International Journal of Academic Research in Business and Social Sciences
2017, Vol. 7, No. 5
ISSN: 2222-6990

Applied Statistical Models of Assessment and Evaluation of Instructional Efficiency

Mahmud A. Mansaray (PhD)
Department of Research, Evaluation, and Planning, North Carolina Central University
1801 Fayetteville St., Durham, NC 27707, USA
Email: mmansara@nccu.edu

Philliph M. Mutisya (EdD)
Professor, Department of Curriculum and Instruction, School of Education
North Carolina Central University, 1801 Fayetteville St., Durham, NC 27707
USA
Email: pmmutisya@nccu.edu

DOI: 10.6007/IJARBSS/v7-i5/2928 URL: http://dx.doi.org/10.6007/IJARBSS/v7-i5/2928

Abstract
The purpose of the contemporary research was the determination of the validity and reliability of the graduating senior exit survey (GSS) as a possible supplementary instrument, to assess faculty on teaching effectiveness at a university in the southeastern US. The research approach was quantitative, and the dataset was the 426 students' responses on specific elements of teaching effectiveness for the fall 2016 semester GSS survey at the university. Applying the Cronbach's alpha statistic, the findings revealed the alpha for the 9 elements on the rating scale, which together assessed teaching effectiveness at the university, was .95, an indication the elements established a scale that had practical internal stability reliability for the GSS instrument. In addition, applying the principal axis factoring with varimax rotation, indicated the three indexes of teaching effectiveness, including motivation, competence, and expertise had strong positive loadings of > .40, and interconnected significantly with teaching effectiveness, thus endorsing the construct validity of the GSS instrument in assessing teaching efficacy. However, the research was limited to one semester analysis of the GSS instrument, which may be inadequate to establish the generalizability of the instrument as a measurement of teaching efficiency. Subsequent studies must include additional semesters to determine generalizability.

Keywords Teaching Effectiveness, Cronbach's Alpha, Principal Axis Factoring, Varimax Rotation, and Faculty Assessment.

1. Introduction

The twenty-first century appears to have reflective challenges in the antiquity, principles, and administration of higher education in the United States (US). Recently, several

International Journal of Academic Research in Business and Social Sciences
2017, Vol. 7, No. 5
ISSN: 2222-6990

US states' governments have noticeably disinvested in public education (Mitchell & Leachman, 2015), and a few states have recognized performance-based allocation to improve state universities regarding degree production (Dougherty & Natow, 2015; Ellis & Bowden, 2014; Miao, 2012; O'Shaughnessy, 2013; Umbricht, Fernandez, & Ortagus, 2015). In financing colleges based on their productivities instead of inputs, state legislators hold that the inducements will inspire colleges to upsurge degree production (Hillman, Tandberg, & Fryar, 2015). However, performance-financing policies by state legislators have not expressively associated with the expansion of total degrees conferred (Hillman et al., 2015; Rutherford & Rabovsky, 2014; Sanford & Hunter, 2011; Shin, 2010). One effect of this policy appears to be that, several US public universities subject to performance-based financing are presently receiving reduced funding from their state legislatures. Subsequently, some universities are progressively into self-funding some of their programs, including raising tuitions, and requiring library fees and Ed Tech fees, inter alia, from their students. Given this, the exclusive burden of searching for additional funds to subsidize university programs progressively falls within the realm of the university administrators (chancellors, provosts, deans, and chairs). This comes alongside the relentless desire of enrolled students for superior college grades. Consequently, the desire to upsurge student enrollment and retention, and the burden to fulfil the students' craving for excellent grades (Agbetsiafa, 2010), maybe critical in influencing college administrators to increasingly apply distinct techniques of evaluating the quality of education, to articulate resolutions on teaching and learning quality, college programs, and faculty employment and tenure, among others.

An assessment instrument often exploited by college administrators in public universities for personnel policy resolution-makings, including supplementary summative purposes (tenure, merit increase, retention for non-tenured), is the student evaluation of teaching (SET), also called the student ratings of instruction (SRI) (Agbetsiafa, 2010; Seyedeh & Kamariah, 2013; Taylor, Grey, & Satterthwaite, 2013). Agbetsiafa (2010) argued SRIs were gradually attaining significance in summative and formative processes in the universities because they present strategic, methodical, and valued methods of obtaining feedback on students' responses to instructors and courses. Nevertheless, even when several academics may realize SRI designs are dependable apparatuses, there is fewer unity concerning their overall validity and reliability regarding the level at which the implements correctly evaluate concrete terms (e.g. teaching quality), or present a comprehensive scoring of the course or instructor (Agbetsiafa, 2010; Clayson, 2009; Osler & Mansaray, 2013, 2015). The conflict among the academics to agree on the validity and reliability of the SRI in assessing faculty on teaching effectiveness has encouraged others to sanction the advancement of the design, to embrace supplementary measurements of assessing teaching efficacy (e.g. self-rating student interviews, peer reviews, etc.) (El Hassan, 2012; Marsh, Ginns, Morin, Nagengast, & Martin, 2011). Embracing supplementary measurements of teaching efficiency is to overcome some of the preferences possibly inherent in the SRI.

Nevertheless, because of the growing engagement of SRI in university policy resolutions, and the inconsistency of the researchers to agree on the validity and reliability of the design in evaluating faculty and teaching quality, it is essential to sanction the use of the design to

include supplementary measurements of evaluating teaching effectiveness and student learning. Thus, the key purpose of the current research is the exploration of the significance of the graduating senior exit survey (GSS) as a supplement to the SRI model in evaluating teaching effectiveness and student learning outcomes for administrative policy resolutions. This would involve investigating the validity and reliability of selected elements on the GSS for possible application in assessing teaching effectiveness and student learning. The ultimate purpose of the current research is the possibility of validating the GSS as an instrument applicable in teaching and public administrative policies, to apprise and affect instruction in higher education at a traditionally black university in the southeastern US.

The need to conduct the contemporary research is significant because, the GSS instrument, like the SRI, is also applicable in assessing faculty for effective teaching and student learning. However, the employment of the GSS ratings for faculty assessment on teaching quality is unsubstantiated, even when some elements on the design are relevant for teaching assessment determinations. In addition, there are hardly any acknowledged peer-reviewed journal articles on the singular impact of the GSS ratings on teaching effectiveness and student learning outcomes. This provides the motivation and support for a theoretical examination of the validity and reliability of the GSS survey as an assessment tool for teaching effectiveness and student learning outcomes. Ultimately, the current research is perhaps one of the first to explore the significance of the GSS design as a supplement to the SRI model, to evaluate teaching effectiveness and student learning for administrative policy resolutions, and applicable generalizability in higher education.

In total, it is obvious the GSS design is applicable to evaluate teaching effectiveness, but information is lacking in the current journals concerning the validity and reliability of the GSS tool. This deficit aids in reinforcing the importance of an empirical and theoretical investigation on the validity and reliability of the GSS instrument for faculty evaluation and student learning outcomes. The research is significant because of its projection to fill the prevailing gaps regarding supplementary designs on faculty assessment in teaching efficacy, in addition to producing the principal position of exploring a novel design on teaching effectiveness for summative and formative resolutions. The outcomes hold practical importance for educational policy makers, faculty, and students, along with adding a new knowledge to the outwardly infinite journals on the evaluation of teaching effectiveness in higher education.

