

THE CONCURRENT AND PREDICTIVE VALIDITY OF COVENANT UNIVERSITY STUDENTS' EVALUATION OF LECTURERS' TEACHING COMPETENCE

J. Odukoya, A. Atayero, A. Alao, A. Afolabi

Covenant University (NIGERIA)

Abstract

The evaluation of a lecturer by a student is an essential step in determining the success of the learning process. This often requires the development and standardization of data gathering instruments. The development and standardization of a data gathering instrument is hardly complete without the establishment of its reliability and validity. Of more importance is the validity index. An instrument can be reliable without being valid, but hardly can the instrument be valid without being reliable. Validity, therefore, subsumes the concept of reliability. The practical implication of this submission is that validity is an imperative requirement in the standardization and hence utilization of all data gathering instruments. Using an un-validated instrument could be disastrous. It is against this background it was deemed necessary to ascertain the validity of the recently developed Covenant University Students' evaluation of Lecturers' teaching competence instrument tagged Lecturers' Teaching Competence Evaluation Form – Students' Version [LTCEF-SV]. To achieve this objective the 7083 students and 21 Heads of Departments of Covenant University participated in the study. After a close scrutiny of recent submission in the literature on the indicators of effective teaching at the secondary and tertiary levels, the LTCEF-SV was developed using the Participatory Research Approach [PAR]. Consequently, Students, Lecturers and Management were actively involved in the process of developing the LTCEF-SV. This step, together with expert's review of the instrument served to establish the content validity of the instrument. Nonetheless, it was deemed imperative that the Criterion-related validities [i.e. concurrent and predictive] be established. For this study, the Concurrent validity was established with correlation of scores from student's evaluation and Heads of departments' evaluation of the same set of lecturers around the same period while the Predictive validity was established with correlation of scores from students' evaluation of lecturers and students' performance score in the 2013/14 Alpha semester's examination. The results mostly showed weak and insignificant Concurrent and Predictive validities. The findings were discussed while relevant recommendations were made.

Keywords: Criterion-related validity, Concurrent validity, Predictive validity, Evaluation, Instruments and Standardization.

1 INTRODUCTION & LITERATURE REVIEW

Utilization of a psychological test without due validation could be catastrophic, especially if the outcome of assessment is targeted at making sensitive decisions that border on the recipients' life and destiny. Students' Evaluations of Lecturers tend to fall within this domain. The outcome of such evaluation is often used to judge the value, relevance and retention of Lecturers. This in turn does not only affect the Lecturer's overall wellbeing, but that of his family as well. Consequently, it is imperative that the validity of such psychological instruments be empirically ascertained before formal usage. This, therefore, is the *justification* and *significance of this study*.

There are many *types of validity* indices. What informs choice of validity for a particular psychological instrument is the purpose and nature of the test. For instance, for many questionnaires seeking people's opinion, basic *face* and *content* validity might just be enough. These have been established for the Covenant University Lecturers' Teaching Competence Evaluation Form – Students' Version [LTCEF-SV]. However, for a summative achievement test like that of the West African Examinations Council or a University Semester Examination, the validities cited above may not be enough. When the purpose of an assessment is sensitive, there is need to step up to more robust validity indices like *Construct*, *Concurrent* and *Predictive* validities. Concurrent and Predictive validities are *Criterion-related validities*. It is against this *background* it was deemed necessary to ascertain the Concurrent and Predictive validity of the LTCEF-SV.

As described in the *Standards for Educational and Psychological Tests* (APA, 1974), 'Questions of validity are questions of what may properly be inferred from a test or instrument's score; Validity refers to appropriateness of inferences that test/instrument's scores'. According to Cherry (2014), 'Validity is the extent to which a test measures what it claims to measure. It is vital for a test to be valid in order for the results to be accurately applied and interpreted'. Validity cannot be ascertained by a single statistic, but by a body of research that demonstrates the relationship between the instrument and the behavior it is intended to measure. There are three basic types of validity: Content, Construct and Criterion. When a test has *Content* validity, the items on the test represent the entire range of possible items the test should cover and should be appropriately structured to elicit the desired responses. An instrument has *construct* validity if it demonstrates an association between its scores and the prediction of a theoretical trait or construct, such as intelligence. An instrument is said to have *criterion-related* validity when it has demonstrated its effectiveness in predicting its envisaged criterion.

Basically, there are two types of *Criterion-related* validities – *Concurrent* and *Predictive* Validities. *Concurrent validity* refers to the degree to which a measure is correlated with other measures that it is theoretically predicted to correlate with (Dardouri, Gharbi, & Selmi, 2014). *Predictive validity* tests the degree to which the test score predict the expected outcomes (Whetzel, Rotenberry, Paul, McDaniel, 2014). The focus of this study is on Criterion-related validities.

