RESEARCH PAPER

ESET: [EAGLES STUDENT EVALUATION OF TEACHING] - AN ONLINE ANDRAGOGICAL STUDENT RATINGS OF INSTRUCTION TOOL THAT IS AN IN–DEPTH SYSTEMIC STATISTICAL MECHANISM DESIGNED TO INFORM, ENHANCE, AND EMPOWER HIGHER EDUCATION

By

JAMES EDWARD OSLER II * MAHMUD A. MANSARAY **

** Faculty Member, North Carolina Central University, USA.*
*** Research Analyst, North Carolina Central University, USA.*

ABSTRACT

This paper seeks to provide an epistemological rationale for the establishment of ESET (an acronym for: "Eagles Student Evaluation of Teaching") as a novel universal SRI [Student Ratings of Instruction] tool. Colleges and Universities in the United States use Student Ratings of Instruction [SRI] for course evaluation purposes (Osler and Mansaray, 2013a). This research investigation is the third part of a post hoc study that psychometrically examines the reliability and validity of the items used in an Historic Black College and University (HBCU) SRI instrument. The ESET sample under analysis consisted of the responses to 56,451 total items extracted from 7,919 distributed Student Ratings Instruments that were delivered electronically (at HBCU) to students who completed the ESET tool. The ESET methodology provides a statistically valid SRI/SET survey instrument along with a variety of post hoc statistical measures to determine the efficacy of collegiate instruction. This research is also the continuation of research conducted on innovative statistical metrics introduced in the i-managre's Journal on Mathematics.

Keywords: Advanced Statistics, Construct Validity, Cronbach's Alpha, Efficacy, Formative Assessment, Goodman & Kruskal's Lambda, Principal Component Factor Analysis, Reliability, Statistics, Student Ratings of Instruction [SRI],

INTRODUCTION

With the increased use of performance-based allocation (O'Shaughnessy, 2013) by many state legislatures as a benchmark of student learning in funding many United States Public Universities, the burden to promote quality learning and effective teaching in public universities increasingly falls within the ambit of the university administrators (Chancellors, Provosts, Deans, and Chairs). Simultaneously, the demands to satisfy the students' longing for superior grades (Agbetsiafa, 2010), is influencing college administrators to progressively make use of diverse methods of evaluating the quality of teaching, to formulate decisions on college programs, retention and graduation rates, student learning outcomes, and faculty employment, among others. In truth, strategic plans, like faculty-student advising; for example, that would enhance student success are applicable in some universities, and others are constantly exploring ways that would further improve effective teaching and student success. Indeed, an assessment tool recurrently used by college administrators in state universities to assess teaching effectiveness and student learning for personnel policy decision-making, including other summative purposes (tenure, merit increase, retention for non-tenured), is the "Student Evaluations of Teaching" (SET) (Agbetsiafa, 2010; Safavi, Bakar, Tarmizi, & Alwi, 2012; Seyedeh & Kamariah, 2013; Taylor, Grey, & Satterthwaite, 2013). Agbetsiafa (2010) noted that, the SETs are ever more becoming noteworthy in summative and formative procedures in the Universities because they present a strategic, methodical, and valued means of obtaining feedback on students' responses to instructors and courses. It is noteworthy that, administrators in state Universities are progressively making inferences on SET to compose personnel decisions regarding curriculum

improvement, teaching effectiveness and student learning outcomes (Agbetsiafa, 2010; Donnon, Delver, & Beran, 2010; Kozub, 2010; Madden, Dillon, & Leak, 2010). Arguably, even though several academics may conclude that SETs are dependable tools, there is a gap between the researchers concerning the trustworthiness of the evaluation design for faculty assessment. There is obviously less unity among the academics about their general validity (soundness) and reliability (trustworthiness) with respect to the level at which the evaluation designs fittingly assess tangible terms (e.g., teaching quality), or present a complete rating of the course or educator (Agbetsiafa, 2010; Beran & Rokosh, 2009; Clayson, 2009; Osler & Mansaray, 2013b). In effect, while some researchers maintained, there was hardly any proof of an association between student ratings and teaching effectiveness (Madden et al., 2010; Otani, Kim, & Jeong-IL, 2012), others regarded the ratings to be a significant evaluation of teaching efficiency and student learning (Frick, Chadna, Watson, Wing, & Green, 2009; Zhao & Gallant, 2012). However, due to recurrent utilization of student ratings in university administrative policy resolutions, and the inconsistency of the academics to reach an agreement on the validity and reliability of the current design to assess faculty and teaching quality, it is crucial to endorse the extension of the model to include other measurements of evaluating teaching quality and student learning. Thus, the objective of this paper is to promote the expansion of the student ratings of instruction to include the ESET, to effectively and efficiently evaluate teaching quality (and its antecedent: teaching effectiveness) and thereby provide a new mechanism that will be contributing to the metrics use of teaching evaluation by the collegiate administrative sector.

1. Key Terminology

Validity: The validity of an assessment design is the degree to which the device measures what it is proposed to mention (Agbetsiafa, 2010; Chen & Watkins, 2010; Leedy & Ormrod, 2009; Zhao & Gallant, 2012)

Reliability: The reliability of an assessment design is the consistency with which the evaluation instrument generates outcome when the entity that is being evaluated stays unchanged (Agbetsiafa, 2010; Chen & Watkins, 2010; Leedy & Ormrod, 2009; Zhao & Gallant, 2012).

Formative Assessment: Student ratings are used by many Universities to assess faculty at the end of every semester for teaching effects which may be used in curriculum improvement, teaching effectiveness and student learning outcomes (Agbetsiafa, 2010; Donnon, Delver, & Beran, 2010; Kozub, 2010; Madden et al., 2010).

Summative Assessment: Student ratings are used by many Universities to assess faculty at the end of every semester for teaching effects which may be used for faculty promotion, tenure and faculty hiring (summative) (Agbetsiafa, 2010; Safavi et al., 2012; Seyedeh & Kamariah, 2013; Taylor et al., 2013).

2. Operational Definition of ESET Variables

The operational variables in the ESET research include the following independent and dependent (construct) variables:

Independent (Internal Construct) Variables: The independent variables for this study were the questions on the survey instruments envisioned to assess faculty on course programs and teaching effectiveness contained within the ESET instrumentation infrastructure. The questions included in ESET were on: course design, instructional method of the teacher, examinations and feedbacks, instructor's enthusiasm and communication, among others. Indeed, Osler and Mansaray (2013b) applied reading scripts and instructor's passion, among others, as independent variables in their analysis of teaching effectiveness and student learning.

Dependent (External Construct) Variables: The dependent variables measured the effect of the independent variables. The dependent variable is teaching quality. The global overall teaching effectiveness of the instruments is assumed to indicate teaching effectiveness and gives the perception of student learning. Agbetsiafa (2010) applied effective teaching as the dependent variable in his construct validity of teaching effectiveness and student learning.

3. Literature Review

The term "andragogy" describes a theoretical framework in adult learning conceptualized by Knowles (1974) to isolate adult learning from conventional pedagogy. Indeed, the pedagogical approach is grounded on a few assumptions, including the proposition that since students hardly have adequate rudimentary knowledge; they rely on the teacher for bearing concerning their education requirements (Forrest & Peterson, 2006). Knowles (1980) described the andragogy as the art and science of assisting adults in learning, which he compared to pedagogy as the art and science of assisting children in learning. Knowles countered his theory against the backdrop that children and adult learn inversely and on his reflection, they appear to retort in a different way to their teachers.

The five assumptions underlying Knowles' (1980) andragogy define the adult student as:

1. An individual that has an independent, self-image, and that guides his individual education.
2. An individual that has accrued a basin of lifetime experiences, which essentially offer a rich foundation for learning.
3. An individual that has learning requirements thoroughly associated to varying social parts.
4. A problem-centered individual who is engrossed in the instantaneous presentation of knowledge.
5. An individual enthused to study by interior rather than exterior factors.

Based on these assumptions, Knowles (1980) postulated a program-development model for creating, applying, and assessing educational experiences with adults. The implication of the first assumption, for example, is that as adults developed, they become further self-regulating and self-directing. Concerning this assumption, Knowles (1980) noted the classroom environment ought to be adult as such both substantial and expressively. He further said in the mature classroom, adults feel recognized, appreciated, and sustained, and there is an existing atmosphere of affinity between instructors and students as combined questers. He surmised that these assumptions form the foundation of adult learning. Forrest and Peterson (2006) who noted that, contrary to children, adults learn from their massive collection of lifetime experiences, which reinforces the capability to self-determination on their deficiency and, therefore, requires absorbing, also support this notion. In addition, the adult students are likewise possible to want a superior feel of collaboration between the student and the instructor as they advance through the Educational procedure (Zmeyov, 1998).