2. Literature Review

2.1. Selected Models on Teaching Effectiveness

A key objective of the current research is the examination of the importance of the GSS design as a supplemental procedure to the SRI model in assessing teaching efficiency and student learning outcomes for administrative policy decisions. Teaching effectiveness was the magnitude at which an instructor empowered learners to effect their academic goals (Mckeachie, 1979). Given this, there are several existing theoretical models on teaching effectiveness (Apodaca & Grad, 2005; Chen & Hoshower, 2003; Mittal & Gera, 2013; Seidel & Shavelson, 2007; Shevlin, Banyard, Davies, & Griffiths, 2000), espoused in the current literature. Chen and Hoshower (2003), for instance, utilized the expectancy theory, initially advanced by

Vroom (1964) in their exploration on student evaluation of teaching. Chen and Hoshower (2003) noted that expectancy models were cognitive clarifications of human conduct that ascribe an individual as a dynamic, thoughtful, foretelling being in his/her location. The authors progressed that the individual unceasingly assessed the results of his or her conduct and intuitively evaluated the prospect that each of his or her conceivable activities led to distinct conclusions. Grounded on this systematic analysis, Chen and Hoshower (2003) surmised that the student would regulate the level of effort he or she would want to utilize in partaking in the evaluation structure.

Apodaca and Grad (2005), alternatively, argued on the theory of teaching effectiveness from a student learning methodology, in particular, the learning theory. Entwistle (1987) was an earlier pioneer of the learning approach design through his postulation of the heuristic model for analyzing the teaching-learning process at higher education. Apodaca and Grad (2005) noted that the heuristic model centered on the features that may affect the learning approaches, procedures, and outcomes of the student, the teaching style, and the institutional framework, including learning resources, and instruction feedback, among others. Thus, it seems using the heuristic model to evaluate teaching and learning outcomes would compel a multifaceted methodology of assessing teaching efficacy. Apodaca and Grad (2005) also affirmed the significance of the learning theory in evaluating teaching efficiency when they noted the assessment of teaching efficacy centered on students' ratings must consider the philosophies for effective teaching resulting from the learning theory, inter alia.

Shevil et al. (2000) postulated a conceptual model of teaching effectiveness and charisma factors, which examined the underlying feature of the charisma of the faculty as a valued element in the students' forecast of teaching efficiency ratings. Shevil et al. argued that charisma was such a leading quality in students' opinion of educators that it influenced evaluation of teacher efficiency. Mittal and Gera (2013) adapted Shevil et al.'s (2000) model of teaching effectiveness and charisma features in their research on student evaluation of teaching effectiveness in higher education in India. Applying both exploratory factor and confirmatory factor analyses, Mittal and Gera (2013) established that students' discernment of the appeal of their instructor described a substantial percentage of the disparity of student assessment of instruction instead of the individual scores of measurements of "lecturer ability", and "module attributes", the two measurable sectors in the model. Mittal and Gera's (2013) research is consistent with Shevil et al.'s (2000) model on teaching effectiveness.

Seidal and Shavelson (2007) reviewed a few teaching effectiveness models, in particular, the Scheerens and Bosker (1997), and Fraser, Walberg, Welch, and Hattie (1987) process-product models, including the Bolthuis (2003) cognitive model of teaching and learning. Fraser et al. (1987), and Scheerens and Bosker's (1997) models together underlined a number of teaching progression elements, including reinforcement, prompts and feedback, tutoring, and teaching expectancy, among others, which positively influenced student learning aftermaths. The Fraser et al's. (1987) paradigm engaged on five teaching apparatuses with the maximum outcome extents, including reinforcement, acceleration, reading training, prompts and feedback, and science proficiency. Correspondingly, Scheerens and Bosker's (1997) model underlined teaching responsibilities, including reinforcement, feedback, cooperative education,

differentiation/adaptive teaching and time on assignment, which generated the supreme outcome extent. In all the given models, effective teaching and student learning seems the core focus and product, even if they diverged in their methodologies and applications. Thus, the current research will incorporate a few core elements of teaching effectiveness applied in some of the reviewed effective teaching models, including inferences on feedbacks, and motivating students to learn, inter alia.

2.2. Teaching Effectiveness Measurement

There is an evolving listing of research journals on teaching effectiveness, and the measurement of teaching efficacy in higher education. Nevertheless, the striving is the pertinent delineation of teaching effectiveness. Cohen (1981), and Feldman (1989), for example, acknowledged that, teaching effectiveness was the capacity of information realized by learners in a course. Mckeachie (1979) similarly noted teaching effectiveness was the scale upon which a mentor sanctioned learners to upshot their academic objectives. Barry (2010), meanwhile, noted teaching efficacy embraced a detailed understanding of subject, learning perception and student divergences, organization, classroom instructional methods, recognizing apparent learners, and assessment of student understanding and capability of learning outcomes. Correspondingly, Hassel (2009) said the fundamental requirement of teacher efficacy must be the student education results. This is evidently the dimension of the student learning achievement, including supplementary valued impacts. There is obviously the existence of a deliberation in the current literature on the subject of the complete delineation of teaching effectiveness in higher education and its measurement methods.

Nevertheless, several research explorations on the evaluation of college faculty on teaching excellence and students' achievements exploit the student ratings of instruction (SRI) survey design (Agbetsiafa, 2010; Chen & Watkins, 2010; Donnon, Delver, & Beran, 2010; El Hassan, 2009; Hatfield & Coyle, 2013; Jones, 2010; Keeley, English, Irons, & Henslee, 2013; Osler & Mansaray, 2013, 2015; Zhao & Gallant, 2012). Donnon et al. (2010), for example, said college faculty engaged SRIs to accumulate students' responses regarding their courses and document growth in their tutoring shares and accountabilities, which may hold a notable consequence on their careers (Sprinkle, 2008). Keeley et al. (2013) had also argued SRIs were in general a relevant tool engaged by higher education institutes in evaluating their lecturers' teaching efficiency. In addition, Osler and Mansaray (2013) noted the use of SRI was to effect knowledge on the students, including the establishment of administrative resolutions, for instance, the award of continued tenure and promotion. Furthermore, SRIs were relevant as administrative contraptions to assess faculty progression, decisions, and tenure determination, because they measure characteristics of lecturers' teaching propensity and the features of the offered course (Beran & Violato, 2005; Heckert, Latier, Ringwald-Burton, & Drazen, 2006). Zhao and Gallant (2012) likewise advanced that a number of personnel functioning groups at US universities and colleges utilized SRIs to generate answers regarding tenure, promotion, merit reimbursement, or faculty proficiency development. In the present, Liu (2012) had advanced that, SRI was a noteworthy component in indicating the dependability of distance learning, and

International Journal of Academic Research in Business and Social Sciences
2017, Vol. 7, No. 5
ISSN: 2222-6990

was primarily relevant in higher education for strategic planning, program improvement, and faculty evaluation.

Other academics have similarly advanced that, SRIs are relevant in sharing knowledge to the learners, including the origination of administrative resolutions like, for instance, the award of tenure position and advancement (Marsh, 2007; McKeachie, 2007). Moreover, a number of teachers of higher education identify that SRI is a significant tool, since the ratings received from it aid them in augmenting the supremacy of their teachings because the ratings afford educators with an understanding of their authorities and shabbiness of their instruction techniques, founded on the perception of the student (Spooren, Brockx, & Mortelmans, 2013). In total, it seems the SRI design is a significant device appropriate for evaluating faculty on teaching efficacy in advanced education.