Criterion Validity evidence involves the correlation between the test and criterion variable(s) taken as representative of the construct. In other words, it compares the test with other measures or outcomes (the criteria) already held to be valid. For example, employee selection tests are often validated against measures of job performance (the criterion), and IQ tests are often validated against measures of academic performance (the criterion). Criterion-related validity looks at the relationship between a test score and an outcome. (Amatachaya, Naewla, Srisim et al, 2014). For example, SAT scores are used to predict whether a student will be successful in college. First-year grade point average becomes the criterion for success. Looking at the relationship between test scores and the criterion can tell you how valid the test is for determining success in college.

Bachman & Palmer (1996) and Sawaki & Nissan (2002) noted that an investigation of the criterion validity of an instrument can be conceptualized as a *predictive* validity study, where the focus is on investigating the extent to which the given assessment predicts candidates' future performance in the targeted criterion; or a *concurrent* validity study, where the focus is on investigating the degree to which a given assessment serves as an indicator of candidates' performance on a criterion measure collected at the same time. Schneider & Schmitt (1992) and Horey, Harvey, Curtin, Keller-Glaze, Morath & Fallesen (2007) affirmed that Criterion-related validity applies when a relationship is hypothesized to exist between an instrument's scores and performance on related criterion measure.

Theoretically, it is expected that when a lecturer is rated high on teaching competence by majority of the students he or she taught, then majority of such students should also score high in his or her course. In essence, students who rate a Lecturer high should also score high in a course handled by the Lecturer and vice versa. Thus, the students' evaluation score is expected to predict their performance in the course handled by the evaluated Lecturer. This is the logic behind *Predictive* validity. It is also expected that students' evaluation score should strongly correlate with evaluation score by a more mature and objective personality like Head of Department [HOD], using the same instrument. This is the premise for *Concurrent* validity. These points were reiterated by Nishiyama, Mizuno, and Kojima (2014).

Higgins (2014) submitted that the key to validation using predictive validity is the use of "correlation coefficient" where each examinee's assessment score is correlated with a criterion score. If the correlation coefficient equals or exceeds $r=0.20$, it means the assessment score is sufficiently related to criterion score to make judgments about a candidates likelihood of success criterion task. A correlation of $r=0.20$ is the minimum that should be considered acceptable. The operational definitions for this study were partly derived against this *theoretical framework*.

Covenant University (CU)

Covenant University (CU), a dynamic vision driven institution was founded on Christian ethos and is committed to achieving excellence in the academia. CU is driven by the compelling vision of *raising a new generation of leaders for the African Continent* via human resource development and integrated learning curriculum. CU is located at kilometre 10 along Idi-iroko way in Ota, Ogun state, Nigeria. Presently, CU operates the collegiate system. There are two colleges – the College of Development Studies (CDS) and the College of Science and Technology (CST). The CDS is made of the School of Social Sciences (SSS), School of Business Studies (SBS) and School of Human Resources Development (SHRD). The CST is made up of School of Engineering, School of Environmental Studies (SES), and School and Natural and Applied Sciences (SNAS).

The core **problem** that prompted this study is the apparent danger of using un-standardized life sensitive psychological instruments.

The **core objectives** of this study are: to *ascertain the Concurrent validity of the CU-LTCEF-SV* and to *ascertain the Predictive validity of the CU-LTCEF-SV*. The **research questions** raised for this study are as follows: 'What is the Concurrent validity of the CU-LTCEF-SV?' and 'What is the Predictive validity of the CU-LTCEF-SV?'

This **study is significant** because it is targeted at enhancing the standardization of the CU-LTCEF-SV, and so make the instrument a more reliable tool for making sensitive decisions on CU lecturers. Ultimately, the instrument is apt to significantly enhance the overall quality of learning at Covenant University [CU].

Decision Rules/Operational Definitions

- *Validity Indicator 1* - Only positive correlation coefficients shall be accepted as valid indicators of the validity being measured.
- *Validity Indicator 2* - Correlation coefficients below 0.20 shall not be accepted as significant enough to report as indices of validity.
- *Low Validity* - Correlation coefficients ranging from 2.0 to 3.9 [for df = 50 and $p < 0.05$] are regarded as low validity indices in this study.
- *Moderately or Fairly High Validity* – Correlation coefficients from 4.0 to 5.9 [for df = 50 and $p < 0.05$] are regarded as fairly high validity indices in this study.
- *High Validity* – Correlation coefficients from 6.0 to 1.0 [for df = 50 and $p < 0.05$] are regarded as high validity indices in this study.