Knowles' (1980) andragogy developed into one of the greatest contentious and argued concepts in the arena of adult education (Brookfield, 1986; Davenport & Davenport, 1985; Hartree, 1984). Arguably, many scholars find it difficult to accept andragogy as a theory. Indeed, in one of these debates, Davenport & Davenport (1985) noted that, andragogy was categorized, on different occasions, as a procedure of adult education; a theory of adult learning; a method of adult education; and, a theory of adult education, among others. Hartree (1984) even questioned that, if there was a theory of any kind, signifying that perhaps the assumptions were merely philosophies of virtuous exercises, or explanations of what the adult learner ought to be. Additionally, a region of ongoing disagreement is the level at which the assumptions are features of adult learners alone. Certainly, though it may be factual that, several adults are autonomous learners, certain adults, arguably, are exceedingly reliant on an instructor for configuration. Alternatively, certain children are autonomous, self-directed learners. Additionally, even when reflecting on, the more apparent assumption that adults have added and subterranean lifespan experiences, this possibly will not serve promising in a learning condition. Definitely, certain life experiences can serve as obstacles to knowledge (Merriam, Mott, & Lee, 1996). In addition, children in assuring circumstances may have an array of experiences qualitatively better-off than some grown-up.

The fluidity of the contentious arguments among the scholars that the andragogy assumptions were not essentially accurate of all adults spurred Knowles (1980) to revise his individual intelligence regarding whether andragogy was only for adults and pedagogy only for

children. This induced him to move away from an andragogy as opposed to pedagogy locus to demonstrating the two on a continuity extending from teacher-directed to student-directed learning. Knowles recognized that both methodologies are opposite with children and adults, contingent on the circumstances. Certainly, an adult who discerns little or naught about a subject, for example, will be heavily reliant on the instructor for knowledge path (Knowles, 1984). At the same time, Knowles also recognized that children who are logically enquiring and are actually self-directing in their scholarship external of school, among others, possibly will also be more self-directed in school. Knowles reviewed assumptions have resulted in a new definition of andragogy, which relied more on the learning condition than on the learner. Today, the application of andragogy in teaching and learning may depend on whether the academic focuses on the teacher-directed or student-directed aspect of learning.

Indeed, in spite of the continued contention about the theoretical connotation of andragogy, it is widely prevalent among the academics and researchers worldwide, and its research group is increasingly mounting (Amrein-Beardsley & Haladyna, 2012; Gilstrap, 2013; Savicevic, 1991; Young, 2012 and Gilstrap, 2013, for example, applying a quantitative technique, produced a synopsis of the andragogy theory in relation to teaching philosophies among librarians ACRL members, using Hadley's Educational Orientation Questionnaire. Based on the theoretical framework, the result found nonlinear and negative correlation between librarians with an understanding of the ACRL Standards and their adult learning orientation ratings, $p = .047$, $t < .05$. The results also highlighted the significance of adult learning as well as the assessment of teaching philosophies. Similarly, Forrester and Peterson (2006) also applied andragogy in management. The authors noted the anagogical approach was essential in Management Education, to help and prepare students for their working environment. Furthermore, the authors surmised that current management desires the application of skills learned, and not rule of ethics. Therefore, without application, the student cannot acclimatize to the evolving place of work. Amrein-Beardsle & Haladyna (2012), meanwhile, referenced the andragogy theory of adult learning to generate and validate a survey to assess teaching effectiveness. The authors also noted that, an assessment survey based on a theory that defines an effective teacher upsurges the likelihood for validation, and that bringing into line the survey items with a theoretical based explanation of effective teaching that decreases the quantity of total items required. Consequently, there is less chance that halo rating errors will decrease subscore validity.

4. Purpose of the Study

For the purpose of this research, the applicability of andragogy is noteworthy in determining the validity and reliability of both Student Evaluations of Teaching (SET) and Graduating Senior Survey (GSS) combined as a measurement of teaching effectiveness and student learning. First, the theory takes into account that the participating students in both survey designs are adult learners, and that learning takes place within the classroom. Depending on the situation, the classroom scene may operate as either teacher-directed or student-directed learning. In addition, the first assumption in the andragogy model has to do with adults need to know why they are learning before they could participate in the learning process (Knowles, 1984). This assumption correlates with a few individual items on both SET and GSS survey designs where students are asked about their objects in learning as they relate to their learning outcomes. In addition, the SET design takes into account the professional relevance of the course to the degree program (Amrein-Beardsley & Haladyna, 2012) in relation to the instructional effectiveness. This aspect is also connected with the first assumption of the andragogy model. This notwithstanding, both the SET and GSS designs also require students to self-rate themselves on a few survey items concerning their learning efforts and motivation. Such questions have to do with the fourth and fifth assumptions of the models. In sum, the andragogy model is a proper theoretical fit for this research.

5. Background of ESET Research

5.1 Existing SET Research

The existing gap between the researchers in connection with the validity and reliability of the SET instrument as a

measurement of teaching effectiveness means, SET alone cannot guarantee the truthfulness of instrument as a measure of teaching effectiveness. This has encouraged some researchers to endorse other methods of assessing teaching effectiveness (El Hassan, 2012; Marsh et al., 2011). Even so, most of the available literature on the measurement of teaching effectiveness is on SET because of its popularity in many universities, in spite of its questionable reliability by some academics. It is against this background that the literature review of this research will include a discussion of SET. Additionally, while there is hardly any information on the GSS survey as a measurement of teaching effectiveness, SETs are widely applicable in many Universities worldwide as a proxy for teaching effectiveness and by implication, the student perception of learning (Carrell & West, 2010; Hatfield & Coyle, 2013; Madden et al., 2010; Spooren, Brockx, & Mortelmans, 2013). In reality, there is a growing list of literature on the utility, validity and reliability of the student evaluations of faculty. In reality, several Journals on the evaluation of university faculty on teaching effectiveness and student success make use of the Student Evaluations of Teaching (SET) design (Clayson, 2009; Donnon et al., 2010; Otani et al., 2012; Pritchard & Potter, 2011; Ruppert & Green, 2012; Zhao & Gallant, 2012).

5.2 The Utility of the ESET Design

An exercise that is predominant in several universities and colleges in the United States and elsewhere, is the utilization of SET to assess instructional effectiveness and student success (Donnon et al., 2010; Spooren, Mortelmans, & Denekens, 2007; Stowell, Addison, & Smith, 2012). This notion is supplemented by the agreement among some scholars who noted SETs appear sufficiently useful and effective in evaluating what they want to define: teaching efficiency, student learning satisfaction, educational knowledge, and program curriculum (Agbetsiafa, 2010; Beran, Violato, & Kline, 2007; Skowronek, Friesen, & Masonjones, 2011; Zhao & Gallant, 2012). Obviously, the use of student evaluations of faculty in Colleges and Universities is to deliver a helpful statement to faculty for teaching improvement, as well as a superficial assessment of teaching efficiency for personnel or administrative decisions, in addition to supplying information to students for the choosing of courses and instructors (Marsh and Roche, as cited in Zhao & Gallant, 2012; Beran, Violato, Kline, & Frideres, 2009). Additionally, the application of SETs as a measurement of teaching effectiveness is significant; because, they offer a planned, systematic, and effective medium of receiving feedback on students' reactions to teachers and courses (Agbetsiafa, 2010), and have been about ever since the middle of 1920s (Cohen, as cited in Donnon et al., 2010; Apollonia & Abrami, as cited in Safavi et al., 2012). This notwithstanding, faculty also applies SETs to acquire students' responses concerning their courses and record development in their instruction parts and responsibilities, which is an important influence in their occupations. Beran and Rokosh (2009), for example, informed from a survey of 262 University teachers that 84% of the respondents supported the application of SET overall, and 62% of the respondents felt that departmental heads and deans properly applied SET results. However, the method in which instructors apply SET differs about background and experience (Sprooren et al., 2007). To this end, Arthur (2009) argued that replying to feedback was a multifaceted procedure, and as a result, he established a typology of factors (e.g. Personality, student characteristics, teaching and learning strategies) that impacted teachers' individual responses to undesirable feedbacks (i.e. blame, shame, etc.). Meanwhile, Aleamoni (as cited in Zhao & Gallant, 2012) had also suggested the application of student ratings because students can suggest information on the accomplishment of essential Educational objectives; empathy with the instructor; and, rudiments of a classroom, such as instructional provisions, assignment, and instructional processes.

Additionally, the student ratings are also applied to express understanding to the students and to institute administrative resolves, such as offering life-term tenure and advancement, evaluation of curriculum programs, faculty hiring, and improvement in teaching performances (Beran et al., 2007; Beran et al., 2009; Kozub, 2010; Seyedeh, Kamariah, Rohani, & Alwi, 2012; Seyedeh & Kamariah, 2013). Furthermore, some studies revealed that students have the tendency to consider teaching assessments sincerely, and are eager to contribute and

offer expressive answer when they contemplate and comprehend that, their contributions are replicated and integrated by their instructors and the University (Agbetsiafa, 2010). Beran et al. (2007) also noted that by helping instructors to increase their instruction through rating feedback, administrators may be able to supervise definitive course developments. Such information may be important to the administrators who can apply it to supervise and improve teaching efficacy largely. The integration of such information is also likely; to regulate teaching quality in a department or program concerning other programs (Beran et al., 2007). The authors further added procedural resolutions may also be knowledgeable by ratings, that may define the instructors' course assignments in succeeding terms. Spooren et al. (2013), argued that, the appropriate gathering an interpretation of SET data depends upon administrators having thorough methodological preparation and systematic briefing on the major outcomes and trends in the study field.