Notwithstanding the expanding journals on the practice of student rating designs in assessing faculty on teaching effectiveness, it hardly eliminates the inconsistency regarding their reliability, validity, generalizability, and their evaluation ability of university teaching efficacy (Agbetsiafa, 2010; Beran & Rokosh, 2009; Marsh, 2007; Osler & Mansaray, 2013, 2015). Pointedly, even when some scholars have contended there is barely any indication of an association between the ratings of student and teaching efficiency (Madden, Dillion, & Lack, 2012; Pounder, 2007), others hold that SRIs are significant in assessing the teaching effectiveness of faculty (Agbetsiafa, 2010; Osler & Mansaray, 2013; Schrodt et al., 2008; Zhao and Gallant, 2012). Generally, Chen and Watkins (2010), and Zhao and Gallant (2012), for example, had earlier advanced that the reliability of SRI was about the dependability, constancy, and trustworthiness of the evaluation tool through the time. Specifically, reliability is about the internal uniformity, and steadiness of the instrument applied to evaluate teaching efficiency. Even with this deviation among the scholars concerning the reliability and validity of the SRI, several research studies realized SRIs are dependable, constant across objects, raters, and session, and valid (Anastasiadou, 2011; Beran & Rokosh, 2009; Kneipp, Kelly, Biscoe, & Richard, 2010; Osler & Mansaray, 2013, 2015). This seems to illustrate the significance for the continued use of SRIs to assess teaching effectiveness in some
universities in the US and elsewhere.

A number of research studies have utilized diverse statistical models in establishing the reliability of the SRI design, to support it continued usage to assess teaching effectiveness in higher education. Most of these examinations focused on the internal reliability of the ratings completed by the students (Anastasiadou, 2011; Beran & Rokosh, 2009; Donnon et al., 2010; Kneipp et al., 2010; Osler & Mansaray, 2013, 2015; Zuberi, Bordage, & Norman, 2007). The ubiquitous statistical reliability measurement is the Cronbach's alpha statistic, with the alpha ranging from 0 to 1, and the 1 signifying the maximum reliability score. Osler and Mansaray (2013), for example, researched the applied mathematical statistical methodology to establish the validity and reliability of independent instructional measures of teaching effectiveness at a historical black university in the US. Applying the Cronbach alpha reliability statistic on student survey responses (*N* = 7,919) from the 2012 spring semester SRI survey, Osler and Mansaray found the internal reliability of the 15 elements on the rating scale at .95, an indication of a high level of internal consistency of the ratings on the tool. Kneipp et al. (2010) had similarly

found an inter-rater dependability assessment of beyond .91 in the Cronbach's alpha statistical test, in their research on the determination of the impact of the lecturer's characteristics on the instructional excellence. Donnon et al. (2010) also recognized a superior level of internal consistency with a Cronbach's alpha coefficient estimate of .93, in their exploratory study on student ratings of lecturers in medical sciences graduate curricula.

Correspondingly, Beran and Rokosh (2009) researched the determination of students' perceptions of the value of student ratings. Beran and Rokosh (2009) developed a psychometric measurement of the value of student ratings, utilizing the survey responses from (*N* = 1229) learners at a primary Canadian university. Applying the Cronbach's alpha statistical test, Beran et al. found a superior internal reliability for the 16 elements on the rating scale at .93, thus endorsing the superiority of the internal dependability of the SRI design. Even with the relatively high internal consistency of the SRI, other academics, including Galbraith and Merrill (2012), had hardly realized any underpin for the reliability of SRI as an inclusive measurement of teaching efficiency. Thus, the reliability of the SRI design appears to have a mixed review.

In continuity, notwithstanding the contradiction among the academics regarding the reliability of the SRI, academics are also incongruity regarding the validity of the design in evaluating teaching effectiveness (Agbetsiafa, 2010; Dodeen, 2013; Donnon et al., 2010; Galbraith & Merrill, 2012; Osler & Mansaray, 2013; Safavi, Bakar, Tarmizi, & Alwi, 2012; Shevlin et al., 2000; Skowronek, Friesen, & Masonjones, 2011). Altogether, the validity of an evaluation tool (SRI) is the level upon which the instrument estimates its projected estimation (Agbetsiafa, 2010; Chen & Watkins, 2010; Leedy & Ormrod, 2015; Zhao & Gallant, 2012). Agbetsiafa (2010) also realized if SRIs were applicable to evaluate teaching competence and student-learning upshots, the instruments must be perceptible to exacting validity testing and examination. Arguably, there are different kinds of validity research studies in the current journals, including construct validity, content validity, criterion-related validity, and external validity, inter alia, even though Kane (as cited in Haladyna & Amrein-Beardsley, 2009) had argued there was an interconnected discourse to validity and that validity was a recent trend.

Aside the recent interconnected discourse on validity, there are academics who remain narrowly focused on some specific aspects of validity, principally the construct validity of SRI as a measurement of teaching efficiency (Agbetsiafa, 2010; Donnon et al., 2010; Osler & Mansaray, 2013; Skowronek et al., 2011; Sprinkle, 2008; Zhao & Gallant, 2012). Zhao and Gallant (2012) had earlier cited Cronbach and Meehl on their argument that, construct validity was the level whereupon a professed measurement reflected the dominant conjectural construct, which the scholar had programmed to evaluate. Skowronek et al. (2011) similarly noted the importance of debating issues on the SRI associated with construct validity, including responding to whether the essentiality of the student rating procedure was coherent for the evaluated construct.

Agbetsiafa (2010) was a leading proponent of the establishment of the construct validity of the rating instrument in the determination of the association between teaching effectiveness and student learning outcomes in the University of Indiana degree level course in Economics. Applying the factor analysis on 1300 sampled students, Agbetsiafa realized the Kaiser-Meyer-

Olkin (KMO) statistical estimate on the rating scale was about .91, an indication of the fitness of the factor analysis for the data. Furthermore, the Bartlett statistic for the presence of interrelation among the variables was significant at $p < .0001$. In full, the results indicated positive connections between student responsiveness of teaching efficiency, education support, communication effectiveness, clarity of the course modules, and course assessment and feedback, thereby establishing the construct validity of the SRI instrument. Donnon et al. (2010) similarly applied the factor analysis, to examine the instructional quality of faculty in medical sciences graduate programs at a Canadian university. Using a sample of 1738 enrolled graduate students in medical science courses for the period from 1999 to 2006, and the application of the principal components analysis with varimax rotation, the results realized one-factor model, which described about 58.4% of the complete variance. Concluding, Donnon et al. established the construct validity of the research with the results supporting a modest to strong associations among the 11 elements on the rating scale with high element loadings extending from .58 to .84. Similarly, Sprinkle (2008) employed the factor analysis to evaluate the association between student prejudices and their understandings of teaching efficiency, making use of student ratings. Applying 202 sampled students, the KMO statistic for the data adequacy was .79, which was conventional, including the Bartlett's test of sphericity, which was significant at $p < .0001$. The results established the construct validity of the test and research. Likewise, Osler and Mansaray (2013) applied the principal component factor analysis on 7919 sampled students, to define the construct validity of the essential configuration of the 15 elements on the rating scale. Osler and Mansaray recognized the Kaiser-Meyer-Olkin (KMO) statistical estimate on the rating scale was about .96, a signal of the fitness of the factor analysis for the data. Furthermore, the Bartlett statistic for the presence of interrelation among the variables was significant at $p < .0001$. Moreover, the factor loading for the 15 elements completely had loadings >= .82, thus confirming the construct validity of the SRI.

