VALIDATION OF INSTRUMENTS FOR STUDYING TEACHERS' CLASSROOM ASSESSMENT PRACTICES: A PILOT STUDY

Million Yilma, Dilla College of Education
Menna Olango, Hawassa University

ABSTRACT

A valid, reliable and practical instrument of data gathering is needed to conduct a sound research. The aim of this pilot study therefore was to assess the validity and reliability of two sets of instruments to measure teachers' classroom assessment practices in public higher education institutions in Ethiopia. One is the Assessment Practices Inventory which is concerned with teachers' self-perceived assessment skills and teachers' use of assessment practices in two different scales, "skill" and 'Use'. The other is developed by the researcher concerning 'factors shaping teachers' classroom assessment practices'. Here, a variety of factors at different levels that shape teachers' classroom assessment practices in some way are dealt with. The questionnaire was distributed to 76 randomly selected teachers/lecturers from 3 randomly selected public universities in the country. However, the analysis was done based on data obtained from 66 of the respondents who have properly completed and returned the questionnaire. The findings from this pilot study established face validity, content validity, construct validity as well as internal consistency of the instruments. Moreover, the factors/components from each of the scales are determined. Therefore, on the basis of this pilot study, the validity and reliability of the instruments to be used for the more comprehensive study are ensured.

Keywords: Classroom Assessment, Survey Instruments, Validity, Reliability, Factor Extraction.

INTRODUCTION

An instrument is valid when it is measuring what is supposed to measure or, in other words, when an instrument accurately measures any prescribed variable, it is considered a valid instrument for that particular variable. Reliability on the other hand is defined as *"the extent to which test scores are free from measurement error"*. It is a measure of stability or consistency of an instrument in measuring certain concepts (Altheide & Johnson, 1994; Shekhar Singh, 2014; Field, 2005).

When it comes to validity, several varieties have been described which include face validity, content validity, construct validity, and criterion validity. Face validity refers to subjective assessments of the presentation and relevance of the measuring instrument as to whether the items in the instrument appear to be relevant, reasonable, unambiguous, and clear (Oluwatayo, 2012). It evaluates the appearance of the questionnaire in terms of feasibility, readability, consistency of style and formatting, and the clarity of the language used. This is the least scientific method of validity, as it is not quantified using statistical methods. Content validity refers to the degree to which items in an instrument reflect the content universe to which

Citation Information: Yilma, M., & Olango, M. (2023). Validation of instruments for studying teachers' classroom assessment practices: A pilot study. Academy of Educational Leadership Journal, 27(1), 1-10.

the instrument will be generalized (Straub et al., 2004). It assesses whether a test is representative of all aspects of the construct. To produce valid results, the content of a test, survey or measurement method must cover all relevant parts of the subject it aims to measure. If some aspects are missing from the measurement (or if irrelevant aspects are included), the validity is threatened (Middleton, 2022). Construct validity is about ensuring that the method of measurement matches the construct (a trait or mental process) you want to measure. If you develop a questionnaire to diagnose depression, you need to know: does the questionnaire really measure the construct of depression? Or is it actually measuring the respondent's mood, self-esteem, or some other construct (Middleton, 2022; Straub et al., 2004)? Lastly, criterion or concrete validity is the extent to which a measure is related to an outcome. It measures how well one measure predicts an outcome for another measure (Taherdoost, 2016; Middleton, 2022). A test has this type of validity if it is useful for predicting performance or behavior in another situation (past, present, or future).

This pilot study was conducted so as to validate the set of instruments to study teachers' classroom assessment practices in public higher education institutions. Classroom assessment is an integral part of the ongoing teaching learning process throughout the academic year (Ewell, 2009). Teachers spend considerable amount of their professional time on assessment activities that inform a wide variety of decisions made daily which directly influence students' learning experiences in Harvey & Williams (2010). Ethiopian teachers are no exception in this regard as they devote a considerable amount of their time and energy to assessment and related activities. It ranges from devising an assessment technique to practically putting that in to practice as well as to giving feedback to students for improving their achievement. Here, teachers dwell on a variety of decisions related to what sorts of assessments are appropriate for students, when and how to undertake assessments, and how to make an effective use of the information obtained from the assessments, among others things. As a result, it is extremely important to use valid and reliable instruments so as to effectively study classroom assessment.