Nevertheless, in spite of its widespread utility, the students' perceptions concerning the utility of SET ratings were lacking, ambivalent, and not properly understood (Beran et al., 2009; Spooren et al., 2013). Spooren et al. (2013) further said that, the students' ambivalent of the utility of SET was the comprehension in their assessments may not be taken earnestly either by the faculty or administration for enhancing teaching quality. In addition, many faculty members also remained apprehensive about SETs' summative application in personnel decisions, such as faculty retention, promotion, salary increases, tenure (Beran & Rokosh, 2009; Beran et al., 2007). This apprehension had a connection with the absence of knowledge about the efficacy of the ratings data, or the unease that administrators were exploiting the ratings data in personnel decisions (Beran & Rokosh, 2009). This notwithstanding, the authors further added that SET ratings also had imperfect usage in refining detailed aspects of instruction, such as choosing course resources and organizing assignments and exams.

5.3 Psychometric Properties: ESET as a SET Model—Items and Measurement

There are several existing tools, both online and direct, which are applicable for assessing faculty on teaching effectiveness and the perception of student learning. Several questions on SETs aim to measure the instructor on teaching effectiveness and program outcomes (Anastasiadou, 2011; Baker, Pollio, & Hudson, 2011; Chulkov & Van Alstine, 2012; Fitzpatrick & Miller-Stevens, 2009; Keeley, Furr, & Buskist, 2010; Skowronek et al., 2011; Spooren, 2010; Zhao & Gallant, 2012). The SET designs are independent student survey instruments utilized to assess teaching excellence and student learning outcomes. Examples include the British Noel-Levitz Student Satisfactory Inventory, the Course Perception Questionnaire, the Student Evaluations of Educational Quality, the CoursEval, the Course Experience Questionnaire, the Student Instructional Report, and the Student Teaching Evaluation Instruments, among others (Liu, 2012; Young & Duncan, 2014). Indeed, even with the dissimilar labels, the majority of the instruments have comparable characteristics and individualized learning items, which were summed to generate a teaching effective, score (Skowronek et al., 2011). SETs usually contain a number of Likert-scale (between 4-points to 10-points scales) based questions that request students to assess several aspects of the instructor's teaching and course design. The questions are placed on Likert-scales. These forms or surveys are finalized by the students at the end of every semester and frequently function as a summative measure in administrative decisions about faculty tenure promotion, and merit pay (Johnson, Narayanan, & Sawaya, 2013; Mau & Opengart, 2012; Venette, Sellnow, & McIntyre, 2010), and as a formative measure for enhancing teaching abilities and course design (Donnon et al., 2010;Dorasamy & Balkaran, 2013; Osler & Mansaray, 2013).

SETs usually contain Likert-scale based questions on teaching effectiveness. The Universal Student Ratings Instrument, for example, was a SET design introduced at a Canadian Graduate University, and had 12 items on a Likert scale independent of the others and applied courses of comparable kind and scope as a basis for assessment. The authors further said 11 of the 12 items were designed to produce definitive ratings on module of the course, using a 7-point scale from 1 = strongly disagree to 7 = strongly agree. The 12th item on the design was an inclusive global

rating of the quality of the course instruction computed on a different 7-point scale, which ranged from 1 = unacceptable to 7 = excellent. Osler and Mansaray (2013) also noted that, the applicability of the Cours Eval design, to assess faculty at a University in the Southern United States. This was a rating design on a 5-point Likert scale, administered online every semester, and requires students to assess their instructors on 12 items, which together summed into teaching effectiveness. The questions on the rating scale include:

1. The identified goals and objectives for this course are in accordance with what was actually taught.
2. The subject material of this course is soundly organized.
3. The instructor plainly delivers his/her subject substance.
4. The instructor is passionate and stimulates interest in his course.
5. My supremacy to reason, censure, and/or construct have been enriched as a result this course.
6. The texts and other readings allocated for this course were supportive.
7. The instructor applies instructional methods (for example, discussions, lectures, audio, visuals, field work, demonstrations, computer programs, etc.), which effectively improve learning in this course.
8. The examinations are in accordance with the course objectives and the instruction.
9. Quizzes, examinations and/or written assignments are delivered often enough to help me assess my growth.
10. The instructor is sincerely concerned with students' advancement.
11. I am able to acquire assistance from the instructor when I require it.
12. This instructor is effective in endorsing learning.

Indeed, the 5-point Likert scale on each item ranged from 1 = strongly disagree to 5 = strongly disagree. The design further has three sub-items on student efforts on the course whose values ranged from 1 = never to 5 = all of the time. In reality, many of the research journals assume the Likert scale on SETs as an interval scale (Chulkov & Van Alstine, 2012; Zhao & Gallant, 2012), thus making it possible to apply quantitative techniques to examine the reliability and validity of the instruments.

Even with the widespread use of the SET design, it is argued that the majority of them have a single-item methodology of faculty measurement, and measuring instructional skills on a single-item methodology generates a more confusing interpretation of the given responses (Spooren et al., 2007). In addition, there is still the disagreement among the academics about the reliability and validity of SET as a measurement of teaching effectiveness (Anastasiadou, 2011; Galbraith & Merrill, 2012; Galbraith, Merrill, & Kline, 2012; Skowronek et al., 2011; Zhao & Gallant, 2012).

5.4 The Reliability of ESET and SET

As the reliance on student ratings has augmented over time, so has the number of research studies on the psychometric properties of ratings, particularly reliability and validity (Beran & Rokosh, 2009). In truth, the reliability of SET refers to the consistency of ratings among distinct raters and the steadiness of such ratings over time. In other words, reliability is concerned with the internal consistency, stability and dependability of the design applied to assess teaching effectiveness. Therefore, given the realization that, the SET design is extensively used in several Universities, it is noteworthy that, the data resulting from these instruments function as a reliable measure of teaching superiority and course improvement (Agbetsiafa, 2010). McMillan (as cited in Zhao & Gallant, 2012) also noted a dependable result is one that has equivalent performance at different times. Notwithstanding the enduring controversy surrounding the reliability of SETs, several studies realized that student evaluations of teaching are reliable, stable across items, raters, and period. Certainly, as reliability is a principal foundation of validity, having superior reliability is critical for summative and formative evaluations (Haladyna & Amrein-Beardsley, 2009).

Several different statistical models are applicable in the determination of the reliability of the rating tools utilized to assess faculty. Most of these assessments focused on the internal consistency (stability) of the ratings. The most established is the Cronbach's alpha statistics, with the alpha varying from 0 to 1; the 1 being the maximum reliability score. In their study, Dorasamy and Balkaran

(2013), for example, assessed student ratings of teaching aptitudes for use in program evaluation, making use of a sample of 3,060 within the Faculty of Management Sciences at the Durban University of Technology. Using teachers' questionnaires on a Likert scale, the authors completed a reliability test by directing several measurements on the same subject, and a general reliability score of 0.949 was achieved, indicating a high degree of internal consistency of the responses. Beran et al. (2009), for example, conducted a study to determine what students find valuable in student ratings. With the utilization of survey responses from (n=1229) students at a prominent Canadian University, the authors established a psychometrically extensive measurement of the utility of student ratings. Using the Cronbach's Alpha model, the results confirmed the internal consistency reliability of the 16 items on the rating scale at 0.93, thus signifying a superior level of internal consistence for the SET ratings. Donnon et al. (2010) correspondingly realized a superior level of internal consistency with a Cronbach's alpha coefficient of 0.93 in their research on SET in medical sciences graduate programs. Osler and Mansaray (2013) also noted a Cronbach's Alpha of 0.954 for the 15 items on the SET scale for their study on the validity and reliability of independent instructional measures in a Southern state University, confirming the high internal consistency of the measuring instrument. Agbetsiafa (2010) also had a high level of internal consistency of the student answers to the rating items in his study. Anastasiadou (2011) similarly had a Cronbach's alpha of 0.908, which was over the threshold, endorsing a strong internal consistency of the designing instrument. Still, there are disagreements among scholars concerning the validity and reliability of SET. Galbraith et al. (2012), for example, found little or no support for the soundness of SET as an overall gauge of teaching effectiveness or student learning.