Scholars have similarly engaged in alternative statistical procedures, aside factor analysis, to validate the SRI design. Some of these procedures include linear regression, analysis of variance, structural equation, Pearson correlation, hierarchical regression, and the *t*-test (Beran & Rokosh, 2009; Donnon et al., 2010; Héfer, 2009; Kneipp et al., 2010; Schrodt et al., 2008; Stowell, Addison, & Smith, 2012). Héfer (2009), for example, exploited the Pearson correlation and hierarchical regression procedures, to authenticate the validity of the SRI instrument in the determination of the association among teaching efficiency, course assessment, and educational performance. Using the Pearson correlation procedure on 113 undergraduates enrolled student in six introductory classes, Héfer recognized associations among deferment of desire, motivational features of deferment of desire, predicted and definite final class grade, and ratings of the class and the teacher. The findings also recognized a positive and significant correlation ($r = .68$) between the students' ratings of the class and teaching efficiency.

El Hassan (2009), however, was interested in the substantive and consequential validity of SRIs. His research exploration answered the disquiets of substantive and consequential validity, and upheld the disquiets were completely acknowledgeable with the complete stratagem and implementation of the assessment procedures, together with an efficient

communication to learners and teachers regarding the sturdiness of the evaluation processes. Employing a descriptive statistical procedure on the SRI ratings, El Hassan recognized 70% of the respondents agreed the SRI was the normal mode to reveal proposals for enhancement, and about 50% of the respondents said faculty valued their input in generating teaching enhancement. The study also found several instructors valued the contributions from the ratings and applied them for course progression. Stowell et al. (2012), alternatively, explored the possibility of a variance in the student response rate and the validity between online and traditional student assessments of professors. Employing the *t*-statistics and correlation matrices on 2057 sampled students, Stowell et al. recognized insignificant variations in the mean ratings between the online arrangement and the traditional classroom student ratings, thereby confirming that the two evaluation designs engendered equivalent data and validity findings. Meanwhile, Skowronek et al. (2011) research investigation was more about the content validity of the rating instrument, because the tool was the focus of only a relatively small reflection on the topic of content validity in the current literature. Content validity is the level at which a tool principally gathers a detailed social construct.

Despite the far-reaching utilization of the SRI as a measurement of teaching efficiency in higher education, it continued use is abound with debates and prejudices, which seems to contest its validity on a number of borders. Donnon et al. (2010), and Smith, Cook, and Buskist (2011), for example, had argued learners who projected a higher letter grade in the class seemed to afford high ratings of teaching, and a teacher who encompassed compassionate grading system when employing intuitive testing procedures may earn superior student evaluations of teaching achievement. An appendage to this prejudice was the findings of Slocombe, Miller, and Hite (2011) who argued learners were disposed to award grander assessments to lecturers that employed humor in teaching, and to lecturers they revered. Galbraith et al. (2012), in their research on efficient teaching, equally realized negligible or no patronage for the validity of SRI as a complete signal of teaching efficiency or student learning. Given this, it is obvious the validity of the SRI remains a debatable theme among the academics.

Concluding, the continued debate concerning the complete validity of the SRI singularly as a measuring tool of teaching efficiency clearly accentuates the need to bolster the SRI design, to embrace supplementary measurements of teaching competence, principally for personnel resolutions (El Hassan, 2012). Therefore, the import of the inclusion of the GSS design to supplement the SRI model as a measurement of effective teaching for personnel resolutions is substantial for additional consistency and validity. This is especially true because the GSS design has similar rating scales as the SRI, and Berk (2005) had argued varied scale elements regarding the superiority of teaching, courses, curriculum admissions, and supplementary subject can afford novel material.

3. Methodology

The methodology for the current research exploration is the quantitative design and the data are the students' responses from the GSS survey. The methodology embraces some model specifications.

3.1. Model Specification: The leading conjecture in the current research assumes the applicability of the graduating senior survey (GSS) instrument as a supplement to the student rating of instructor (SRI) design exhibits internal consistency in the evaluation of teaching effectiveness in higher education. The model for this assumption is the Cronbach's alpha (α), after Cronbach (1951). Thus, drawing from Field (2013), the model specification for the Cronbach's alpha is the Equation (1), and is applicable in determining the reliability of the GSS instrument:

$$\frac{G^2 Covariance}{\sum s^2_{element} + \sum Cov_{element}} \quad (1)$$

The model in Equation (1) obviously stipulates a rating scale encompassing elements, and it is possible to compute the variance enclosed in an individual element, including the covariance among a specific element and any supplementary element on the rating scale. Given this, a variance-covariance matrix computation of the complete elements is a possibility. Referencing Field (2013), the transverse principles in the matrix recognize the variance enclosed in a specific element, and the off-transverse principles embrace covariances within the collections of elements. The top half of the model is the square of the number of elements (G) multiplied by the mean covariance among the elements. The bottom half of the model appears as the complete element variances and element covariances. The magnitude of the Cronbach's alpha statistic extends from 0 to 1. Field (2013) even accentuated the higher the magnitude, the exceptionality the selective elements coordinated as a body in assessing the implement construct, and thereby, the exceptionality the reliability of the evaluation implement. Consequently, a Cronbach's coefficient alpha of 1 principally suggests an impeccably dependable rating device, and a coefficient estimate of 0 suggests an undependable rating device.

The accompanying conjecture in the current research assumes the applicability of the GSS instrument as a supplement to the SRI design exhibits construct validity in the evaluation of teaching effectiveness in higher education. The model for this assumption is the factor analysis, in particular, the principal axis factoring (PAF), following similar applications by de Winter and Dodou (2012), and Ngure, Kihoro, and Waititu (2015). de Winter and Doduo (2012), for example, argued the PAF was a least-squares valuation of the communal factor model. PAF creates no supposition concerning the kind of error and reduces the unweighted sum of squares or ordinary least squares of the residual matrix. Therefore, drawing from de Winter and Doduo (2012), the PAF model specification is the Equation (2), and is applicable to determine the construct validity of the GSS instrument:

$$T_{OLS} = \frac{1}{2} tr[(G - \Sigma)^2] = \sum_k \sum_l (G_{kl} - \delta_{kl})^2, \quad (2)$$

where G_{kl} and δ_{kl} are components of the perceived sample correlation matrix, including the implicit correlation matrix, correspondingly. Ngure et al. (2015) also reinforced the realization that, the PAF model was a kind of exploratory factor analysis, which limited the variance that was communal among elements, that is, it did not reallocate the variance that was exclusive to any individual element.

International Journal of Academic Research in Business and Social Sciences
2017, Vol. 7, No. 5
ISSN: 2222-6990

4. Data

The applicable data for the current research are students' responses from the fall 2016 semester graduating senior survey (GSS) in a historically black university located in the southeastern part of the US. The GSS is a mandatory survey administered online to undergraduate senior students of the university, with several feasible sections, including academic advising, teaching effectiveness, library efficiency, and information technology, among others, on a 5-point Likert scale. The scores on the 5-point Likert scale span from 5 (very satisfied) to 1 (very dissatisfied). The research and planning department of the university holds the responsibility of administering the online survey to senior undergraduate students of the university each semester. The fall 2016 GSS survey had 426 student responses, and the section on teaching effectiveness with 9 questions on motivation, competence, and expertise, is significant to the current research in establishing the reliability and construct validity of the instrument in its determination of teaching effectiveness and student learning.

