment practices have four variables which contribute to the academic staff's assessment practices. These studies found out that assessment practices include; (a) designing assessment tools, (b) gathering of learning evidence about the students, (c) analysing the evidence, reporting and discussing of the results, and (d) identifying improvement area in learning (Sadler, 2009; Gibbs, 2006; Martell & Calderon, 2005; Alexander et al., 2003; Astin et al., 1993). It has been suggested in this research that assessment design is the first stage of the assessment process, followed by administration, interpretation, and then application later. As well, these stages govern the assessment practices of the university staff at the various stages. The findings of this study are also similar to the findings of NIE (2008) in which it was revealed that favourable assessment practices among academic staff should be based on the Design-Implement-Review-Improve (DIRI) process. This means that academic staff with the adequate assessment competencies should be able to design proportionate assessments, administer them, revisit the assessment process, and be able to improve on students' learning (O'Donovan et al., 2001). This also highlights that the assessment competencies or process is not a single shot element but, it has various components which all the academic staff should possess as basic knowledge in order to undertake proper student assessment.

References

Ainsworth, L., & Viegut, D. (2006). *Common formative assessments: How to connect standards-based instruction and assessment*. California: Corwin.

Alexander, W. A., Trudy, W. B., Patricia, K, C., El-Khawas, E., Ewell, P. T., Hutchings, P., Theodore, J. M., Kay, M. M., Mentkowski, M., Miller, M. A., Thomas, M. E., & Wright, B. D. (2003). *Best Practices in Academic Assessment.* AAHE Assessment Forum, Exxon Education Foundation.

Alkharusi, H. (2011). Teachers' classroom assessment skills: Influence of gender, subject area, grade level, teaching experience and in-service assessment training. *Turkish Science Education, 8*(2), 39-47.

Alkharusi, H. (2012). Educational assessment attitudes, competences, knowledge, and practices: An exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning, 1*(2), 217-232.

Arter, J. A. (2003). Assessment for learning: Classroom assessment to improve student achievement and well-being. *RIEPR2004, 24.*

Astin, A. W., Banta, T. W., Cross, K. P., El-Khawas, E., Ewell, P. T., Hutchings, P., Marchese, T. J., McClenney, K. M., Mentkowski, M., Miller, M. A., Moran, E. T., & Wright, B. D. (1993). Principles of good practice for assessing student learning. *Leadership Abstracts*, *6*(4), 1-2.

Banta, T. W. & Associates. (1993). *Making a difference: Outcomes of a decade of assessment in higher education*. San Francisco: Jossey-Bass.

Banta, T. W., Lund, J. P., & Oblander, F. W. (1996). *Assessment in practice: Putting principles to work on college campuses*. San Francisco: Jossey-Bass.

Benson, A. D. (2003). Assessing participant learning in online environments. *New Directions for Adult and Continuing Education, 100*, 69-78.

Bentler, P. M. (1992). On the fit of models to covariance and methodology. *Psychology Bulletin, 112*, 400-404.

Biggs, J. (2003). Aligning teaching and assessment to course objectives. Paper presented at ICHEd Conference: Teaching and Learning in Higher Education: New Trends and Innovations, April 13-17, in University of Aveiro, Portugal. Retrieved January 10, 2012. http://www.event.ua.pt/iched/main/ invcom/p182.pdf

Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, *5* (1), 7-74.

Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, *80*(2), 139-148.

Black, P., & Wiliam, D. (2003). In praise of educational research: Formative assessment. *British Educational Research Journal*, *29*(5), 623-637.

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). *Evaluation to improve learning.* New York: McGraw-Hill.

Boud, D., & Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council. Retrieved

February 13, 2012. http://www.iml.uts.edu.au/assessment-futures/Assessment-2020_propositions_final.pdf

Boud, D., Cohen, R., & Sampson, J. (1999). Peer learning and assessment. *Assessment and Evaluation in Higher Education*, *24*(4), 413-426.

Braskamp, L. A. (1991). Purposes, issues and principles of assessment. *NCA Quarterly, 66*(2), 417-429.