5.5 The Validity of ESET and SET Research

There are several Journals on the validity of the SET design. However, the validity study has always been a contentious subject in SET, and the problem is still unresolved among the academics. Certainly, in all-purpose, validity of SETs indicates the level at which the student ratings in effect assess what they are intended to evaluate (Zhao & Gallant, 2012). There are several types of validity studies, including content validity, construct validity, external validity, and criterion-related validity, among others. However, there is an integrated approach to validity and validation was recent (Kane, as cited in Haladyna & Amrein-Beardsley, 2009). Thus, in the novel cohesive view, validity refers to the truth of an explanation of a score, for instance, the results from student evaluations of teaching (Haladyna & Amrein-Beardsley, 2009). Moreover, the authors noted that, the validity may also relate to the application of student ratings, citing their research whose objectives were to assist the instructor to improve the methods of instruction. Then again, the validity of this information for formative uses was questionable.

Indeed, in spite of the recent unified validity view, some scholars remain to be interested in a narrow focus of validity, particularly construct validity, in connection with SET as a measurement of teaching effectiveness. Surely, Agbetsiafa (2010) had argued for the use of construct validity by noting that to apply student ratings to evaluate teaching effectiveness and student learning outcomes, then the instruments must be visible to inspiring validity trials and examination. Cronbach and Meehl (as cited in Zhao, 2012) who had said construct validity was the degree to which an apparent measurement mirrored the central theoretical construct that the academic had planned to assess initially postulated Agbetsiafa's argument. Skowronek et al. (2011) also said it was crucial to discuss concerns that were associated with construct validity, including rejoining whether the nature of the student rating technique was rational for the construct that was being assessed.

In reality, there are several research studies with applicable distinct statistical models, including factorial analysis, stepwise linear regression, multivariate analysis of variance, and structural equation models, among others, to validate the SET tool. Agbetsiafa (2010), for example, utilized the factorial analysis to examine the construct validity of the SET in determining the association between teaching effective and student ratings in a university level course, in Economics at the University of Indiana. Using (N=1300)

sampled students, the result of the Kaiser-Meyer-Olkin (KMO) statistics on the rating scale was 0.912, signifying the suitability for the use of factor analysis to the data. In addition, the Bartlett's for the presence of interaction amongst the variables was significant at $p < 0.0001$. In sum, the results found positive associations between student perception of teaching effectiveness, education support, effective communication, and clarity of course works, and course assessment and feedback, therefore confirming the construct validity of the rating tool. Similarly, Zhao and Gallant (2012) applied the confirmatory factor analysis via the structural equation model to examine the validity of the SET design in a Midwestern university in the United States for both administrative and instructional decisions. Using (N = 73,500) sampled students who completed the assessment, the results revealed the model was an acceptable fit for the data, and that instruction effectiveness was appropriately and satisfactorily evaluated by the 10 observed variables in the SET survey. That notwithstanding, using the principal component factor analysis with varimax rotation, Osler and Mansaray (2013) realized a significant relationship amongst the 15 items on the rating scale of $p < .000$. Furthermore, the factor loading for the 15 items all have loading of $>= 0.82$, thus establishing the construct validity of the ESET.

El Hassan (2009) on the other hand, was particular about substantive and consequential validity of the student ratings. His research discussed concerns of substantive and consequential validity, and upheld these could be effusively discussed where the assessment techniques were completely planned and realized, including effective communique to students and faculty concerning the tenacity of the evaluation techniques. Exploiting a descriptive statistics on a 5-point Likert scale, the author noted that, about 70% of the students acknowledged the student ratings to be the standard for demonstrating suggestions for improvement, and about 50% of students also said faculty values their input to generate teaching development. The study also found several instructors value which is the contribution from the ratings and applied them for course progression. Stowell, Addison, and Smith (2012), meanwhile, examined whether there was a variance in the response rate and validity between online and classroom-based student evaluations of faculty. Using (N = 2057) sampled students who responded to the SET survey, the authors utilized the t-statistics and correlation matrices, and realized no significant differences in the mean ratings between the online format and classroom-based student ratings, thus establishing that the two assessment formats generated comparable data and validity results.

5.6 Alternative Arguments Regarding ESET and SET

Even with the extensive application of SET as a measurement of effective teaching in many colleges and universities, it continues to be fraught with controversies and biases, which appeared to challenge its validity on several fronts. Arguably, student ratings of faculty differ in accordance with several student characteristics. Indeed, students who anticipated a superior letter grade in the course appeared to offer high ratings of instruction, and an instructor who embraced a more compassionate grading standard when applying subjective testing techniques may receive higher student assessments of teaching performance. A supplement to this bias was the findings of Slocombe, Miller, and Hite (2011) who noted students inclined to offer higher evaluations to instructors who applied humor and to instructors they adored. The authors also noted students failed to offer higher evaluations to male instructors or those below 55 years. In addition, Galbraith et al. (2012), in their study of effective teaching, found minimal or no backing for the validity of SET as an overall pointer of teaching effectiveness or student learning. Other studies also noted the influence of gender and race on student ratings of faculty, while Ibrahim (2011) said that class size had a positive influence on the dependability of student evaluations of instruction, and that ratings received from bigger classes were more dependable than ratings of reduced classes.

In sum, the ongoing enquiry surrounding the overall validity of SET alone as a measurement of teaching effectiveness evidently underscores the necessity to expand SET to include other measures of teaching efficiency, particularly for personnel decisions (El Hassan, 2012). To this end, the significance of the addition of the Graduating Senior Survey design (GSS) to the SET design as a combined measurement of teaching effectiveness and student

learning for personnel decisions may be crucial for added reliability and validity. This is because granting the GSS is similar to the initial SET ratings on the same scale, diverse scale items concerning the quality of instruction, courses, curriculum admittances, and other subject can offer fresh information (Berk, 2005).

6. Statement of the Problem

The current impasse among the academics is their failure to reach an agreement in connection with the reliability and validity of SET in assessing teaching quality and student learning in universities (Agbetsiafa, 2010; Beran & Rokosh, 2009; Clayson, 2009; Dorasamy & Balkaran, 2013; Otani et al., 2012). Beran and Rokosh (2009) and Madden et al. (2010) hardly found any evidence of the connection between SET and teaching effectiveness observed in this inherent variation. However, some positive relations have been perceived between SET and teaching effectiveness. Because of this essential disparity among the academics, it is arguable that SET alone cannot adequately endorse instructional excellence and student learning across universities, notwithstanding its popularity. In reality, Beran and Rokosh (2009) who noted that, instructors did not consider SET to be a perfect model for teaching effectiveness inherently perceived this as well. This may be true concerning class size, and a SET that is obtainable from small classes is probable unreliable (Ibrahim, 2011). Thus, the research problem is to explore whether the combination of SET and GSS together can produce robust reliability and validity of teaching quality, and student-learning outcomes than the conflicting results derived from SET alone. The knowledge realized will help perfect the current flawed information base concerning coalescing SETs with other evaluation designs for assessing teaching effectiveness (Berk, 2005; Marsh, Ginns, Morin, Nagengast, & Martin, 2011). It will also produce a new perception, generate robust validity and consistency, and expand on the existing methods of evaluating teaching effectiveness and student outcome (El Hassan, 2009; Kozub, 2010; Skowronek, Friesen, & Masonjones, 2011). Without this research, it is probable for the significance of the GSS as a supplementary teaching assessment design may remain indistinct.

7. Purpose of the ESET Tool

The purpose of the ESET is to provide a statistically valid student rating instrument that can be used to determine instructional efficacy. The ESET is completed by currently enrolled students every academic semester at a HBCU via online surveys with set questions on a 5-point Likert scale. The instrumentation is independently administered near the close of every semester by the 'Research, Evaluation, and Planning' division of the HBCU. The tool is electronically released and independently administered to all students at the university every semester. In this manner, every semester courses and faculty are evaluated. Thus the university administration has the students evaluate faculty on teaching effectiveness and overall engagement that leads to student learning.

8. Research Questions

The research questions (Osler and Mansaray, 2013a) listed below were developed to examine the validity and reliability of the Student Ratings of Instruction [SRI/ESET] instrument used in the study to evaluate teaching quality.

Q1: Do ratings completed by students engender internal reliability [consistency] in their measurement of teaching effectiveness?

This question calls for a quantitative research design. The ratings from the survey data at HBCU were used to determine the reliability of the question. This is specified in hypothesis H_{10} and H_{1a}. The statistical tool that will be realistic is the Cronbach's Alpha to verify the reliability [consistency] of the instrument used to evaluate teaching effectiveness.

Q2: Do ratings completed by student ratings produce augmented validity in measuring teaching effectiveness?

The question calls for a quantitative research design. Again, the research made use of existing data at HBCU to determine this question. This is again outlined in hypothesis H_{20} and H_{2a}. The study will apply the factor analysis to determine the validity [construct validity] of the instrument used to determine teaching effectiveness.