When it comes to Ethiopian public higher education institutions/universities, there seems to be a lack of understanding in relation to teachers' classroom assessment practices as to for example what are their perceived skills in relation to classroom assessment and what exactly determines their classroom assessment practices, among other things. To this end, to make a thorough understanding of these issues, appropriate data must be gathered from the target population. In order to do so, it is essential that appropriate instruments for data gathering are devised. Finding out such instruments requires a series of steps in order to make sure that they are up to the required standard in terms of such essential parameters as validity and reliability. Therefore, this pilot study made an attempt to validate a set of instruments to examine various aspects of teachers' classroom assessment practices in public higher education institutions in Ethiopia in line with school/institutional, class, teacher and student level variables.

Objectives

This pilot study aimed to validate two sets of instruments: the Assessment Practices Inventory (API) and the instrument designed to determine factors shaping teachers' classroom assessment practices. It aimed to come up with valid and reliable set of instruments for making a comprehensive study of various aspects of teachers' classroom assessment practices in public higher education institutions in Ethiopia.

METHODS

Citation Information: Yilma, M., & Olango, M. (2023). Validation of instruments for studying teachers' classroom assessment practices: A pilot study. Academy of Educational Leadership Journal, 27(1), 1-10.

In this section, some important details of the respondents who were included in this pilot study and the other methods employed such as information with regard to the data collection instruments and the data analysis techniques are presented.

Participants of the Study

When it comes to selecting the samples for this study, similar contexts to that of the actual study population were taken into account. In this regard, a total of 76 teachers from three different universities (Dilla, Wolkite, and Bulehora) were randomly selected and included as respondents to fill out the survey questionnaire. However, 5 respondents didn't return the questionnaire while 5 other respondents didn't complete it properly. Therefore, the study was undertaken with the data obtained from the 66 respondents who properly filled out and returned the questionnaire. In this respect, 21 respondents from Wolkite University, 20 from Bulehora and 25 from Dilla University were included in this pilot study. Important demographic information of the respondents is presented as follows.

Table 1
DEMOGRAPHIC CHARACTERISTICS OF RESPONDENTS

Variable	Frequency	Percent
Sex		
Male	56	84.8
Female	10	15.2
Level of education		
BSc/BA	9	13.6
Masters	53	80.3
PhD	4	6.1
Field of specialization		
Engineering & technology	1	1.5
Natural & computational science	16	24.2
Agricultural science	9	13.6
Health and medical science	7	10.6
Computing & informatics	3	4.5
Social science & humanities	10	15.2
Education & behavioral science	20	30.3
Teaching level		
Graduate assistance	9	13.6
Under graduate	51	77.3
Graduate	5	7.6
Post graduate	1	1.5
Assessment training		
No training	8	12.1
Some lessons on CA	14	21.2
A course on CA	8	12.1
More than one course on CA	7	10.6
In-service training on CA	5	7.6
Adequate training on CA	24	36.4

When it comes to teaching experience of respondents, it ranged from 1 to 30 years. This indicated that the respondents included those newly recruited ones to those who stayed in the teaching profession for plenty of years already.

Citation Information: Yilma, M., & Olango, M. (2023). Validation of instruments for studying teachers' classroom assessment practices: A pilot study. Academy of Educational Leadership Journal, 27(1), 1-10.

Instruments of Data Collection

With regard to the instruments of data collection used for this pilot study, the API-(Assessment Practices Inventory) was employed along with some other items dealing, for example, with demographic characteristics (6 items) and one (1) item asking for assessment training teachers received. The API was used for assessing concerns regarding classroom assessment practices. It was developed within the theoretical framework delineated by the literature on classroom assessment. It consisted of 67 items measured on two rating scales: *"Use scale"* and *"Skill scale"*. The *"use"* scale was meant to measure teachers' assessment practices on a scale from 1 (*not at all used)* to 5 (*used very often*). The *"skill"* scale on the other hand was designed to measure teachers' self-perceived assessment skills with a scale ranging from 1 (*not at all skilled) to* 5 (*very skilled*). The Cronbach alpha reliabilities of the scales in the API ranged from 0.89 to 0.77 for assessment practices and from 0.91 to 0.85 for self-perceived assessment skills.