5. Results

The primary supposition in the current research, revealed in Equation (1), assumed the application of the GSS device as a supplement to the SRI design demonstrated reliability in the evaluation of teaching effectiveness on motivation, competence, and expertise overall. The analysis included and satisfied a preliminary examination of the assumption of normality, linearity, and moderate level of correlation among the variables. Equation (1) applied the Cronbach's alpha test statistic, to resolve the issue of the reliability of the GSS instrument in assessing the three areas of teaching efficiency in totality. The findings of the test embraced an introductory element descriptive statistics (e.g. mean, standard deviation), to reveal any conceivable difference among the 9 elements on the rating scale. Table 1 is a summary of the descriptive element statistics of the variables applied in the current research, and it is self-explanatory. In Table 1, there was hardly any disparity and spreading between the mean for element A1 (*Their ability to motivate me to do my best*) (M = 4.31, SD = .773) and the mean for element A2 (*How carefully they explain the expectations of student performance in the course* (M = 4.26, SD = .802) on the rating scale, for example. Because of the absence of any infinite difference among the elements on the scale, the applicable unstandardized Cronbach's alpha in Table 2 was significant in the interpretation of the reliability coefficient of the GSS instrument, following a similar argument by Leech, Barrett, and Morgan (2014). Therefore, in Table 2, the alpha for the 9 elements on the rating scale, which together assessed teaching effectiveness at the university, was .95, an indication the elements established a scale that had practical internal stability reliability for the GSS instrument. In addition, Table 3 shows a summary of element total statistics, which were the reliability findings for the individual

Table 1
Summary of Element Statistics

Element	Mean	SD	N
A1	4.31	.773	426
A2	4.26	.802	426
A3	4.04	.922	426
A4	4.22	.789	426
A5	4.31	.762	426
A6	4.21	.778	426
A7	3.95	.950	426
A8	4.13	.847	426
A9	4.29	.822	426

Note: A1 = Their ability to motivate me to do my best; A2 = How carefully they explain the expectations of student performance in the course; A3 = The extent to which they consider different learning styles; A4 = How well they explain course material; A5 = The extent to which they encourage class discussion; A6 = How effectively they use instructional technology in teaching and learning activities; A7 = How quickly they provide feedback on my work; A8 = The helpfulness of their feedback on my work; A9 = Overall satisfaction with instructors in your major.

Table 2
Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	*N* of Items
.945	.946	9

elements on the GSS scale. The two supreme important sectors in Table 3 were the corrected-element-total- correlation, and the alpha-if-element-deleted. The previous sector in the table was the connection of the individual explicit element with the entire aggregate of the outstanding elements on the GSS scale, and the association must be >= .40 to be deemed a decent module for this summated assessment scale (Leech et al., 2014). The findings in Table 3 indicated all the elements on the scale had significant inter-relationship with each other of >=.75. Additionally, the latter sector was the alpha coefficient estimate for the individual element on the GSS scale, which showed the elements were dependable and exceedingly inter-related and collectively, they generated excellent internal constancy reliability. Consequently, the results in Tables 2 and 3 for all 9 elements on the GSS scale collectively specified the students presented progressive evaluations of the instructional quality they received at the university, and the elements engendered a scale that held convincing internal uniformity reliability. Given this, the GSS instrument exhibited strong internal reliability.

International Journal of Academic Research in Business and Social Sciences
2017, Vol. 7, No. 5
ISSN: 2222-6990

In continuity, the accompanying conjecture in the current research, revealed in Equation (2), assumed the applicable GSS instrument demonstrated construct validity in the evaluation of teaching effectiveness on motivation, competence, and expertise at the university. The

Table 3
Summary Element-Total Statistics

	Scale Mean if Element Deleted	Scale Variance if Element Deleted	Corrected Element-Total Correlation	Squared Multiple Correlation	Cronbach's Alpha if Element Deleted
A1	33.40	31.299	.802	.696	.938
A2	33.46	31.162	.784	.685	.939
A3	33.68	29.961	.794	.651	.939
A4	33.50	30.994	.821	.703	.937
A5	33.41	31.809	.750	.608	.941
A6	33.50	31.474	.773	.633	.940
B4	33.76	30.002	.761	.646	.941
A7	33.59	30.384	.827	.732	.937
A8	33.43	30.928	.791	.640	.939

principal axis factoring (PAF) with varimax rotation was applicable in Equation (2), which measured the essential configuration for the 9 elements on the GSS rating scale on teaching effectiveness. The application of three factors on the PAF analysis was significant because of the realization the intention of the elements on the GSS design was to table three constructs of teaching effectiveness, which were motivation, competence, and expertise. The results of the PAF test were revealed in Tables 4, 5 and 6.

Table 4
Correlation Matrix

		A1	A2	A3	A4	A5	A6	A7	A8	A9
Correlation	A1	-	.768	.630	.744	.662	.635	.614	.650	.691
	A2	.768	-	.623	.754	.590	.623	.597	.647	.687
	A3	.630	.623	-	.664	.643	.680	.684	.741	.656
	A4	.744	.754	.664	-	.663	.671	.623	.680	.709
	A5	.662	.590	.643	.663	-	.710	.573	.621	.611
	A6	.635	.623	.680	.671	.710	-	.606	.672	.610
	A7	.614	.597	.684	.623	.573	.606	-	.779	.639
	A8	.650	.647	.741	.680	.621	.672	.779	-	.708
	A9	.691	.687	.656	.709	.611	.610	.639	.708	-

Note: Determinant = .001

International Journal of Academic Research in Business and Social Sciences
2017, Vol. 7, No. 5
ISSN: 2222-6990

Table 4 shows the estimated correlation and the level of significance of the 9 elements on the GSS scale on teaching effectiveness. All the 9 elements were significant (p = .001) and in correlation with one another, thus suggestion they may establish one or more factors. Table 5

Table 5
KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.944
Bartlett's Test of Sphericity	Approx. Chi-Square	3131.332
	df	36
	Sig.	.000

shows the Kaiser-Meyer-Olkin (KMO) statistic, which was an assessment of the sampling aptness, that is, whether there were sufficient elements projected by individual factors. The KMO statistical estimation ranges from 0 to 1, and a value close on 1 is a signal of the alignment of associations that is practically compacted and, consequently, the factor analysis be supposed to generate an exclusive and trustworthy factor (Agbetsiafa, 2010). Leech et al. (2014) had even suggested the KMO statistic must be > .70 to endorse acceptable elements for individual factors. In Table 5, therefore, the KMO statistic for the 9 elements on the GSS scale was .944, thus sanctioning the sampling suitability for the KMO analysis. Table 5 also displays the Barlett's test of sphericity, which is a measurement of the *null hypothesis* that, the primary correlation matrix appears to be an identity matrix. Leech et al. (2014) had noted the Barlett's test of sphericity must be significant (p < .05) to endorse that the correlation matrix significantly differed from an identity matrix, where associations among variables were exclusively 0. Therefore, in Table 5, the Barlett's test of sphericity, χ^2 (36) = 3131.332, p < .0001 was

Table 6
Total Variance Explained

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.306	70.1	70.1	2.5	27.7	27.7
2	.594	6.6	76.7	2.4	26.2	53.9
3	.519	5.8	82.4	1.8	20.0	73.9
4	.324	3.6	86.0			
5	.313	3.5	89.5			
6	.283	3.1	92.7			
7	.248	2.8	95.4			
8	.212	2.4	97.8			
9	.201	2.2	100.0			

Note: Extraction Method: Principal Axis Factoring.

significant, suggesting the associations among the 9 elements on the GSS instrument were sufficiently robust for PAF analysis.