9. Research Hypotheses

The following hypotheses (Osler and Mansaray, 2013a) were used to assess the research questions Q1 and Q2. Each research question addresses a null hypothesis with

anticipation of a non–significant association, and an alternative hypothesis that suggests that a significant association does occur between the variables.

H_{10}: The student ratings do not increase the reliability of the instrument used to assess teaching effectiveness.

H_{1a}: The student ratings significantly increase the reliability of the instrument used to assess teaching effectiveness.

H_{20}: The ratings completed by students do not create any validity of the rating instrument used in evaluating teaching effectiveness.

H_{2a}: The ratings completed by students generate increased validity of the rating instrument used in evaluating teaching effectiveness.

9.1 Tri–Squared Test Mathematical Hypotheses

The first sets of Mathematical Hypotheses used in the study in terms of Tri–Squared to determine SRI item efficacy, validity, and reliability were as follows:

H_{10}: $\boldsymbol{Tri^2 = 0}$

H_{1a}: $\boldsymbol{Tri^2 \neq 0}$

9.2 Cronbach's Alpha [α] Mathematical Hypotheses

The second sets of Mathematical Hypotheses used in the study in terms of Cronbach's Alpha to determine reliability were as follows:

H_{10}: $\alpha \leq 0$;

H_{1a}: $\alpha > 0$.

The second sets of Mathematical Hypotheses used in the study in terms of Cronbach's Alpha to determine validity were as follows:

H_{20}: $\alpha \leq 0$;

H_{2a}: $\alpha > 0$.

10. Procedure of Data Analysis: Statistical Models

10.1 Tri–Squared Test [Tri^2]

Tri–Squared comprehensively stands for "The Total Transformative Trichotomous–Squared Test" (or "Trichotomy–Squared"). The Total Transformative Trichotomous–Squared Test provides a methodology for the transformation of the outcomes from qualitative research into measurable quantitative values that are used to test the validity of hypotheses. It is based on the mathematical "Law of Trichotomy". The Total Transformative Trichotomous–Squared Test provides a methodology for the transformation of the outcomes from qualitative research into measurable quantitative values that are used to test the validity of hypotheses. The advantage of this research procedure is that, it is a comprehensive holistic testing methodology that is designed to be static way of holistically measuring categorical variables directly applicable to educational and social behavioral environments where the established methods of pure experimental designs are easily violated. The unchanging base of the Tri–Squared Test is the 3×3 Table based on Trichotomous Categorical Variables and Trichotomous Outcome Variables. This emphasis of the three distinctive variables provide a thorough rigorous robustness to the test that yields enough outcomes to determine if differences truly exist in the environment in which the research takes place (Osler, 2013). The Tri–Squared research procedure uses an innovative series of mathematical formulae that do the following as a comprehensive whole: (1) Convert qualitative data into quantitative data; (2) Analyze inputted trichotomous qualitative outcomes; (3) Transform inputted trichotomous qualitative outcomes into outputted quantitative outcomes; and (4) Create a standalone distribution for the analysis possible outcomes and to establish an effective—research effect size and sample size with an associated alpha level to test the validity of an established research hypothesis. Osler (2012) defined Tri–Squared as:

$$\boldsymbol{Tri^2 = T_{Sum}\left[\left(Tri_x - Tri_y\right)^2 : Tri_y\right]}$$

10.2. Cronbach's Alpha [α]

One of the significant statistical models in this research is the Cronbach's Alpha [α]. It is a valuable coefficient for examining the internal consistency and has been named after Lee Cronbach who first developed it in 1951. Bland and Altam (1997) defined Cronbach's Alpha as:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k} s_i^2}{s_T^2}\right)$$

where, k is the amount of objects s_i^2 is the variance of the i^{th} object and s_T^2 is the variance of the final total created by adding all the objects. In addition, they also said if the objects were not simply added to make the total, but were

initially multiplied by weighting coefficients, then the object must be multiplied by its coefficient ahead of the analysis of the variance s_i^2 Certainly, the formula must contain at least two objects, that is, $k>1$ or α cannot be distinct. Field (2009) also defined the Cronbach's Alpha somewhat differently from that stated by Bland et al. (1997), even when the ideas are similar. Field defined the Cronbach's Alpha as:

$$\frac{N^2 Cov}{\sum s^2_{item} + \sum Cov_{item}}$$

The author noted that, for every object on the scale, two things can be computed: the variance contained in the object, and the covariance amongst an explicit object and any additional object on the scale. Thus, a variance–covariance matrix of the whole objects can be computed. In addition, the author also said, in the matrix, the diagonal rudiments establish the variance contained in an exact object, and the off–diagonal rudiments comprise covariances amid sets of objects. The upper half of the formula is the quantity of objects (N) squared multiplied by the mean covariance amongst objects. The lower half is only the total of all the object variances and object covariances. The arrays of the alpha statistic are between zero and one. Greater the coefficient, better the select items organized together in evaluating the instrument construct, and thus the better the statistical reliability of the assessment tool. An alpha of 1.00 would imply a seamlessly consistent instrument, while a coefficient of zero would imply an untrustworthy tool (Osler and Mansaray, 2013).

10.3 Factorial Analysis

The factorial model used in this study is derived from Agbetsiafa (2010), and Field (2009). Concisely, factor analysis allows the delineation of an essential or hidden configuration in a data set. It accelerates the analysis of the configuration of the associations (correlation) among an outsized number of variables by describing a set of shared essential measurements, commonly termed factors (Agbetsiafa, 2010). Field (2009) noted that, factorial is a mathematical model, resembling a linear equation but without the intercept because the lines intersect at zero and, therefore, the intercept is also zero. Field (2009) defined factorial as:

$Y_i = b_1X_{1i} + b_2X_{2i} + \ldots + b_nX_{ni} + \varepsilon_i$ [The values of b are the loading factors].

Agbetsiafa's (2010) was more detailed in his description of the factorial model in his research than Field (2009). According to Agbetsiafa, it is conceivable to reorient the data to allow the first small number of measurements to explain for much of the existing data. Assuming there is any idling in the data set; it is also conceivable to explain for most of the evidence in the original data with a significantly condensed amount of measurements. Adapting his template, this study also assumes that the 15 items on the student evaluation survey instrument bears relationships with a series of functions working linearly, and they may be represented by the following mathematical formulas:

$Y_1 = \alpha_{10} + \alpha_{11}X_1 + \ldots + \alpha_{1n}X_n + \varepsilon_i$

$Y_2 = \alpha_{20} + \alpha_{21}X_1 + \ldots + \alpha_{2n}X_n + \varepsilon_i$

$Y_3 = \alpha_{30} + \alpha_{31}X_1 + \ldots + \alpha_{3n}X_n + \varepsilon_i.$

$Y_{15} = \alpha_{150} + \alpha_{151}X_1 + \ldots + \alpha_{15n}X_n + \varepsilon_i$

where, Y = a variable with recognized data; α = constant; X_i = the fundamental factors; and, ε_i= the error terms, which help to point out the conjectured associations, are not exhaustive. Thus, applying the technique to the recognized 15 items on the student rating survey instrument, factor analysis describes the unidentified X utilities. The developing loadings from the analysis are the constants, and the factors are the X utilities. The scope of the individual loading for every utility assesses the degree to which the definite utility is associated with the explicit variable (Y). Thus, for any of the 15 variables in equation one of the proposed study, the model may be written as: $Y1 = \alpha_1X_1 + \alpha_2X_2 + \alpha_3X_3 + \ldots + \alpha_nX_n + \varepsilon_i$, where X_{is} denote factors, and α_{is} signify the loadings (Osler and Mansaray, 2013a).

11. ESET Research at an HBCU

The backdrop for this research study is the Research, Evaluation, and Planning Department (REP) of Historic Black College and University. The study will make use of the student rating responses conducted during the spring semester of 2012 by REP. This is a cross–sectional data set. Arguably, REP has the responsibility to coordinate all student, faculty and administrative surveys on behalf of the university. 7,919 students responded to the spring semester

survey. The students were graduates and undergraduates and were from the various departments, schools and colleges within the university, with the exclusion of the Law School. The ratings from the Law School have different components, which are incompatible with the ratings of the rest of the schools and colleges in the University (Osler and Mansaray, 2013b).

11.1 The SRI Measurement Scale

The student ratings of instruction survey is employed to evaluate course instructors and is administered online during the spring and fall semesters of each academic year, with the CourEval assessment tool in a 5–point Likert scale, to all registered students of the University. The rating survey requires students to assess their instructors on 15 items in the assessment tool. The instrument has two subscales. Items 1 to 3 measure the student's efforts in the course, where the scale comprises the following: 1 = never, 2 = not much of the time, 3 = about half of the time, 4 = most of the time and, 5 = all of the time. Items 4 to 15 evaluate the instructor, where the scale comprisesthe following: 1 = strongly disagree, 2 = disagree, 3 = no opinion, 4 = agree and, 5 = strongly agree. This research considers the evaluation of the instructor as an assessment of teaching effectiveness or teaching quality. Also, the instrument has a section where students can make open–ended statements about the instructors, when these are requested by the individual colleges or departments within the University. For the 2012 spring semester ratings survey, the 7,919 responders evaluated instructors on 18,817 courses and course sections offered at HBCU. In addition, the only variables included in this study are the 15 items on the survey instrument with theirrespective ratings.