Brown, G. T. L. (2003). *Teachers' instructional conceptions: Assessments relationship to learning, teaching, curriculum, and teacher efficacy*. Paper presented at the Joint Conference of the Australian and New Zealand Associations for Research in Education (AARE/NZARE). Auckland: University of Auckland. Retrieved March 18, 2012. http://www.aare.edu.au/ 03pap/bro03022.pdf

Brown, G. T. L. (2011). *Intermediate topics in confirmatory factor analysis and structural equation modelling.* Presentation to Research Development Office, Continuing Professional Development Programme. The Hong Kong Institute of Education. Retrieved February 15, 2012. http://www.academia.edu/ 1684096/Intermediate_Topics_in_Confirmatory_Factor_Analysis_CFA_and_Structural_Equation_Modeling_SEM_

Brown, T. A. (2007). *Confirmatory factor analysis for applied research.* New York: Guild Ford Press.

Burry-Stock, J. A., & Frazier, C. H. (2008). *Revision of the Assessment Practice Inventory Revised (APIR): A combined exploratory factor analysis and polytomous IRT approach.* Paper Presented at the American Educational Research Association. New York: NY.

Byrne, B. M. (2006). *Structural equation modelling with EQS: Basic concepts, applications, and programming* (2nd edn.). New Jersey: Lawrence Erlbaum.

Byrne, B. M. (2010). *Structural equation modelling with AMOS. Basic concepts, applications, and programming* (2nd edn.). New York: Routledge.

Connoley, R. (2004). *Criterion-referenced assessment.* Faculty of Business & Law, Deakin University. Retrieved August 21, 2011. http://deakin.edu.au/itl/assets/resources/pd/tl-modules/scholarly/studyguides/criterion-ref-assessment.pdf

Crook, C., Gross, H., & Dymott, R. (2006). Assessment relationships in higher education: The tension of process and practice. *British Educational Research Journal, 32*(1), 95-114.

Dixon, T. P. (1994). A discussion of non-traditional outcomes assessment models. Stillwater, OK: National Conference of Academic Deans. (ERIC Document Reproduction Service No. ED 372 698).

Dunca, C. R., & Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *The Alberta Journal of Educational Research, 53*(1), 1-21.

Duvall, B. (1994). Obtaining student cooperation for assessment. In T. H. Bers, & M. L. Mittler (Eds.), *Assessment and testing: Myths and realities* (New Directions for Community Colleges (pp. 47-52). San Francisco: Jossey-Bass.

Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning.* Thousand Oaks, CA: Corwin Press.

Ebersole, T. E. (2009). Postsecondary assessment: Faculty attitudes and levels of engagement. *Assessment Update*, *1*(2), 13-14.

Eisenman, C. D. (1991). Faculty participation in assessment programs. *NCA Quarterly, 66*(2), 458- 464.

Elwood, J., & Klenowski, V. (2002). Creating communities of shared practice: The challenges of assessment use in learning and teaching. *Assessment & Evaluation in Higher Education*, *27*(3), 243-256.

Ewell, P. T. (1987a). *Assessment, accountability and improvement: Managing the contradiction.* Boulder, CO: National Centre for Higher Education Management Systems.

Ewell, P. T. (1987b). Strengthening assessment for academic quality improvement. In M. W. Peterson, D. D. Dills, & L. A. Mets (Eds.), *Planning and management for a changing environment: A handbook on redesigning postsecondary institutions* (pp. 360-381). San Francisco: Jossey-Bass.

Ewell, P. T. (1988). Implementing assessment: Some organizational issues. In T. W. Banta (Eds.), *Implementing outcomes assessment: Promise and perils* (pp. 15-28). San Francisco: Jossey-Bass.

Ewell, P. T. (1991). Assessment and public accountability: Back to the future. *Change*, *23*(6), 12-17.

Ewell, P. T. (1994). Developing state-wide performance indicators for higher education: Policy themes and variations. In S. S. Ruppert (Eds.), *Charting higher education accountability: A sourcebook on state-level performance indicators* (pp. 147-166). Denver, CO: Education Commission of the States.