In addition to the API, an instrument meant to gather data in relation to factors that shape teachers' classroom assessment practices (35 items initially) in Ethiopian higher education institutions was developed. The items were prepared considering variables that range from the school/institution level to teacher, classroom and finally to student levels. Therefore, these items were pilot-tested to find out the final set of items to be used for the actual data collection as reported here under by going through all the important steps.

Data Analysis Techniques

The data were processed and analyzed using IBM SPSS Statistics 25. When it comes to the specific techniques used, descriptive statistics was used to indicate the demographic characteristics of the respondents. The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) was conducted to ensure the adequacy of the samples per item. Initially, correlation between items and the effect of mulitcollinearity were determined using Bartlett's Test of Sphericity and determinant of the R-matrix respectively. Principal component analysis (PCA) was used to determine the components/factors within the scales. For reliability, internal consistency of the instruments was determined using Cronbach's alpha. Finally, Principal Component Analysis (PCA) with varimax rotation method was used to determine construct validity of the instruments.

RESULTS

In this section, the major issues in line with the findings of the preliminary are included. In this regard, such issues as sampling adequacy and correlation between items, mulitcollinearity information, reliability and validity as well as factor extraction issues are outlined.

Sampling adequacy and correlation between items

The skills scale

It is outlined that the larger the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) value is, the better (or a value of 0.5 and above is acceptable). In this particular case, for the skills scale, the KMO value is found to be 0.81, which is a good indication that an adequate

Citation Information: Yilma, M., & Olango, M. (2023). Validation of instruments for studying teachers' classroom assessment practices: A pilot study. Academy of Educational Leadership Journal, 27(1), 1-10.

sample size per item is incorporated. On the other hand, in relation to correlation between items, two points are worth considering. The first one is Bartlett's Test of Sphericity. If this test is significant, then the R-matrix is not approximately I or correlation is non-zero. Here, since the test was found to be significant ($p<0.001$), then there was no such problem. The second one is the effect of multi-collinearity which can be detected by looking at the determinant of the R-matrix whose value is expected to be greater than 0.00001 which in this case is 3.96E-009. Therefore, no problem of multi-collinearity was detected.

The use scale

When it comes to the use scale, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) value was found to be 0.817 which indicated that an adequate sample size per item was incorporated. On the other hand, in relation to correlation between items, Bartlett's Test of Sphericity was found to be significant (($p<0.001$). This is a good indication that the items in this particular scale were acceptably correlated to each other. Furthermore, there was no effect of multi-collinearity as the determinant was greater than 0.00001.

The 'factors' scale

With regard to the items in the factors scale, the Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO) value was found to be 0.66 which is an acceptable value indicating that an adequate sample size per item was incorporated in the study. On the other hand, Bartlett's Test of Sphericity was found to be significant ($p<0.001$) which is a good indication that the items in this particular scale were acceptably correlated to each other. Furthermore, there was no effect of multi-collinearity as the determinant (1.294E-005) was found to be greater than 0.00001.

Factor structures of the instruments

Principal component analysis (PCA) is used to reduce a large number of items into a smaller number of factors that can be analyzed and interpreted with ease (Field, 2005; Thomson, 2005).

In this regard, in order to extract the factors/components for this particular study, PCA (Promax oblique factor rotation method specifically), which is commonly used when factors are correlated (Field, 2005; Thomson, 2005), was used on the *"skill"* and *"use"* assessment practices scales as well as for the newly developed instrument related to factors shaping assessment practices. Therefore, based on scree plots and Kaiser' Little Jiffy (eigenvalues >1), the items from *"skill"* subscale converged into six factors that accounted for 67.64% of the variance in item responses after extraction. On the other hand, the items in the *"use"* scale converged into seven factors. However the seventh factor was composed only of a single item which in factor extraction is not acceptable because at least two items are expected to be loaded into a given factor. Therefore, the six factors were retained to be of value for the *"use"* subscale as well. These factors accounted for 67.85% of the variance in item responses after extraction. Factor loading values greater than 0.40 had shown that an item loaded on to a particular factor.