12. Research Methodology

The choice to use the qualitative into quantitative mixed research methodology is due to its alignment with the problem and purpose statement of the planned study. The study assessed whether student ratings and senior ratings, together can produce robust validity and reliability in assessing teaching efficiency and student learning. The designs for these analyses are the SET and GSS ratings. Evidently, these are student surveys, which are applicable in assessing faculty, and are administered online each semester by the selected university for the projected research. The surveys are on a 5-point Likert scale, which definitely afford them a quantitative position. Arguably, quantitative research hypothesizes the comprehension of the features of a perceived event that can ordinarily be measured, or the evaluation of a probable association between two essentials (Cozby, 2012; Leedy & Ormrod, 2010). In other words, qualitative into quantitative mixed research methodology is a recognized, impartial, methodical procedure in which statistical data are used to obtain data on a phenomenon. Surely, the procedure is employed to classify variables and assess the association between variables. It is against this background that this approach is the most applicable research method for the planned dissertation because the study will only apply numerical data in all its analyses.

13. ESET Statistical Tools used to Analyze Data and Report Results

13.1 Post Hoc Tri–Squared Test Analysis

The application of the Cronbach's Alpha Reliability Model on three factors as qualitative outcomes was used to determine the ESET SRI Efficacy using the Tri–Squared Test statistic.

Data Analyzed using the Trichotomous–Squared Test Standard a Three by Three Table is designed to analyze the research questions and data extracted from an Inventive Investigative Instrument designed with the following Trichotomous Categorical Variables: a_1 = Is the Student Rating of Instruction Instrument effective?; a_2= Is the Student Rating of Instruction Instrument valid?; and a_3= Is the Student Rating of Instruction Instrument consistent? The 3×3 Table has the following Trichotomous Outcome Variables: b_1 = Yes; b_2 = No; and b_3 = No Opinion. The Inputted Qualitative Outcomes are reported as shown in Figure 1 (for $56451_{[\text{Total Items}]} \in 7919_{[\text{Grand Total SRI}]}$) (all results are from Osler and Mansaray, 2013a).

The Tri–Square Test Formula for the Transformation of Trichotomous Qualitative Outcomes into Trichotomous Quantitative Outcomes to Determine the Validity of the Research Hypothesis:

$n_{Tri} = 56451_{[Total\ Items]}$

a = 0.001

TRICHOTOMOUS TESTING

TRICHOTOMOUS INPUT VARIABLES

TRICHOTOMOUS RESULTS		a_1	a_2	a_3
	b_1	15629	15751	15125
OUTPUT VARIABLES	b_2	1222	1058	1406
	b_3	1966	2008	2286

$$Tri^2 d.f. = [C-1][R-1] = [3-1][3-1] = 4 = Tri^{2}_{[\bar{x}]}$$

Figure 1. Post Hoc Tri-Squared Test

$$Tri^2 = T_{Sum}\left[\left(Tri_x - Tri_y\right)^2 : Tri_y\right]$$

Tri^2 Critical Value Table = 18.467 (with d.f. = 4 at α = 0.001). For d.f. = 4, the Critical Value for p >0.001 is 18.467. The Calculated Tri–Square value is 92.531, thus, the null hypothesis (H_0) is rejected by virtue of the hypothesis test which yields the following: Tri–Squared Critical Value of 18.467<92.531 Calculated Tri–Squared Value. Results: (1) Tri–Squared Calculated Value = 92.531; (2) Tri–Squared Degrees of Freedom = 4; (3) Tri–Squared Probability = 0.0096; 4) Tri–Squared Alpha Level = 0.001 [for n_{Tri} = 56451[Total Items] Maximized Test Critical Value].

13.2 Tri–Squared Percentage Deviations

Percentage deviation and standardized residual are both measures of the degree to which an observed Tri–Squared cell frequency differs from the value that would be expected on the basis of the null hypothesis. Figure 2 shows the Tri–Squared Percentage Deviations.

13.3 Tri–Squared Standardized Residuals

The standardized residual for a cell in a Tri–Squared table is a version of the standard normal deviate, "z_{Tri}", calculated as follows

$$z_{Tri} = \frac{Tri_x - Tri_y}{\sqrt{Tri_y}}$$

	a_1	a_2	a_3
b_1	+0.8%	+1.6%	–2.4%
b_2	–0.5%	–13.8%	+14.4%
b_3	–5.8%	–3.8%	+9.6%

Figure 2. Tri–Squared Percentage Deviations

	a_1	a_2	a_3
b_1	+1.02	+2.00	–3.03
b_2	–0.18	–4.85	+5.06
b_3	–2.64	–1.72	+4.36

Figure 3. Tri–Squared Standardized Residuals

Where, z_{Tri} = The Tri–Squared Calculated Standard Normal Deviate;Tri_x = Trichotomous Qualitative Outcomes; andT_{riy} = Trichotomous Quantitative Outcomes. Assuming the null hypothesis to be true, values of the standardized residual belong to a normally distributed sampling distribution with a mean of zero and a standard deviation of ±1.0. Figure 3 shows the Tri–Squared Standardized Residuals.

13.4 Goodman & Kruskal's Lambda (λ) Tri–Squared Results

Goodman & Kruskal's Lambda (λ) is a cross tabulation analysis measure of proportional reduction in error. Lambda indicates the extent to which the modal categories and frequencies for each value of the independent variable differ from the overall modal category and frequency. The Goodman–Kruskal Values for Lambda range from zero (indicating that there is "no association" between independent and dependent variables) to one (indicating a "perfect association" between independent and dependent variables). It is calculated with the following equation: $\lambda = \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_1}$ where,ε_1 = is the overall non–modal frequency; and ε_2 = is the sum of the non–modal frequencies for each value of the independent variable. Table 1 shows the Goodman & Kruskal's Lambda (l) Tri–Squared Results.

Cross Tabulation of Variables		Independent Variables: Categorical Variable1 = a_1	Categorical Variable 2 = a_2	Categorical Variable 3 = a_3	Results: =
Dependent Variables	Outcome Variable 1 = b_1	15629	15751	15125	46505
	Outcome Variable 2 = b_2	1222	1058	1406	3686
	Outcome Variable 3 = b_2	1966	2008	2286	6260
Results:		18817	18817	18817	56451

Table1. Goodman & Kruskal's Lambda (λ) Tri–Squared Table

14. Summaryand Recommendations

The research employed the responses from the ESET ratings to assess instructional (teaching) effectiveness and student success (its inherent outcome). Students were requested to rate the numerical variables on the ESET survey on a 5–point Likert scale. Exploratory data analysis as a means of measurement is significant for the numerical identified independent variables which were used to determine the level at the research hypotheses. In addition, the Cronbach's Alpha statistic was used to measure and assess the internal consistency (reliability) of the ESET instruments employed to students to evaluate faculty teaching efficacy. In addition, Factorial Analysis is another applicable statistic used in the study, to determine the construct validity of the ESET instrumentation. The authors make the following recommendations regarding the use of ESET as a novel universal SRI (Student Ratings of Instruction) tool:

(1) The ESET and associated measurement procedures be officially support by the institution of its origin as a base operating procedure for the continuous improvement of ongoing collegiate teaching and learning.

(2) The ESET structure needs to be commercialized and branded as supportive research regarding magnitude clearly illustrates that the ESET has great value in regards to learning measurement and assessment.

(3) The ESET model be adopted by more institutions so that it becomes an "industry standard" at more similar or like institutions of higher learning.

Conclusion

Student ratings are immensely common to evaluate teaching quality in many Universities worldwide. There are, however, some disagreements relating to their validity and reliability as a measurement of teaching effectiveness. The ESET design offers a statistically quantifiable methodology that can be readily used by any institution seeking to effectively evaluate its faculty teaching efficacy. This novel instrumentation is also applicable for faculty assessment in the self–assessmentof learner performance. When the ESET instrument is effectively employed in conjunction with its multiplicative statistical methodology; the joint analytical designs reveal the actively robust validity and reliability of the ESET instrument. The ESET (as a measurement of teaching effectiveness and student learning) is an exceedingly usable arithmetical implement. Information on data gathering, suppositions, hypotheses, and the nature of the research methodologies and statistical models were revealed in this study. The use of the ESET in future research will generate new ideas on the uses of instructional effectiveness metrics to determine curricular outcomes for formative program evaluation (as "curriculum development") and summative evaluation for faculty assessment purposes (expressed as: tenure, faculty merit pay increases, and the retention of instructors for vital non-tenured [adjunct] faculty positions).