Ewell, P. T., & Boyer, C. M. (1988). Acting out state-mandated assessment: Evidence from five states. *Change, 20*(4), 40-47.

Fatmawati, A. (2011). *Using portfolio to assess student learning of problem solving.* Geelong-Victoria: Deakin University.

Frazier, C. H. (2007). Investigating teachers' self-perceptions applications of classroom assessment practices using Assessment Practice Inventory Revised (API_R). Doctoral Dissertation, University of Alabama.

Gardner, H. (1999). Assessment in context. In P. Murphy (Edn.), *Learners, Learning and Assessment*. London: Paul Chapman.

Gibbs, G. (2006). Why assessment is changing. In C. Bryan & K. Clegg, (Eds.), *Innovative Assessment in Higher Education* (pp. 11-22). London: Routledge.

Gipps, C. (1998). Socio-cultural aspects of assessment. *Review of Research in Education, 24,* 355-392.

Gore, J., Ladwig, J., Elsworth, W., & Ellis, H. (2009). *Quality assessment framework: A guide for assessment practice in higher education.* The Australian Learning and Teaching Council Ltd. Retrieved December 10, 2012. http://www.academia.edu/655848/QUALITY_ASSESSMENT_FRAMEWORK_A_Guide_for_Assessment_Practice_in_Higher_Education

Gronlund, N. E. (2006). *Assessment of student achievement* (8th edn.). Boston: Pearson.

Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning.* California: Corwin Press.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education*, *4*(3), 365-379.

Higher Education Academy (HEA). (2007). *Good practices in assessment of NQF BTEC higher national in further education colleges. Higher education academy.* United Kingdom. Retrieved December 12, 2011. http://www.heacademy.ac.uk/resources/detail/subjects/engineering/good-practice-in-assessment-of-nqf-btec-higher-nationals

Hindi, N., & Miller, D. (2000). A survey of assessment practices in accounting departments of colleges and universities. *Journal of Education for Business, 75*(5), 286.

Hutchings, P., & Marchese, T. (1990). Watching assessment: Questions, stories, prospects. *Change, 22*(5), 12-38.

James, R., McInnis, C., & Devlin, M. (2002). Assessing learning in Australian universities: Ideas, strategies and resources for quality in student assessment. Centre for the Study of Higher Education. www.cshe.unimelb.edu.au/assessinglearning

Kells, H. R. (1992). An analysis of the nature and recent development of performance indicators in higher education. *Higher Education Management*, *4*(2), 131-138.

Kirkwood, A., & Price, L. (2008). Assessment and student learning: A fundamental relationship and the role of information and communication technologies. *Open Learning: The Journal of Open and Distance Learning, 23*(1), 5-16.

Knight P., & Yorke, M. (2003). Employability and good learning in higher education. *Teaching in Higher Education, 8*(1), 3-16.

Lei, P. W., & Wu, Q. (2007). An NCME Instructional Module on Introduction to structural equation modelling: Issues and practical considerations. Pennsylvania: The Pennsylvania State University.

Luth, R. W. (2010). Assessment and grading at the University of Alberta: Policies, practices, and possibilities. A report to the Provost and the University. Alberta: University of Alberta.

Madaus, G., Raczek, A., & Clark, M. (1997). The historical and policy foundations of the assessment movement. In A. Goodwin (Edn.), *Assessment for equity and inclusion: Embracing all our children,* (pp. 1-34). New York: Routledge.

Martell, K., & Calderon, T. (2005). Assessment of student learning in business schools: What it is, where we are, and where we need to go next. In K. Martell & T. Calderon (Eds.), *Assessment of student learning in business schools: Best practices each step of the way* (pp. 1-22). Tallahassee, Florida: Association for Institutional Research.

McKeachie, W. J., & Svinicki, M. (2006). *McKeachie's teaching tips: Strategies, research, and theory for college and university teachers.* Boston: Houghton Mifflin.