When it comes to the items in the *"Factors shaping assessment practices"* scale, Kaiser' Little Jiffy (eigenvalues >1) criterion outlined that the items converged into 6 factors which accounted for 64.8% of the variance in item responses after extraction. However, the scree plot indicated that 4 factors are of value to this particular scale. This was supported by the fact that

Citation Information: Yilma, M., & Olango, M. (2023). Validation of instruments for studying teachers' classroom assessment practices: A pilot study. Academy of Educational Leadership Journal, 27(1), 1-10.

the 5th and the 6th factors contained only one item loaded to each which in factor extraction is not acceptable because at least two items are expected to be incorporated in a factor or construct. As a result, the 4 factors were taken to be important for the scale, *"factors shaping teachers"* classroom assessment practices.

Reliability and validity

Reliability and validity are the two most important and fundamental features in the evaluation of any measurement instrument or tool for a good research. Therefore, these two issues were focused up on in this study. *Reliability* can be thought of as consistency. It intends to make sure whether a given instrument consistently measures what it is intended to measure. On the other hand, *validity* is the extent to which an instrument measures what it is supposed to measure and performs as it is designed to perform.

Reliability

Reliability was one of the things given due attention in this preliminary study as a very important aspect of research instrument. In this regard, the entire 67 items in the API were given to respondents to fill out categorized under *"skills"* and *"use"* scales. Such issues as internal consistency, item-total correlations as well as inter-item correlations were checked for. With regard to internal consistency for example, Cronbach's alpha for all items in each scale (use and skill) as well as for items within *"factors shaping assessment practices"* scale were taken into account and checked for. Consequently, for the *"Skills*" scale, the overall Cronbach's alpha value was found to be 0.94, which indicates that was strong internal consistency of the items. This value was obtained after few items were left out of the original items in the API because their values were found to be beyond the overall alpha value and they had lower (below 0.30) corrected item-total correlation values. However, the remaining items were found to have acceptable levels of item-total correlation values. A few items on the other hand were reshaped accordingly by rewording them in a way that suits in context to Ethiopian teachers in higher institutions. This was done on the basis of experts' consultation especially in the area of English language and classroom assessment in addition to the researcher's own personal experience in these particular areas. Therefore, 30 items from which 6 factors had been extracted were identified to be off value to assess teachers' perceived assessment skills for this particular study. The six factors were extracted based on the screen plot and eigenvalues for the initial solution.

Cronbach's alpha for the six skills factors were found to be as follows. The first factor which was about *"Perceived skillfulness in Standardized Testing, Test Revision, and Instructional Improvement"* had 5 items loaded to it. It had a Cronbach's alpha value of 0.87 which is way above the minimum standard required, α=0.70 (Nunnally, 1978). Each of the items in this factor was acceptably correlated to the total, with the minimum value being 0.63. In addition, the items were acceptably correlated to each other with the minimum value being 0.50. On the other hand, the second factor which is about 'Perceived skillfulness in Assessment application' included 7 items with a Cronbach's alpha value of α=0.86. For this factor, items were acceptably correlated to the total with the minimum value being 0.62 and also these items were acceptably correlated to each other with the minimum value being, 0.40. The third factor which was about perceived skillfulness in assessment planning had 4 items loaded to it having a Cronbach's alpha value of α=0.80. Furthermore, each of the items was acceptably correlated to

Citation Information: Yilma, M., & Olango, M. (2023). Validation of instruments for studying teachers' classroom assessment practices: A pilot study. Academy of Educational Leadership Journal, 27(1), 1-10.

the total with a minimum value of 0.55 and also the items in the factor were correlated to each other with the minimum value being 0.40.