References

[1]. Agbetsiafa, D. (2010). "Evaluating effective teaching in college level Economics using student ratings of instruction: A factor analytic approach", *Journal of College Teaching and Learning,* 7(5), 57-66. Retrieved from http://journals.cluteonline.com/index.php/TLC

[2]. Amrein-Beardsley, A., and Haladyna, T. (2012). " Validating a theory-based survey to evaluate teaching effectiveness in higher education", *Journal on Excellence in College Teaching*, Vol. 23(1), pp. 17-42. Retrieved from http://celt.muohio.edu/ject/issue.php?v=20&n=2

[3]. Anastasiadou, S. D. (2011). "Reliability and validity testing of a new scale for measuring attitudes toward learning statistics with technology", *Acta Didactica Napocensia*, Vol. 4(1), 1-10. Retrieved from http://adn.teaching.ro/

[4]. Arthur, L. (2009). "From performativity to professionalism: lecturers' responses to student feedback", *Teaching in Higher Education*, Vol.14(4), pp. 441-454. doi:10.1080/13562510903050228

[5]. Baker, L. R., Pollio, D. E., and Hudson, A. (2011). "Assessing student perception of practice evaluation knowledge in introductory research methods". *Journal of Social Work Education*, Vol. 47(3), pp.555-564. Retrieved from http://www.tandfonline.com/toc/uswe20/current#.U5yC8WXD8dU

[6]. Beran, T., and Rokosh, J. (2009). "Instructors' perspectives on the utility of student ratings of instruction",

Instructional Science, Vol. 37(2), pp. 171-184. doi:10.1007/s11251-007-9045-2

[7]. Beran, T., Violato, C., and Kline, D. (2007). "What's the "Use" of student ratings of instruction foradministrators? One university's experience", *Canadian Journal of Higher Education*, Vol. 37(1), pp. 27-43. Retrieved from http://ojs.library.ubc.ca/index.php/cjhe/index

[8]. Beran, T., Violato, C., Kline, D., and Frideres, J. (2009). "What do students consider useful aboutstudent ratings?", *Assessment & Evaluation in Higher Education*, Vol. 34(5), pp. 519-527. doi:10.1080/02602930802082228

[9]. Berk, R. A. (2005). "Survey of 12 strategies to measure teaching effectiveness", *International Journal of Teaching Learning in Higher Education*, Vol. 17(1), pp. 48–62. Retrieved from http://www.isetl.org/ijtlhe/

[10]. Brookfield, S. (1986). *Understanding and facilitating adult learning*. San Francisco CA: Jossey-Bass.

[11]. Carrell, S. E., and West, J. E. (2010). "Does professor quality matter? Evidence from random assignment of students to professors", *Journal of Political Economy*, Vol. 118(3), pp. 409-432. Retrieved from http://www.jstor.org/action/showPublication?journalCode=jpoliecon&

[12]. Chen, G., and Watkins, D. (2010). "Stability and correlates of student evaluations of teaching at a Chinese university", *Assessment & Evaluation in Higher Education*, Vol. 35(6), pp.675-685. doi:10.1080/02602930902977715

[13]. Chulkov, D. V., and Van Alstine, J. (2012). "Challenges in designing student teaching evaluations in a business program", *International Journal of Educational Management*, Vol. 26(2), pp.162-174. doi:10.1108/09513541211201979

[14]. Clayson, D. E. (2009). "Student evaluations of teaching: Are they related to what student learn?", *Journal of Marketing Education*, Vol. 31(1), pp.16-30. Retrieved from http://jmd.sagepub.com

[15]. Cozby, P. C., and Bates, S. C. (2012). *Methods in behavioral research,* 11th ed., New York, NY: McGraw-Hill.

[16]. Davenport, J., and Davenport, J. A. (1985). "A chronology and analysis of the andragogy debate", *Adult Education Quarterly,* Vol. 35(3), pp. 152-159. Retrieved from http://aeq.sagepub.com

[17]. Donnon, T., Delver, H., and Beran, T. (2010). "Student and teaching characteristics related to ratings of instruction in medical sciences graduate programs", *Medical Teacher*, Vol. 32(4), pp.327-332. doi:10.3109/01421590903480097

[18]. Dorasamy, N., and Balkaran, R. (2013). "Role of student ratings of lecturers in enhancing teaching at higher education institutions: A case study of the durban university of technology", *Journal of Economics and Behavioral Studies*, Vol. 5(5), pp. 268-281. Retrieved from http://ifrnd.org/JournalDetail.aspx?JournalID=2

[19]. El Hassan, K. (2012). "Investigating substantive and consequential validity of student ratings of instruction". *Higher Education Research & Development*, Vol. 28(3), pp. 319-333. doi:10.1080/07294360902839917

[20]. Frick, T., Chadha, R., Watson, C., Wang, Y., and Green, P. (2009). "College student perceptions of teaching and learning quality", *Educational Technology Research & Development,* Vol. 57(5), pp.705-720. doi:10.1007/s11423-007-9079-9

[21]. Fitzpatrick, J. L., and Miller-Stevens, K. (2009). "A case study of measuring outcomes in an MPA program", *Journal of Public Affairs Education*, Vol. 15(1), pp.17-31. Retrieved from http://www.naspaa.org/initiatives/jpae/jpae.asp

[22]. Forrest III, S., and Peterson, T. (2006). "It's called andragogy. *Academy of Management Learning and Education*", Vol. 5(1), pp. 113-122. doi:10.5465/AMLE.2006. 20388390

[23]. Galbraith, C. S., and Merrill, G. B. (2012). "Predicting student achievement in university-level business and economics classes: Peer observation of classroom instruction and student ratings of teaching effectiveness", *College Teaching*, Vol. 60(2), pp. 48-55. doi:10.1080/87567555.2011.627896

[24]. Galbraith, C., Merrill, G., and Kline, D. (2012). "Are student evaluations of teaching effectiveness valid for measuring student learning outcomes in business related classes? A neural network and bayesian analyses", *Research in Higher Education*, Vol. 53(3), pp. 353-374. doi:10.1007/s11162-011-9229-0

[25]. Gilstrap, D. (2013). "Why do we teach? Adult learning theory in professional standards as a basis for curriculum development". *College and Research Libraries*, Vol. 74(5), pp. 501-518. Retrieved from http://crl.acrl.org/

[26]. Haladyna, T., and Amrein-Beardsley, A. (2009). "Validation of a research-based student survey of instruction in a college of education", *Educational Assessment, Evaluation & Accountability*, Vol. 21(3), pp. 255-276. doi:10.1007/s11092-008-9065-8

[27]. Hartree. A. (1984). "Mialcolm Knowles' theory of androgogy: A critique", *International Journalof Lifelong Education*, Vol. 3(3), pp. 203-210. Retrieved from http://www.tandfonline.com

[28]. Hatfield, C. L., and Coyle, E. A. (2013). "Factors that influence student completion of course and faculty evaluations", *American Journal of Pharmaceutical Education,* Vol. 77(2), pp.1-4. Retrieved from http://www.ajpe.org/

[29]. Ibrahim, A. M. (2011). "Using generalizability theory to estimate the relative effect of class sizeand number of items on the dependability of student ratings of instruction", *Psychological Reports*, Vol. 109(1), pp. 252-258. doi:10.2466/03.07.11.PR0.109.4.252-258

[30]. Johnson, M. D., Narayanan, A., and Sawaya, W. J. (2013). "Effects of course and instructor characteristics on student evaluation of teaching across a college of engineering", *Journal of Engineering Education,* Vol. 102(2), pp. 289-318. doi:10.1002/jee.20013

[31]. Keeley, J., Furr, R., and Buskist, W. (2010). "Differentiating psychology students' perceptions of teachers using the teacher behavior checklist", *Teaching of Psychology,* Vol. 37(1), pp.16-20. doi:10.1080/00986280903426282

[32]. Knowles, M. S. (1974). "Human resources development in OD", *Public Administration Review*, Vol. 34(2), pp. 115-123. Retrieved from http://onlinelibrary.wiley.com/ journal/10.1111/(ISSN)1540-6210

[33]. Knowles. M . S. (1980). *The Modernl Practice of Adult Education: From Pedagogy to Androgogy*. (2nd ed.) New York, NY: Cambridge Books.

[34]. Knowles, M. (1984). *The adult learner: A neglected species*. 3rd ed., Houston, TX: Gulf.

[35]. Kozub, R. M. (2010). "Relationship of course, instructor, and student characteristics to dimensions of student ratings of teaching effectiveness in business schools", *American Journal of Business Education*, Vol. 3(1), pp.33-40. Retrieved from http://journals.cluteonline.com/index.php/AJBE

[36]. Leedy, P. D., and Ormrod, J. E. (2009). *Practical research: Planning and design,* (9th ed.). UpperSaddle River, NJ: Pearson Education Inc.