McMillan, J. H. (2003). Understanding and improving teachers' classroom assessment decision making: Implications for theory and practice. *Educational Measurement: Issues and Practice, 22*(4), 34-43.

Mentkowski, M. (1991). Creating a context where institutional assessment yields educational improvement. *Journal of General Education*, *40,* 255-283.

Ministry of Education [MOE], (2010). *Assessment and evaluation in Saskatchewan: Assessment and evaluation.* Unpublished. Retrieved January 13, 2012. http://www.prairiesouth.ca/staff/curriculum-a-learning/assessment.html

National Council for Curriculum and Assessment [NCCA]. (2005). Supporting assessment in schools - 1: Assessment in primary schools. USA: National Council for Curriculum and Assessment.

Netherlands Inspectorate of Education (NIE). (2008). *Assessment matters: The quality assurance of student assessment in higher education*. Report of an international working group. Retrieved November 12, 2011. www.enqa.eu/.../QA%20of%20Student%20Assessment%20Report.pdf

O'Connor, K. (2009). Reforming grading practices in secondary schools. *Principal's Research Review, 4*(1), 1-7.

O'Donovan, B., Price, M., & Rust, C. (2001). The student experience of criterion referenced assessment through the use of a common criteria assessment grid. *Innovations in Learning and Teaching International, 38*(1), 74-85.

O'Grady, G. (2006). *Guide for good practice in assessment.* Ngee Ann Polytechnic, Singapore. Retrieved February 12, 2012. http://www.learnerstogether. net/PDF/Guide-for-Good-Practice-in-Assessment.pdf

Ong, S. L. (2010). Profiling classroom teacher's assessment practice. The 36th International Association of Education Assessment Conference. Assessment for the Future Generation, Bangkok, Thailand. Retrieved December 16, 2011. www.iaea.info/documents/paper_4d32f2cd.pdf

Orlich, R. J., Harder, R. C., Callahan, R. C., & Gibson, H. W. (2001). *Teaching strategies: A guide to better instruction* (6th edn.). New York: Houghton Mifflin.

Orzolek, D. C. (2006). The paradox of assessment: Assessment as paradox. *Research and Issues in Music Education, 4*(1), 1-5. Retrieved September 12, 2011. http://www.stthomas.edu/rimeonline/vol4/orzolek.htm

Paris, S., & Hoffman, J. (2004). Reading assessments in kindergarten through third grade: Findings from the centre for the improvement of early reading achievement. *The Elementary School Journal, 105*(2), 199-217.

Peterson, M. W. (1988). The organizational environment for student learning. In J. S. Stark & L. A. Mets (Eds.), *Improving teaching and learning through research.* San Francisco: Jossey-Bass.

Peterson, M. W. (1997). *Inventory of institutional support for student assessment.* National Centre for Research to Improve Postsecondary Teaching and Learning. Michigan, Ann Arbor: University of Michigan, NCPI.

Peterson, M. W., & Spencer, M. G. (1990). Understanding academic culture and climate. In W. G. Tierney (Eds.), *Assessing academic climates and culture*. San Francisco: Jossey-Bass.

Peterson, M. W., Cameron, K. S., Knapp, A., Spencer, M. G., & White, T. H. (1991). *Assessing the organizational and administrative context for teaching and learning: An institutional self-study manual.* Michigan, Ann Arbor: University of Michigan, NCPI.

Peterson, M. W., Cameron, K. S., Mets, L. A., Jones, P., & Ettington, D. (1986). *The organizational context for teaching and learning: A review of the research literature.* Michigan, Ann Arbor: University of Michigan, NCPI.

Peterson, M. W., Einarson, M. K., Augustine, C. H., & Vaughan, D. S. (1999a*). Institutional support for student assessment: Methodology and results of a national survey.* National Centre for Postsecondary Improvement. Stanford, CA: Stanford University, NCPI.