The fourth factor ('Perceived skillfulness in Crieterion referenced testing') to which 7 items were loaded had a Cronbach's alpha value of α=0.79. Similarly, the items in this specific category were acceptably correlated to the total with the minimum value of 0.50. Each of the items was also found to be acceptably correlated to the other ones with the minimum value being 0.32. With regard to the fifth factor (Perceived skillfulness in Grading practices) which contained 4 items loading to it, a Cronbach's alpha value of α=0.76 was obtained. This was accompanied by an acceptable level of item-total correlation with the minimum value being 0.50 and also with items acceptably correlating to each other with a minimum value of 0.40. Finally, the sixth factor which was about perceived skillfulness in writing paper/pencil tests contained 3 items with a Cronbach's alpha value of α=0.71. In this particular factor, the items correlated acceptably to the total with a minimum value of 0.45 and having acceptable levels of inter-item correlation with the minimum value being 0.36. In conclusion, the items for teachers' perceived assessment skills were found to have acceptable level of internal consistency in general and within the specific factors in particular. In addition, they were found to have acceptable levels of item-total as well as inter-item correlations.

When it comes to the items in the *"use"* scale (frequency of use of assessment practices), those same 30 items included in the skills scale were also retained to be of value. In this regard, the items in this particular scale were found to have an overall Cronbach's alpha value of α=0.94 which depicts strong internal consistency of the items. Here too, those items from the API which were found to have very low item-total correlations and inter-item correlations were left out.

In relation to the reliability information of the specific factors within the *"Use scale"*, the following were found to be very important. The first factor (Standardized Testing, Test Revision, and Instructional Improvement) had 5 items loaded to it. A Cronbach's alpha value of 0.84 which is way above the minimum standard required was obtained for this particular factor. Each of the items in this factor was acceptably correlated to the total, with the minimum value being 0.55. In addition, the items were acceptably correlated to each other with the minimum value being 0.37. The second factor (Assessment application) which included 6 items had a Cronbach's alpha value of α=0.84. In this particular factor, the items were found to be acceptably correlated to the total with the minimum value being 0.53 and also they were found to have acceptable level of inter-item correlation with the minimum value being, 0.38. On the other hand, the third factor (4 items) which was about assessment planning had a Cronbach's alpha value of α=0.77. Similarly, each of the items in this factor was acceptably correlated to the total with a minimum value of 0.48 and also the items were found to have acceptable inter-item correlation level with the minimum value being 0.45.

The fourth factor (Crieterion referenced testing) to which 6 items were loaded had a Cronbach's alpha value of α=0.79. Similarly, the items in this specific category were acceptably correlated to the total with the minimum value of 0.50. Each of the items was also found to be acceptably correlated to the other ones with the minimum value being 0.32. With regard to the fifth factor (Grading practices) which contained 4 items loading to it, a Cronbach's alpha value of α=0.76 was obtained. This was accompanied by an acceptable level of item-total correlation with the minimum value being 0.50 and also with items acceptably correlating to each other with a minimum value of 0.40. Finally, the sixth factor which was about 'Using paper/pencil tests' contained 3 items with a Cronbach's alpha value of α=0.71. In this particular factor, the items correlated acceptably to the total with a minimum value of 0.45 and having acceptable levels of

Citation Information: Yilma, M., & Olango, M. (2023). Validation of instruments for studying teachers' classroom assessment practices: A pilot study. Academy of Educational Leadership Journal, 27(1), 1-10.

inter-item correlation with the minimum value being 0.36. In conclusion, the items within the *"use"* scale were found to have acceptable level of internal consistency in general and within the specific factors in particular. In addition, they were found to have acceptable levels of item-total as well as inter-item correlations.

Therefore, a total of 60 items from the API were retained to be the final questionnaire within two subscales namely, teachers' perceived assessment skills and teachers' use of assessment practices. Here, some items were reshaped in wording without losing the original idea in a way that suits to the context of the respondents in the study areas while 7 items were not included.

The similarities and differences between the factor structures of the two scales, *("use"* and *"Skills"*), may be highlighted as follows. With the exception of a few items, the two factor structures were similar in almost all the underlying dimensions such as paper-pencil tests; standardized testing, test revision, and instructional improvement; Criterion referenced testing among others.