[37]. Liu, O. L. (2012). "Student evaluation of instruction: In the new paradigm of distance education", *Research in Higher Education*, 53(4), 471-486. doi:http://dx.doi.org/10.1007/s11162-011-9236-1

[38]. Madden, T. J., Dillon, W. R., and Leak, R. L. (2010). "Students' evaluation of teaching: Concernsof item diagnosticity", *Journal of Marketing Education*, Vol. 32(3), pp. 264-274. doi:10.1177/0273475310377759

[39]. Marsh, H. W., Ginns, P., Morin, A. S., Nagengast, B., and Martin, A. J. (2011). "Use of student ratings to benchmark universities: Multilevel modeling of responses to the Australian Course Experience Questionnaire (CEQ)", *Journal of Educational Psychology,* Vol. 103(3), pp. 733-748. doi:10.1037/a0024221

[40]. Mau, R. R., and Opengart, R.A. (2012). "Comparing ratings: In-class (paper) vs. out of class (online) student evaluations". *Higher Education Studies*, Vol. 2(3), pp. 55. doi:10.5539/hes.v2n3p5

[41]. Merriam, S B. Mott, V. W., and Lee, M. (1996). "Learning that comes from the negative interpretation of life experience", *Studies in Continuing Education*, Vol. 18(1), pp.1-23. Retrieved from http://www.tandfonline.com/

[42]. O'Shaughnessy, L. (2013). *6 Challenges facing state universities*. CBSNews. Retrieved from http://www.cbsnews.com/news/6-challenges-facing-state-universities/

[43]. Osler, J. E. and Mansaray M. A. (2013a). "The Application of Online Psychometric Tests: An Analysis of Student Ratings of Instruction", *i-manager's Journal on Educational Psychology*, Vol. 7 (3), pp. 32–50.

[44]. Osler, J. E., and Mansaray, M. A. (2013b). "Applied

mathematical statistical modeling to determine the validity and reliability of independent instructional measures". *i-manager's Journal on Mathematics,* Vol. 2(3), pp. 12-30. Retrieved from http://www.imanagerpublications.com/JournalIntroduction.aspx?journalid=102

[45]. Otani, K., Kim, B., and Jeong-IL, C. (2012). "Student evaluation of teaching (SET) in higher education: How to use SET more effectively and efficiently in public affairs education", *Journal of Public Affairs Education*, Vol. 18(3), pp. 531-544. Retrieved from http://www.naspaa.org/initiatives/jpae/jpae.asp

[46] Pritchard, R. E., and Potter, G. C. (2011). "Adverse changes in faculty behavior resulting from use of student evaluations of teaching: A case study", *Journal of College Teaching and Learning,* Vol. 8(1), pp. 1-7. Retrieved from http://journals.cluteonline.com/index.php/TLC

[47]. Ruppert, B., and Green, D. A. (2012). "Practicing what we teach: Credibility and alignment in the business communication classroom", *Business Communication Quarterly*, Vol. 75(1), pp. 29-44. doi:10.1177/1080569911426475

[48]. Safavi, S., Bakar, K., Tarmizi, R., and Alwi, N. (2012). "The role of student ratings of instruction from perspectives of the higher education administrators", *International Journal of Business and Social Science*, Vol. 3(9), pp. 233-239. Retrieved from http://www.ijbssnet.com/

[49]. Savicevic, D. M. (1991). "Modern conceptions of andragogy: A European framework". *Studies in the Education of Adults*, Vol. 23(2), pp. 179. Retrieved from http://www.niace.org.uk/publications/academic-journals/studies

[50]. Seyedeh, A. S., and Kamariah, A. B. (2013). "The utility of student evaluations for medical sciences teachers and administrators", *International Journal of Business and Social Science*, Vol. 4(13). Retrieved from http://www.ijbssnet.com/

[51]. Seyedeh, A. S., Kamariah, A. B., Rohani, A. T., and Alwi, N. H. (2012). "The role of student ratings of instruction from perspectives of the higher education administrators", *International Journal of Business and Social Science*, Vol.3(10). Retrieved from http://www.ijbssnet.com/

[52]. Skowronek, J., Friesen, B., and Masonjones, H. (2011). "Developing a statistically valid and practically useful student evaluation instrument", *International Journal for the Scholarship of Teaching and Learning*, Vol. 5(1), pp. 1-19. Retrieved from http://digitalcommons.georgiasouthern.edu/ij-sotl//

[53]. Slocombe, T., Miller, D., and Hite, N. (2011). "A survey of student perspectives toward faculty evaluations", *American Journal of Business Education*, Vol. 4(7), pp. 51-57. Retrieved from http://journals.cluteonline.com/index.php/AJBE

[54]. Smith, D. L., Cook, P., and Buskist, W. (2011). "An Experimental analysis of the relation between assigned grades and instructor evaluations", *Teaching of Psychology*, Vol. 38(4), 225-228. doi:10.1177/0098628311421317

[55]. Spooren, P. (2010). "On the credibility of the judge: A cross-classified multilevel analysis on students' evaluation of teaching", *Studies in Educational Evaluation*, Vol. 36(4), pp. 121-131.

[56]. Spooren, P., Brockx, B., and Mortelmans, D. (2013). "On the validity of student evaluation of teaching: The state of the art", *Review of Educational Research*, Vol. 83(4), pp. 598. doi:10.3102/0034654313496870

[57]. Spooren, P. P., Mortelmans, D. D., and Denekens, J. J. (2007). "Student evaluation of teaching quality in higher education: Development of an instrument based on 10 Likert-scales", *Assessment & Evaluation in Higher Education,* 32(6), 667-679. doi:10.1080/02602930601117191

[58]. Stowell, J. R., Addison, W. E., and Smith, J. L. (2012). "Comparison of online and classroom-based student evaluations of instruction", *Assessment and Evaluation in Higher Education*, Vol. 37(4), pp. 465-473. doi:10.1080/02602938.2010.545869

[59]. Taylor, C. L., Grey, N., and Satterthwaite, J. D. (2013). "Assessing the clinical skills of dental students: A review of the literature", *Journal of Education and Learning*, Vol. 2(1), pp. 20-31. Retrieved from http://www.ccsenet.org/ journal /index.php/jel

[60]. Venette, S., Sellnow, D., and McIntyre, K. (2010).

"Charting new territory: assessing the online frontier of student ratings of instruction". *Assessment and Evaluation in Higher Education*, Vol. 35(1), pp. 97-111. doi:10.1080/02602930802618336

[61]. Young, N. (2012). "Effective learning and teaching of RDA: Applying adult learning theory". *Journal of Library Metadata*, Vol. 12(2/3), pp. 188-198. doi:10.1080/19386389.2012.699836

[62]. Young, S., and Duncan, H. E. (2014). "Online and face-to-face teaching: How do student ratings differ?". *Journal of Online Learning and Teaching*, Vol. 10(1), pp. 70-79. Retrieved from http://jolt.merlot.org/

[63]. Zhao, J., and Gallant, D. J. (2012). "Student evaluation of instruction in higher education: exploring issues of validity and reliability", *Assessment & Evaluation in Higher Education*, Vol. 37(2), pp. 227-235. doi:10.1080/02602938.2010.523819

[64]. Zmeyov, S. I. (1998). "Andragogy: Origins, developments and trends", *International Review of Education/Internationale Zeitschrift Für Erziehungswissenschaft*, Vol. 44(1), pp. 103-108. Retrieved from http://link.springer.com/journal/11159

ABOUT THE AUTHORS

James EdwardOsler II is an accomplished Artist, Teacher, and Researcher. Currently, he is a Faculty Member of the North Carolina Central University. His research foci are: (1) Novel Statistical Metrics; and (2) Instructional Systems Design Informatics. He has authored and published research articles, books, and multimedia on topics designed to encourage the development of Education as a Science. Osler's additional interests include: Ministry, Fundamental Christian Education; Mathematical foundations of Research; Courseware Design; Gamification; Illustration; Invention Metrics; Teaching; Consulting; and Service-based Initiatives. In his professional career he has served as the Program Coordinator of the Graduate Program in Educational Technology and authored the Graduate Program in Online Instructional Design. He has received three of the highest and most respected honors at NCCU: "The Employee Recognition Award for Outstanding Service" and "The University Award for Teaching Excellence", and the first "Chancellor's Award for Innovation".

Mahmud Mansaray is a Research Analyst in the Department of Research, Evaluation and Planning at the North Carolina Central University. Prior to this, he was the Assessment Specialist at the University of North Carolina General Administration (UNCGA) for five years, At UNCGA, Mahmud was in charge of all student surveys received from the 16 campuses within the system. He was also involved in research activities relating to Educational Policy Analyses. Mahmud is a graduate of Fourah Bay College, the University of Sierra Leone and holds a Bachelor of Art degree with Honors in Geography. He also holds a Master's degree in Applied Economics and Policy Analysis from North Carolina State University. He is also currently involved in Doctoral studies in Policy Analysis, Statistics and Research Designs.