2 METHODOLOGY

The ex-post facto and survey **research designs** were used in this study. Existing and fresh data were used in this study.

The **populations** for this study were all the students, academic staff and Heads of Department in Covenant University. On the overall, the population was estimated to be approximately 8000 people.

Sample and Sampling Technique

The purposive sampling technique was used in this study. The students' sample distribution for this study is summarized in Table 1 below:

Table 1: CU Students Sample Distribution [2013/14 Session].

	COVENANT UNIVERSITY, OTA, NIGERIA															
	SUMMARY OF REGISTERED STUDENTS FOR 2013/2014 SESSION															
	100 LEVEL			200 LEVEL			300 LEVEL			400 LEVEL			500 LEVEL			College Total
	F	M	Tot	F	M	Tot	F	M	Tot	F	M	Tot	F	M	Tot	
School of Business Studies	169	109	278	183	100	283	167	123	290	155	122	277	0	0	0	1128
School of Human Resources Develop	123	50	173	117	38	155	95	23	118	118	43	161	0	0	0	607
School of Social Science	153	87	240	177	90	267	140	64	204	112	60	172	0	0	0	883
College of Development Studies (CDS)																2618
Sch. of Engineering	121	432	553	108	387	495	136	423	559	97	304	401	129	416	545	2553
Sch. of Environmt. Studies	49	89	138	41	92	133	50	101	151	49	103	152	34	49	83	657
Sch. of Natural & Applied Sciences	125	191	316	122	160	282	139	175	314	157	186	343	0	0	0	1255
College of Science and Technology (CST)																4465
GRAND TOTAL																7083

Source: Covenant University Data Centre

Instruments

The main instruments used in this study were the Lecturers' Teaching Competence Evaluation Form – HOD's Version [LTCEF-HODV] and the Students Version [LTCEF-SV]

The LTCEF- HODV is divided into 11 sections, namely: Subject Mastery; Human Relations; Communicative Skill; Pedagogical Skill; Class Control/Students' Management; Time Management/Absenteeism; Learning Materials; Testing and Evaluation Skill; Record Keeping & Organisational Skill; Originality, Creativity and Innovation; and ICT and Technology Usage. The instrument ends with an open ended question requesting students to summarize their perception of the Lecturer's teaching competence and comment on other issues not addressed in the form. Each section is a cluster of questions/prompts.

In a previous study by Odukoya, Atayero, Williams, Afolabi and Akande (2014), the face and content validity of the Covenant University LTCEF-SV were established.

Procedure for Data Collection

After creating the Head of Department's [HOD] version of the LTCEF, it was posted on online via the University portal. The Deputy Vice Chancellor [Academics] thereafter requested the HODs to complete evaluation forms online. The students had earlier evaluated their lecturers prior the 2nd semester examinations. Thus, students and HOD's evaluation scores on Lecturers were obtained. Students' achievement scores in the first semester examinations were also obtained from the University data base centre.

Consequently, concurrent validity was established via correlation of score of Students' evaluation of Lecturers in the 2013/14 Omega semester with Head of Departments' evaluation score of the same lecturers in the same semester. The predictive validity was established by correlating students' evaluation score for a lecturer with the students' score in the course undertaken by the lecturer in the 2013/4 Alpha semester.

Data Analysis

Using the Pearson Product Moment Correction Coefficient, the HODs' and Students' evaluation scores were correlated to obtain the Concurrent Validity index while the Students evaluation scores were correlated with their first semester achievement scores to obtain the Predictive Validity index for the LTCEF.

3 RESULT & DISCUSSION

Research Question 1: *What is the Concurrent validity of the CU-LTCEF-SV?*

The hypothesis the Concurrent validity tested was that *there is no significant difference in the evaluation score of Heads of Departments and that of Students*. The result in Table 2 below was obtained as an answer to this question.

Table 2: Correlation Coefficient of Students' and HOH's Evaluation Scores.

	N	r	p
Accounting	16	0.56	Significant
Mass Communication	40	-0.06	Not sigf.
Psychology	14	-0.16	Not sigf.
Civil Engineering	19	0.12	Not sigf.

Source: CU Data Centre [2014] – Data available as at time of Report.

The core findings from these results are as follows:

- Only CU Accounting Department, from the available data, furnished significant indices of concurrent validity. This implies that Students evaluation scores were significantly correlated with their HOD's Lecturers' evaluation scores, using the same instrument. Other Departments reported here did not show similar result.
- Mass Communication and Psychology departments furnished negative indices of correlation coefficient. This suggested that the Students' perception or evaluation or their Lecturers' teaching competence opposed their respective HOD's perception.