Table 6 is the results of the clarification of the total variance of the 9 elements on the rating scale. A preliminary application of three factors on the PAF model was significant, grounded on the recognition the design of the elements on the GSS instrument was to index the three constructs on teaching effectiveness, which were motivation, competence, and expertise. Following the varimax rotation, the first factor accounted for about 27.7% of the

Table 7
Rotated Factor Matrix

	Factor Loading		
	1	2	3
How carefully they explain the expectations of student performance in the course	.76		
The ability to motivate me to do my best	.70		
How well they explain course material	.65		.42
Overall satisfaction with instructors in your major	.56	.50	
The helpfulness of their feedback on my work		.78	
How quickly they provide feedback on my work		.70	
The extent to which they consider different learning styles		.59	.46
The extent to which they encourage class discussion			.68
How effectively they use instructional technology in teaching and learning activities		.41	.63

Note: Removal of Loadings <.40

variance; the second factor accounted for about 26.2% of the variance; and, the third factor accounted for about 20% of the variance. Table 7 displays the 9 elements and factor loadings for the rotated factors, with loadings <.40 removed for clearness. The first factor, which appeared to index motivation, had strong positive loadings on the first three elements of >= .65. The fourth element indexed motivation had a positive loading of .56 compared to the four elements in the motivation factor. *How carefully they (instructors) explain the expectations of student in the course*, which could be an essential ingredient in motivating students to excel in a course, had the highest loading (.76) on the first (motivation) factor. *How well they (instructors) explain course material* had a loading of > .60 on the first factor (motivation), but had a cross-loading of > .40 on the third factor (expertise). The second factor, which appeared to catalog competence in the assessment of teaching effectiveness, had high positive loadings on the ensuing four elements of > .40 in Table 7. *The helpfulness of their (instructors) on my work*, and *How quickly they (teachers) provide feedback on my work*, exemplified the strongest positive loadings of >=.70 on the competence factor. However, *The*

overall satisfaction with instructors in your major had a loading of .50 on the competence factor, but had a cross-loading of > .50 on the motivation factor. In Table 7, the third factor, which appeared to index expertise on teaching effectiveness, had strong positive loadings of >=.63 on two of the three elements in the group. The third element, *The extent to which they (instructors) consider different learning styles*, had a loading of .46 on the expertise factor, but had a cross-loading of .59 on the competence factor. In total, the three factors of motivation, competence, and expertise together had strong loadings of >.40, and interconnected significantly with teaching effectiveness, thereby endorsing the construct validity of the GSS instrument applied in assessing teaching efficacy.

5. Analysis of Results

The focus of the current research is the determination of the reliability and validity of the GSS instrument as a supplementary assessment tool of teaching quality in a historically black university in the southeastern US. Applying the Cronbach's alpha test statistic to resolve Equation (1), which examines the reliability of the GSS instrument, the findings in Table 2 unveils a predominantly high internal homogeneousness of the students' responses to the 9 elements on the GSS rating scale. Consequently, the result endorses the reliability of the instrument as a possible measure of teaching efficacy at the university. The findings are analogous to some studies on the reliability of the SRI, a parallel measuring design on teaching effectiveness with the utilization of the Cronbach's alpha test statistic (Agbetsiafa, 2010; Anastasiadou, 2011; Beran & Rokosh, 2009; Donnon et al., 2010; Kneipp et al., 2010; Osler & Mansaray, 2013, 2015). In addition, the inter-total statistics in Table 3, which is a subsection of the reliability test statistic, indicates all 9 elements on the GSS scale collectively stipulates that, the students present progressive valuations of the instructional efficacy they received at the university, and the elements engendered a scale that has considerable internal evenness reliability. Consequently, the GSS instrument shows strong internal reliability. The results are correspondingly analogous to the findings of Osler and Mansaray (2013, 2015) on the reliability of the SRI, a parallel assessment instrument of faculty on teaching effectiveness.

The subsequent analysis is the determination of the validity of the GSS design as a measure of teaching effectiveness at the university, which encompasses Equation (2) of the current research. Applying the factor analysis, particularly the principal axis factoring (PAF) with varimax rotation, to examine the construct validity, the findings discloses the Kaiser-Meyer-Olkin (KMO) statistic on the GSS rating scale in Table 5 is about .94, soundly beyond the >.70 suggested by Leech et al. (2014) in validating the sampling suitability for the PAF analysis. The KMO result along with the statistically significant Barlett's sphericity in Table 5 confirms that the 9 elements on the GSS scale are sufficiently robust for the PAF analysis. The findings are also equivalent to the KMO and Barlett's sphericity results of the SRI measuring instrument of teaching effectiveness with the application of factor analysis in similar research explorations (Agbetsiafa, 2010; Donnon et al., 2010; Osler & Mansaray, 2013, 2015; Sprinkle, 2008). In addition, the application of three factors on the PAF analysis appears significant because of the recognition that, the objective of the 9 elements on the GSS design is to table three constructs of teaching effectiveness, which are motivation, competence, and expertise. Following the

International Journal of Academic Research in Business and Social Sciences
2017, Vol. 7, No. 5
ISSN: 2222-6990

varimax rotation, the findings in Table 6 suggests the three factors on teaching effectiveness together accounts for about 73.9% of the total variance of the 9 elements on the GSS rating scale. Furthermore, the factor loading for the 9 elements display in Table 7 completely have loadings > .40, and are significantly associated with teaching effectiveness,
thereby sanctioning the construct validity of the GSS instrument as a supplementary assessment tool of faculty on teaching quality at the university. The construct validity findings are consistent with the findings of other academics (Agbetsiafa, 2010; Safavi et al., 2012; Osler & Mansaray, 2013), with the application of the factor analysis in assessing faculty on teaching efficiency. In sum, the study confirms a robust interconnection between the ratings of students on the GSS scale and the three modules of teaching efficiency, including motivation, competence, and expertise.

6. Conclusion

The objective of the current research was the determination of the validity and reliability of the GSS design as a supplementary instrument in assessing faculty on teaching effectiveness at a historically black university in the southeastern US. A preliminary descriptive element statistics of the variables applied in the current research was noteworthy, to reveal any possible disparity and dispersion among the elements of concentration. In addition, the application of the Cronbach's alpha test was also significant because it helped resolve the issue of the reliability of the GSS assessment design as specified in Equation (1). The findings on the Cronbach's alpha reliability statistic showed the alpha for the 9 elements on the GSS, which together assessed teaching effectiveness at the university, was .95 (superior), thus confirming the high reliability of the GSS design for effective teaching assessment. This is in addition to the recognition that, the reliability findings for the individual elements on the GSS scale also showed the 9 elements were dependable and exceptionally interconnected and, jointly, they engendered exceptional internal constancy reliability.