On the other hand, when it comes to the items regarding factors shaping teachers' classroom assessment practices', a pool of items were prepared and pilot tested. In this regard, a total of 35 items were initially distributed to respondents in the selected universities for pilot testing. After going through essential reliability and validity tests as well as factor extraction issues, a total of 27 items were retained to be of great value to collect data regarding factors shaping teachers' classroom assessment practices. Scores for the instrument as a whole had a relatively strong reliability coefficient in this sample (α=0.83) which indicates that there was strong internal consistency amongst them.

With regard to the four components extracted from the actors shaping teachers' classroom assessment practices' subscale, the following reliability information were found to be important. The first component which is about teacher related variables got 7 items loaded to it. These items relate to different characteristics of teachers which in some way shape their classroom assessment practices such as teaching experience, level of education, assessment training received, and field of specialization, among others. The second component which is about student related factors included 5 items dealing with such issues as student gender, level of education of students, and students' field of specialization which are thought to shape teachers' classroom assessment practices in some way. The third component on the other hand is about classroom level variables determining teachers' classroom assessment practices. There are 7 items included in this particular category which reflect such important issues as class size, provision of educational facilities or technology, the course content, and the instructional approaches/pedagogy employed. Lastly, the last component relates to school/institutional variables shaping teachers' classroom assessment practices. Included in this category are items dealing with such important issues as professional community, school emphasis on academic achievement, school leadership at different levels, and school/institutional goals and priorities.

With regard to the reliability information of each of the factors as well as the items within these factors, the following result has been obtained. The first factor (teacher level variables) had Cronbach's alpha value of 0.74 which is above the minimum required standard (7.0). Furthermore, the items in this particular factor are found to have acceptable level of inter-item and item-total correlations. The second factor (classroom level variables) was found to have Cronbach's alpha value of 0.72 which depicts a relatively strong reliability within the items. This was supported also by acceptable level of inter-item correlations with the minimum value being 0.30. Each of the items was also found to acceptably correlate with the total. When it comes to

Citation Information: Yilma, M., & Olango, M. (2023). Validation of instruments for studying teachers' classroom assessment practices: A pilot study. Academy of Educational Leadership Journal, 27(1), 1-10.

the third factor (school level variables) on the other hand, a Cronbach's alpha value of 0.59 was obtained. However, acceptable levels of inter-item as well as item-total correlations were depicted. Finally, the fourth factor (student level variables) was found to have a Cronbach's alpha value of 0.76 that is an indicator of a relatively strong reliability. Here too, acceptable levels of inter-item as well as item-total correlations were displayed.

Validity

Validity concerns what an instrument measures, and how well it does so. As there are various types of validity, different techniques were employed to make sure the instruments fulfilled important validity standards. In this regard, subject specialists in the area of assessment and evaluation, as well as language, along with the researcher's own professional assessment made sure the instruments had face validity fulfilled. On the other hand, to ensure the content validity of the instruments, a table of specifications was used to generate items for each major aspect of classroom assessment in the first place. Expert judges or panels which included assessment and evaluation specialists were also consulted to make sure the instruments have got content validity fulfilled. In this regard, those items rated as essential by a proportional level of experts were retained with the minimum content validity ratio (CVR) being P=0.05 (Lawshe, 1975 in Taherdoost, 2016). Furthermore, a series of literature reviews (Ghaicha, 2016; Mayer et al., 2000; Nilsen & Gustafsson, 2016) were conducted so as to extract the related items to establish content validity of the instruments. This was given due attention especially for the newly developed instruments regarding factors shaping classroom assessment practices.

The other very important aspect of validity is construct validity. A measurement technique has construct validity if it is related to things to which we expect the concept we are trying to measure to be related, and independent of those things of which the concept should be independent. A preliminary estimation of construct validity which is termed as the most critical validity measure (Kerlinger and Lee, 2000) was made in this study. In this regard, the extent to which the data exhibit support of convergent and discriminant validity was checked.