Peterson, M. W., Einarson, M. K., Augustine, C. H., & Vaughan, D. S. (1999b). *Designing student assessment to strengthen institutional performance in comprehensive institutions.* National Canter for Postsecondary Improvement. Stanford, CA: Stanford University, NCPI.

Peterson, M. W., Einarson, M. K., Trice, A. G., & Nichols, A. R. (1997). *Improving organizational and administrative support for student assessment: A review of the research literature.* National Centre for Postsecondary Improvement. Stanford, CA: Stanford University, NCPI.

Popper, E. (2005). Learning goals: The foundation of curriculum development and assessment. In K. Martell & T. Calderon (Eds.), *Assessment of student learning in business schools: Best practices each step of the way, 1*(2), 1-23. Tallahassee, Florida: Association for Institutional Research.

Quality Assurance Agency for Higher Education [QAAHE] (2007). Integrative assessment: Managing assessment practices and procedures, guide 4. Adams way, Mansfield, NG18 4FN Science and Business Media B.V.

Quality Matters (2006). *Quality matters rubric standards, quality matters, inter-institutional quality assurance in online learning.* Retrieved November 8, 2011. http://www.qmprogram.org/Rubric.htm

Roeber, E. D. (2002). How we will gather the data we need to inform policy makers? In R. W. Lissitz & W. D. Schafer (Eds.), *Assessment in educational reform*. Both means and ends. Boston, MA: Allyn and Bacon.

Ryan, G. J. (1993). After accreditation: How to institutionalize outcomes-based assessment. In C. Praeger (Eds.), *Accreditation of the two-year college.* San Francisco: Jossey-Bass.

Sadler, D. R. (2009). Fidelity as a precondition for integrity in grading academic achievement. *Assessment & Evaluation in Higher Education*, *34*(1), 1-17.

Satterly, D. (1989). *Assessment in schools* (2nd Edn.). Oxford: Blackwell.

Schreiber, J. B., Stage, F. K., King, J., Nora, A., & Barlow, E. A. (2006). Reporting Structural Equation Modelling and Confirmatory Factor Analysis: A review. *Journal of Educational Research, 99*(6), 323-337.

Shohamy, E. (2004). Assessment in multicultural societies: Applying democratic principles and practices to language testing. In B. Norton & K. Toohey (Eds.), *Critical pedagogies and language learning* (pp. 72-93). Cambridge: Cambridge University Press.

The Quality Assurance Agency for Higher Education [QAHE] (2006). *Code of practice of academic quality and standards in higher education. Section 6: Assessment of students.* Retrieved January 23, 2012. http://www.qaa.ac.uk/assuringstandardsandquality/code-of-practice/Pages/default.aspx

Toohey, S. (1999). *Designing courses for higher education.* Buckingham: Open University Press.

Warren-Piper, D., Nulty, D. D., & O'Grady, G. (1996). Examination practices and procedures in Australian universities. Higher Education Division, Department of Employment Education and Training. Evaluations and Investigations Program. AGPS, Canberra.

Weddel, K. S., & Van Duzer, C. (1997). Needs assessment for adult ESL learners. ERIC Digest. Washington, DC: National Clearinghouse for ESL Literacy Education.

Young, C. C., & Knight, M. E. (1993). Providing leadership for organizational change. In T. W. Banta, J. P. Lund, K. E. Black, & F. W. Oblander (Eds.), *Making a difference: Outcomes of a decade of assessment in higher education*. San Francisco, CA: Jossey-Bass.

Zhang, Z. (1995). *Investigating teachers' perceived assessment practices and assessment competencies on the API.* Tuscaloosa, AL: The University of Alabama.

Zhang, Z. (1996). *Teacher assessment competency: A Rasch model analysis.* A paper presented at the annual meeting of American Educational Association. USA, New York.

Zhang, Z., & Burry-Stock, J. A. (1994). *Assessment practices inventory.* Tuscaloosa, AL: The University of Alabama.

Zhang, Z., & Burry-Stock, J. A. (2003). Classroom assessment practices and teachers' self-perceived assessment skills. *Applied Measurement in Education, 16*(4), 323-342.