The accompany conjecture in the current research, revealed in Equation (2), was the determination of the validity of the GSS design as a supplementary instrument, to assess faculty on effective teaching at the university. The applicable principal axis factoring (PAF) with varimax rotation model had three factors, which were significant because the objective of the 9 elements on the GSS design was to table three constructs of teaching quality, including motivation, competence, and expertise. The PAF findings indicated the 9 elements on the rating scale had sampling suitability, and that the motivation, competence, and expertise factors of teaching efficiency jointly accounted for about 73.9% of the total variance. Furthermore, the PAF findings also showed the three factors of motivation, competence, and expertise altogether had strong loadings of > .40, and significantly interrelated with teaching effectiveness, thereby endorsing the construct validity of the GSS instrument as a supplementary assessment tool of teaching efficacy at the university.

Given all this, the GSS design is proven to be a valid and reliable instrument on the specific elements of teaching effectiveness relating to motivation, competence, and expertise, and can be applied as a supplement to the SRI instrument currently available for assessing teaching quality at the university. The findings are significant because of the ongoing conflict

among the academics to agree on the validity and reliability of the widely applied SRI as a measuring tool of teaching quality in higher education. El Hassan (2012), and Marsh et al. (2011) had earlier embraced the use of supplementary measurements of assessing teaching efficacy because of the variance among the academics on the use of the SRI alone to measure faculty on teaching effectiveness. Thus, the GSS will serve as a supplement to the SRI instrument in establishing the validity and reliability of the ratings of students concerning teaching quality at the university. Therefore, educational policy makers, and the faculty at the university should engender policies to streamline the elements on the GSS instrument that are specific to teaching effectiveness as a supplementary tool of assessing faculty on instructional quality.

Even with this policy endorsement, the research is not without its confines. In particular, the research was limited to one semester analysis of the GSS instrument, which may be inadequate to establish the generalizability of the instrument as a measurement of teaching efficiency. In addition, the current research also fell short because the exploration of the construct validity of the GSS design was all-purpose, and not tied to a specific course or instructor. In such, the students ratings on the GSS were universal, and may change once indexed to assess a course or an instructor. Still, the research exploration was robust and decisive. It facilitated the closure of the imperfect information in the current literature concerning the validity and reliability of the graduating senior exit survey and its possible use to assess faculty on teaching effectiveness and student learning outcomes. The results also held tangible extrapolations for educational policy makers, administrations, faculty, and students, along with adding a novel knowledge to the apparently boundless journals on the assessment of faculty on teaching effectiveness in higher education. In conclusion, the results realized will add a new knowledge of assessing faculty on teaching effectiveness in higher education with enhanced validity and reliability.

References

Agbetsiafa, D. (2010). Evaluating effective teaching in college level economics using student ratings of instruction: A factor analytic approach. *Journal of College Teaching & Learning, 7*(5), 57-66. Retrieved from https://www.cluteinstitute.com/journals/journal-of-college-teaching-learning-tlc/

Apodaca, P., & Grad, H. (2005). The dimensionality of student ratings of teaching: Integration of uni-and multidimensional models. *Studies in Higher Education, 30*(6), 723-748. doi: 10.1080/03075070500340101

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

Barry, R. A. (2010*). Teaching effectiveness and why it matters.* Marylhurst, OR: Marylhurst University and the Chalkboard Project. Retrieved from https://chalkboardproject.org/sites/default/files/teacher-effectiveness-and-why-it-matters.pdf

Beran, T., & Rokosh, J. (2009). Instructors' perspectives on the utility of student ratings of instruction. *Instructional Science*, *37*(2), 171-184. doi:10.1007/s11251-007-9045-2

Beran T., Violato C. (2005). Ratings of university teacher instruction: How much do student and course characteristics really matter? *Assessment and Evaluation in Higher Education, 30*, 593–601. doi:10.1080/02602930500260688

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

Bolhuis, S. (2003). Towards process-oriented teaching for self-directed lifelong learning: A multidimensional perspective. *Learning and Instruction, 13*(3), 327–347. Retrieved from https://www.journals.elsevier.com/learning-and-instruction/

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

Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation., 28(1), 71-88. doi: 10.1080/0260293032000033071

Clayson, D. E. (2009). Student evaluations of teaching: Are they related to what students learn?. *Journal of Marketing Education*, *31*(1), 16-30. Retrieved from http://journals.sagepub.com/home/jmd

Cohen, P. A. (1981) Student ratings of instruction and student achievement: A meta-analysis of multisection validity studies. *Review of Educational Research, 51*, 281–309. Retrieved from http://journals.sagepub.com/home/rre

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297-334. Retrieved from https://www.psychometricsociety.org/content/psychometrika

Daugherty, K. J., & Natow, R. S. (2015). *The politics of performance funding for higher education: Origins, discontinuations, and transformations*. Baltimore, MD: JHU Press.

de Winter, J. C. F., & Doduo, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics, 39*,(4), 695-710. Retrieved from http://www.tandfonline.com/toc/cjas20/current

Dodeen, H. (2013). Validity, reliability, and potential bias of short forms of students' evaluation of teaching: The case of UAE University. *Journal Educational Assessment, 18*(4), 235-250. doi: 10.1080/10627197.2013.846670

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

El Hassan, K. (2009). Investigating substantive and consequential validity of student ratings of instruction. *Higher Education Research & Development*, *28*(3), 319-333. doi:10.1080/07294360902839917

Ellis, R., & Bowden, R. (2014). Performance based funding: Changing the paradigm for higher education. *British Journal of Education, Society & Behavioral Sciences, 4*(7), 942-952. Retrieved from http://www.sciencedomain.org/journal/21

Entwistle, N. J. (1987*). Understanding classroom learning*. London, UK: Hodder and Stoughton.

Feldman, K. A. (1989) Instructional effectiveness of college teachers themselves, current and former students, colleagues, administrators, and external (neutral) observers. *Research in Higher Education, 30*, 137–193. Retrieved from https://link.springer.com/journal/11162

Field, A. (2013). *Discovering statistics using SPSS* (3rd ed.). Thousand Oaks, CA: SAGE Publications Inc.

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Syntheses of educational productivity research. *International Journal of Educational Research, 11*, 145–252. Retrieved from https://www.journals.elsevier.com/international-journal-of-educational-research

Galbraith, C. S., & 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, 60*(2), 48-55. doi: 10.1080/87567555.2011.627896

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

Hassel, B. C. (2009). *How should state define teacher effectiveness*? Washington, DC: Center for American Progress. Retrieved from https://www.americanprogress.org/

Hatfield, C. L., & Coyle, E. A. (2013). Factors That Influence Student Completion of Course and Faculty Evaluations. *American Journal of Pharmaceutical Education*, *77*(2), 27. http://doi.org/10.5688/ajpe77227

Heckert, T. M., Latier, A., Ringwald-Burton, A., & Drazen, C. (2006). Relations among student effort, perceived class difficulty appropriateness, and student evaluations of teaching: Is it possible to "Buy" better evaluations through lenient grading? *College Student Journal, 40*(3), 588. Retrieved from http://www.projectinnovation.com/college-student-journal.html

Héfer, B. (2009). Teaching effectiveness, course evaluation, and academic performance: The role of academic delay of gratification. *Journal of Advanced Academics, 20*(2), 326-355. Retrieved from http://journals.sagepub.com/home/joa

Hillman, N. W., Tandberg, D. A., & Fryar, A. H. (2015). Evaluating the impacts of "New" performance funding in higher education. *Educational Evaluation and Policy Analysis, 37*(4), 501-519. doi: 10.3102/0162373714560224

Jones, B. (2010). An examination of motivation model components in face-to-face and online instruction. *Electronic Journal of Research in Educational Psychology, 8*(3), 915-944. Retrieved from http://investigacion-psicopedagogica.org/revista/new/english/index.php

Keeley, J. W., English, T., Irons, J., & Henslee, A. M. (2013). Investigating halo and ceiling effects in student evaluations of instruction. *Educational and Psychological Measurement, 73*(3), 440-457. Retrieved from http://journals.sagepub.com/home/epm

Kneipp, L. B., Kelly, K. E., Biscoe, J. D., & Richard, B. (2010). The impact of instructor's personality characteristics on quality of instruction. *College Student Journal, 44*(4), 901-905. Retrieved from http://www.projectinnovation.com/college-student-journal.html

Leech, N. A., Barrett, K. C., Morgan, G. A. (2014). *IBM: SPSS for intermediate statistics use and interpretation* (5th ed.). New York, NY: Routledge.