It is outlined that with the purpose of verifying the construct validity (discriminant and convergent validity), a factor analysis can be conducted utilizing Principal Component Analysis (PCA) with varimax rotation method (Koh and Nam, 2005, Wee and Quazi, 2005). In this regard, items loadeding above 0.40, which is the minimum recommended value in research are considered for further analysis. Also, items cross loading above 0.40 should be deleted. Therefore, the factor analysis results will satisfy the criteria of construct validity including both the discriminant validity (loading of at least 0.40, no cross-loading of items above 0.40) and convergent validity (eigenvalues of 1, loading of at least 0.40, items that load on posited constructs) (Straub et al., 2004). In line with this, out of the 67 items with in the Assessment Practices Inventory (API), 7 items which loaded below the minimum recommended value (0.40) and which cross-loaded above 0.40 were deleted. Therefore, the remaining 60 items were retained to be important. In a similar note, 8 items were excluded from the *"factors shaping assessment practices"* scale which initially included 35 items. As a result, 27 items were considered to be of value.

CONCLUSION

The kind and quality of data gathering instruments are vital to make the objective of a given study a success. In this study attempt was made to come up with valid and reliable survey

Citation Information: Yilma, M., & Olango, M. (2023). Validation of instruments for studying teachers' classroom assessment practices: A pilot study. Academy of Educational Leadership Journal, 27(1), 1-10.

instruments that can be applied to investigate teachers' classroom assessment practices in public higher education institutions. The instruments were of two types. First, important items from the API (Assessment Practices Inventory) were adapted and made ready for use. This instrument includes items related to teachers' perceived assessment skills and items that are related to teachers' frequency of use of different aspects of classroom assessment practices. Second, items regarding factors shaping teachers' classroom assessment practices were developed by the researcher. After going through a series of steps, those items fulfilling reliability and validity issues are retained.

REFERENCES

Altheide, D.L., & Johnson, J.M. (1994). Criteria for assessing interpretive validity in qualitative research.

Ewell, P. (2009). *Assessment, Accountability and Improvement: Revisting the Tension*. National Institute for Learning Outcomes Assessment.

Field, A.P. (2005). *Discovering statistics using SPSS:(and sex, drugs and rock'n'roll)*. Sage.

Ghaicha, A. (2016). Theoretical framework for educational assessment: A synoptic review. *Journal of Education and Practice*, *7*(24), 212-231.

Harvey, L., & Williams, J. (2010). Fifteen years of quality in higher education (Part Two).

Mayer, D.P., Mullens, J.E., & Moore, M.T. (2001). Monitoring school quality: An indicators report. *Education Statistics Quarterly*, *3*(1), 38-44.

Middleton, F. (2022). The 4 types of validity in research | Definitions & examples.

Nilsen, T., & Gustafsson, J.E. (2016). *Teacher quality, instructional quality and student outcomes: Relationships across countries, cohorts and time* (p. 166). Springer Nature.

Oluwatayo, J.A. (2012). Validity and reliability issues in educational research. *Journal of Educational and Social Research*, *2*(2), 391-400.

Shekhar Singh, A. (2014). Conducting case study research in non-profit organisations. *Qualitative Market Research: An International Journal*, *17*(1), 77-84.

Straub, D., Boudreau, M.C., & Gefen, D. (2004). Validation guidelines for IS positivist research. *Communications of the Association for Information Systems*, *13*(1), 24.

Taherdoost, H. (2016). Validity and reliability of the research instrument; how to test the validation of a questionnaire/survey in a research. *How to test the validation of a questionnaire/survey in a research*.

Received: 09-Dec-2022, Manuscript No. AELJ-22-12988; **Editor assigned:** 12-Dec-2022, PreQC No. AELJ-22-12988(PQ); **Reviewed:** 26-Dec-2022, QC No. AELJ-22-12988; **Revised:** 30-Dec-2023, Manuscript No. AELJ-22-12988(R); **Published:** 06-Jan-2023

Citation Information: Yilma, M., & Olango, M. (2023). Validation of instruments for studying teachers' classroom assessment practices: A pilot study. Academy of Educational Leadership Journal, 27(1), 1-10.