These findings tend to reiterate the speculation that either the students did not take their time to conduct proper evaluation of their Lecturers or that the HODs were laden with too many assignments to have ample time to observe and conduct a proper evaluation of the Lecturers.

Research Question 2: *What is the Predictive validity of the CU-LTCEF-SV?*

Table 3: Correlation Coefficient of Students' Course Performance and their Evaluation Score of the Most Senior Lecturer who took the Course.

Accounting		Banking & Fin.				Biochemistry		Architecture	
Course	r	Course	r			Course	r	Course	r
ACC 111	0.05	BFN 111	0.11			BCH211	0.18	ARC 111	0.20
ACC 211	-0.02	BFN 211	0.07			BCH 212	0.06	ARC 112	-0.15
ACC 212	0.16	BFN 311	0.14			BCH 213	0.28	ARC 113	0.23
ACC 310	0.19	BFN 312	0.09			BCH 214	0.30	ARC 114	0.20
ACC 311	0.31	BFN 313	0.20			BCH 311	0.16	ARC 211	0.09
ACC 312	0.20	BFN 316	0.25			BCH 312	0.00	ARC 213	0.11
ACC 313	0.11	BFN 411	0.15			BCH 313	0.20	ARC 214	0.23
ACC 314	0.25	BFN 412	0.12			BCH 314	0.19	ARC 215	0.20
ACC 411	-0.40	BFN 413	0.24			BCH 315	0.16	ARC 216	0.05
ACC 412	-0.30	BFN 415	0.10			BCH 316	0.33	ARC 311	0.14
<i>30% Low Validity</i>		<i>30% Low Validity</i>				<i>40% Low Validity</i>		<i>50% Low Validity</i>	
ACC 413	-0.26	BFN 416	0.05			BCH 317	0.28	ARC 313	0.24
ACC 414	-0.05	BFN 417	0.16			BCH 318	0.15	ARC 314	0.20
ACC 416	0.00	BFN 418	0.11			BCH 411	0.07	ARC 315	0.25
ACC 417	-0.12					BCH 412	-0.04	ARC 316	0.16
ACC 418	-0.09					BCH 413	-1.00	ARC 317	0.04
						BCH 414	0.08	ARC 319	0.02
						BCH 415	0.05	ARC 411	0.21
						BCH 416	-1.00	ARC 412	0.18
						BCH 417	0.33	ARC 413	0.26
						BCH 418	0.40	ARC 414	0.28
						BCH 431	0.27	ARC 415	0.08
						BCH432	0.46	ARC 416	0.03
						BCH 433	0.32	ARC 417	0.19
								ARC 418	0.21
								ARC 419	0.06

Source: Covenant University Data Centre [2014]; For r to be significant [i.e. 2.0+], it must have df 90+ at p = 0.05

Guided by the operational definitions and decision rules submitted for this study, and using the first ten [10] randomly selected courses from two [2] departments that were randomly selected from the two Colleges [CDS and CST] as assessment parameters, the following findings can be deduced from the results in Table 3:

- The College of Science and Technology [CST] furnished more indices of predictive validity for LTCEF-SV [40% and 50% low but significant predictive validities for Biochemistry and Architecture respectively] than the College of Development Studies [CDS] where 30% low but significant predictive validities were obtained for Accounting and Banking & Finance departments.
- No department or course furnished a significant predictive validity index that was above the low validity level.
- Department of Architecture furnished the higher indices of predictive validity than any other department reported in this study.
- The proportion of negative correlation coefficient recorded was more in the Department of Accounting when compared with other departments reported in this study. This was followed by Biochemistry. Negative correlation coefficient implies that students' evaluation scores of a Lecturer were not in direct relation with their scores in the same course taught by the Lecturer.

The implication of these findings is that CST students' evaluations tend to be more valid than CDS students' evaluation. Furthermore, it appears Architecture students offered more valid indices than other departments. This agrees with findings from a previous related study by Odukoya, Atayero, Williams, Afolabi and Priscilla (2014) in which the greatest percentage of Lecturers from CU Department of Architecture were rated highest in terms of manifestation of teaching competence. This is further confirming that CU Architecture students' evaluation of their Lecturers, using the LTCEF-SV, was more valid when compared with other students from other department in the University. Possible explanation for this finding is the highly practical and closeness to 'real life situation' nature of the teaching content and method in this course.

Increasing incidences of negative correlation coefficients is suggesting that some students, which seem to be peculiar to some departments, tend to complete the evaluation carelessly. It could also be suggestive that the LTCEF-SV is not valid; whichever way it is conceived, it is worth further study.