Leedy, P. D., & Ormrod, J. E. (2015). *Practical research: Planning and design* (11th ed.). Upper Saddle River, NJ: Pearson Education Inc.

Liu, O. L. (2012). Student evaluation of instruction: In the new paradigm of Distance Education. *Research in Higher Education, 53*(4), 471–486. doi: 10.1007/s11162-011-9236-1

Madden, T. J., Dillon, W. R., & Leak, R. L. (2010). Students' evaluation of teaching: Concerns of item diagnosticity. *Journal of Marketing Education*, *32*(3), 264-274. doi:10.1177/0273475310377759

Marsh, H. W. (2007). Students' evaluations of university teaching: Dimensionality, reliability, validity, potential biases, and usefulness. In R. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 319-383). Netherlands: Springer.

Marsh, H. W., Ginns, P., Morin, A. S., Nagengast, B., & 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*, *103*(3), 733-748. doi:10.1037/a0024221

McKeachie, W. J. (1979). Student ratings of faculty: A reprise. *Academe, 65*, 384-397. Retrieved from https://www.aaup.org/academe

McKeachie, W. J. (2007). Good teaching makes a difference-And we know what it is. In R. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education. An evidence-based perspective* (pp. 457-474). Netherlands: Springer. doi: 10.1007/1-4020-5742-3_11

Miao, K. (2012*). Performance-based funding of higher education: A detailed look at best practices in 6 states*. Washington, DC: Center for American Progress. Retrieved from

https://www.americanprogress.org/issues/education/reports/2012/08/07/12036/performance-based-funding-of-higher-education/

Mitchell, M., & Leachman, M. (2015). *Years of cuts threaten to put college out of reach for more students*. Washington, DC: The Center on Budget and Policy Priorities. Retrieved from http://www.cbpp.org/research/state-budget-and-tax/years-of-cuts-threaten-to-put-college-out-of-reach-for-more-students

Mittal, S., & Gera, R. (2013). Student evaluation of teaching effectiveness (SET) in higher education in India. *International Journal of Business and Social Science, 4*(10), 289-298. Retrieved from http://www.ijbssnet.com/

Ngure, J. N, & Kihoro, J. M., & Waititu (2015). Principal component and principal axis factoring of factors associate with high population in urban areas: A case study of Juja and Thika, Kenya. *American Journal of Theoretical and Applied Statistics, 4*,(4), 258-263. Retrieved from http://www.sciencepublishinggroup.com/journal /archive.aspx?journalid=146&issueid=-1

O'Shaughnessy, L. (2013, January 16). 6 Challenges facing state universities. *CBSNews*. Retrieved from http://www.cbsnews.com/news/6-challenges-facing-state-universities/

Osler, J. E., & Mansaray, M. A. (2013). Applied mathematical statistical modeling to determine the validity and reliability of independent instructional measures. *Journal on Mathematics, 2(*3), 12-31. Retrieved from http://www.imanagerpublications.com/

Osler, J. E., & Mansaray M. A. (2015). Educational technology assessment: A model for analyzing online psychometric tests for course evaluations. In S. J. Keengwe (Eds.), *Handbook of research on educational technology integration and active learning* (pp. 215-246). Hershey, PA: IGI.

Pounder, J. (2007). Is student evaluation of teaching worthwhile? An analytical framework for answering the question. *Quality Assurance in Education, 18*(1), 47-63. Retrieved from http://www.emeraldinsight.com/loi/qae

Rutherford, A., & Rabovsky, T. (2014). Evaluating impacts of performance funding policies on student outcomes in higher education. *The Annals of the American Academy of Political and Social Science, 655*, 185-208. doi: 10.1177/0002716214541048

Sanford, T., & Hunter, J. M. (2011). Impact of performance-funding on retention and graduation rates. *Education Policy Analysis Archives, 19*(33), 1-30. Retrieved from http://epaa.asu.edu/ojs/

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

Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford, UK: Pergamon.

Schrodt, P., Witt, P. L., Myers, S. A., Turman, P. D., Barton, M. H., & Jernberg, K. A. (2008). *Communication Education, 57*(2), 180-200. doi:10.1080/03634520701840303

Seidel, T., & Shavelson, R. J. (2007). Teaching effectiveness research in the past decade: The role of theory and research design in disentangling meta-analysis results. Review of Educational Research, 77(4), 454-499. doi: 10.3102/0034654307310317

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

Shevlin, M., Banyard, P., Davies, M., & Griffiths, M. (2000). The validity of student evaluation of teaching in higher education: Love me, love my lectures*?. Assessment & Evaluation in Higher Education, 25*(4), 397-405. doi: 10.1080/713611436

Shin, J. C. (2010). Impacts of performance-based accountability on institutional performance in the U.S. *Higher Education, 60*, 47-68. Retrieved from https://link.springer.com/journal/10734

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

Slocombe, T., Miller, D., & Hite, N. (2011). A survey of student perspectives toward faculty evaluations. *American Journal of Business Education*, 4(7), 51-58. Retrieved from https://www.cluteinstitute.com/journals/american-journal-of-business-education-ajbe/

Smith, D.L., Cook, P., & Buskist, W. (2011). An experimental analysis of the relation between assigned grades and instructor evaluations*. Teaching of Psychology, 38* (4), 225-228. Retrieved from http://journals.sagepub.com/home/top

Spooren, P., Brockx, B., & Mortelmans, (2013) On the validity of student evaluation of teaching: The state of the art. Review of Educational Research, 83(4), 598-642. Retrieved from http://journals.sagepub.com/doi/abs/10.3102/0034654313496870

Sprinkle, J. E. (2008). Student perceptions of effectiveness: An examination of the influence of student biases. *College Student Journal, 42*(2), 276-293. Retrieved from http://www.projectinnovation.com/college-student-journal.html

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

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

Umbricht, M. R., Fernandez, F., & Ortagus, J. C. (2015). An examination of the (Un)intended consequences of performance funding in higher education. *Educational Policy,* 1-31. doi: 10.1177/0895904815614398

Vroom, V. C. (1964). *Work and motivation* (1st ed.). San Franscisco, CA: Jossey-Bass.

Zhao, J., & Gallant, D. J. (2012). Student evaluation of instruction in higher education: exploring issues of validity and reliability. *Assessment & Evaluation In Higher Education*, *37*(2), 227-235. doi:10.1080/02602938.2010

Zuberi, R. W., Bordage, G., & Norman, G. R. (2007). Validation of the SETOC instrument – student evaluation of teaching in outpatient clinics. *Advances in Health Sciences Education*, *12*(1), 55-69. doi:10.1007/s10459-005-2328-y