4 RECOMMENDATIONS

Based on the findings and deductions from this study, the following are recommended:

- It is imperative that concerted sensitization seminars be held with Students to give them feedback on the Lecturers' evaluation exercise and more importantly, to enlighten them on the significance of objectively completing the evaluation form
- Heads of Department should be given more time to observe Lecturers in their departments during lectures to enable them make more objective evaluation of their teaching competence.
- Lecturers should regularly be given feedback from the evaluation exercise to allow for positive change, growth and development.
- Other departments should emulate the teaching style of the Lecturers in the Department of Architecture, which is more of teaching students predominantly by '*doing the real things*', that is, professional real life experiences.

5 CONCLUSION

The core objective of this study was to ascertain the criterion related validities of the Covenant University Lecturers' Teaching Competence Evaluation Form – Students' Version [LTCEF-SV]. Specifically, this infers the *Concurrent* and *Predictive* validities. The hypothesis the Concurrent validity tested was that *there is no significant difference in the evaluation scores of the Head of Department and that of Students in the same department*. However, the results obtained tend to reject this hypothesis, thus suggesting that, at it were, in many of the departments, the Students' evaluation scores of Lecturers using the LTCEF-SV did not provide sufficient concurrent validity. The results of the Predictive validities tend to follow the same pattern. The only exception, from the sample of this study, was the Department of Architecture which furnished more indices of *Predictive* validity, though low. The tentative conclusion drawn from these findings, especially against the background of established evidences of face and content validities, is that the challenge of validity may be more human than instrumental. The results obtained from analysis of the responses of CU Architecture students, and related departments, especially from CST, tend to support this speculation. Considering the significance of this study, it is imperative that further studies be undertaken to empirically verify these speculations.

REFERENCES

- [1] American Psychological Association (1974). *Standards for Educational and Psychological Tests*. Washington DC: APA.
- [2] Amatachaya, S.; Naewla, S.; Srisim, K.; et al, (2014). 'Concurrent Validity Of The 10-Meter Walk Test as Compared with the 6-Minute Walk Test in Patients with Spinal Cord Injury at Various Levels of Ability' in *SPINAL CORD* Volume: 52 Issue: 4 Pages: 333-336
- [3] Bachman, L. F., & Palmer, A. (1996). *Language testing in practice*. Oxford, U.K: Oxford University Press.

- [4] Cherry Kendal (2014). What is Validity? Published on the web - <http://psychology.about.com/od/researchmethods/f/validity.htm>
- [5] Dardouri, W.; Gharbi, Z.; Selmi, M. A.; et al (2014). 'Reliability and Validity of a New Maximal Anaerobic Shuttle Running Test' in *International Journal Of Sports Medicine* Volume: 35 Issue: 4 Pages: 310-315.
- [6] Higgins, Jim (2014). *Brief Introduction to Criterion Validation*. United States: Biddle Consulting Group, Inc.
- [7] Horey Jeffrey, Harvey Jennifer, Curtin Pat, Keller-Glaze Heidi, Morath Ray, and Fallesen Jon (2007). *A Criterion-Related Validation Study of the Army Core Leader Competency Model*. Virginia: United States Army Research Institute for the Behavioral and Social Sciences - Technical Report 1199.
- [8] Nishiyama, Takeshi; Mizuno, Tomoki; Kojima, Masayo; et al, (2014). 'Criterion validity of the Pittsburgh Sleep Quality Index and Epworth Sleepiness Scale for the diagnosis of sleep disorders' in *SLEEP MEDICINE* Volume: 15 Issue: 4 Pages: 422-429.
- [9] Odukoya, Atayero, Williams, Afolabi and Akande (2014). '*Overcoming Students' Lethargy In Evaluating Lecturers' Teaching Competence: Covenant University Experience*'. Valencia: INTED 2014 Conference Proceedings.
- [10] Sawaki Yasuyo and Nissan Susan (2002). *Criterion-Related Validity of the TOEFL iBT Listening Section*. United Kingdom: TOEFL iBT Research Report.
- [11] Schneider and Schmitt (1992). Schneider, J. & Schmitt, N. (1992). An exercise design approach to understanding assessment center dimension and exercise constructs. *Journal of Applied Psychology*, 77, 32-41.
- [12] Whetzel, Deborah L.; Rotenberry, Paul F.; McDaniel, Michael A. (2014). 'In-basket Validity: A systematic review' in *International Journal of Selection and Assessment* Volume: 22 Issue: 1 Pages: 62-79.