AJOTE
African Journal of Teacher Education
ISSN 1916-7822. A Journal of Spread Corporation

Volume 14 No. 1 2025 Pages 87-119

Teachers' conceptualization of validity and reliability in classroom assessment in Ghana

Abraham Yeboah, Ato Kwamina Arhin
Akenten Appiah-Menka University of Skills Training and Entrepreneurial Development, Ghana

Abstract

Assessment must be conducted so that the result is reliable and valid. However, in Ghana, standardised tests are not available for use by classroom teachers, therefore, classroom teachers depend on their little knowledge in conducting classroom assessments. The study explored teachers' conceptualization of validity and reliability in classroom assessment in public senior high schools in the Kumasi Metropolis in the Ashanti Region of Ghana. The quantitative descriptive survey design was used for the study. The purposive sampling method was used to select 200 teachers for questionnaire responses on teachers' conceptualization of validity and reliability. The data collection instrument was the questionnaire. The Cronbach's coefficient alpha for the items on the questionnaire was 0.85. The data analysis procedure utilized means and standard deviations. Findings showed that teachers' conceptualization of validity and reliability enabled them to align the content of tests with learning outcomes of the curriculum unit and construct test items to solicit different types of information to make judgements of what students know. It was recommended that heads of senior high schools together with assessment experts give periodic training to teachers on congruent validity. Teachers' conceptualization of validity and reliability will enable them to construct effective test items, and administer, grade and interpret results to enhance students' learning.

Keywords: Teachers, validity, reliability, classroom, assessment, senior high school

Introduction

Effective teaching and learning are enhanced through classroom assessment. Classroom assessment has been described as the cornerstone of the current education system (Brookhart, 2011; Popham, 2014; Kelly, Feistman, Dodge, Rose., & Littenberg-Tobias, 2020). Classroom assessment involves the collection of information from a variety of sources, with the intention

of promoting effective teaching and learning (Kane & Saskia, 2020, Yeboah, 2019), "formal and informal procedures that teachers employ in an effort to make accurate inferences about what their students know and can do" (Popham, 2009, p. 6). Classroom assessment helps to determine students' status with respect to educational variables of interest such as how students vary in how much they know about a subject, how skilled they are in performing such operations as long division, and how positive their attitudes are toward classroom activities and school (Popham, 2011; Black & Wiliam, 2018, Tontus, 2020). Classroom assessments take a variety of forms, such as teacher observations of the students in various contexts Watson (2023), interactions with students, quizzes, tests, assignments, and projects. It is only through assessment that one can discover whether the instructional activities in which one engages students resulted in the intended learning (Wiliam, 2013). Classroom assessments provide rich sources of information on student performance, skills, and achievement.

Classroom assessment plays many essential roles in education (Yeboah, Gyamfi & Sam, 2019), helps to determine students' current status, to monitor students' progress, to assign grades to students and to determine instructional effectiveness (Popham, 2011; Stiggins & Chappuis, 2012), in supporting students' capability to make evaluative judgements about their work and that of others (Fischer, Bearman, Boud & Tai, 2023). Remediation occurs when standard of instruction is ineffective, requiring special instruction or alternative methods or materials. Diagnostic assessments help determine the best learning activities to meet individual student needs and maximize opportunities to achieve targets. Significantly, selection decisions are also dear to the heart of assessment supervisors (OECD, 2010). This is because an organization decides that some persons are acceptable while others are not (Nitko, 2001). For classroom assessment to serve these purposes, the results must be valid and reliable.

For all assessments, validity is an important concern (American Educational Research Association [AERA], American Psychological Association [APA] and National Council on Measurement in Education [NCME], 2014). Kane and Saskia (2020) explain validity as what is assessed and how well this corresponds with the behaviour or construct that it is intended to test or assess. Validity is the soundness of one's interpretation and uses of students' assessment results (Nitko, 2001). This means that teachers in the senior high schools' conceptualization of validity of assessment results would probably mean that results must be relevant and reflect the thinking processes of students. This would imply that students' performance would reflect the results obtained from assessment.

Darling-Hammond, Herman, Pellegrino and Abedi (2013) noted that valid assessments produce data that can be used to inform education decisions at multiple levels, from school improvement and effectiveness to teacher evaluation to individual student gains and performance. Darling-Hammond et al. (2013) further mentioned that for classroom assessment results to be considered valid, an assessment should be a good representation of the knowledge and skills it intends to measure. An assessment instrument must be such that the result is valid. To achieve valid results, assessment instruments should be selected or constructed appropriately. Yeboah (2018) noted that test specification tables are not taken into consideration by the secondary school teachers in Ghana when constructing test items thereby rendering the validity of the test items rather low or unknown. This suggests how senior high school teachers conceptualize validity in educational assessment.

In addition to sound classroom assessment that reflects student current thinking, reliability is crucial. Reliability is the degree of consistency between two measures of the same thing (Liaquat, Asif, Siraji & Maroof, 2012). It deals with the degree to which assessment results would be similar under slightly different measurement conditions. For example, if one assesses a student twice, one hopes that one will obtain almost the same score if one assesses the student one day later. Here, if one measures a person's level of achievement, one hopes that the scores will be similar under different administrators, using different scorers, with similar but not identical items (Yeboah, 2019). Teachers' conceptualization of reliability in classroom assessment would suggest that teachers would construct test items bearing in mind the guidelines regarding test items construction, follow apt means to administer, score and grade students' assessment. This would enable students to demonstrate a particular behaviour consistently upon going through assessment. In Ghana, it appears that senior high school teachers lack skills in test item construction. Quansah, Amoako and Ankomah (2019) opined that senior high school teachers in Ghana are weak in test construction. Teachers, albeit with some principles of test construction followed, most of the critical issues which are related to validity and reliability are overlooked. This raises questions about the validity and consistency of students' results.

Statement of the Problem

Classroom assessment is the lifeblood of teaching and learning processes. Assessment results become useful and purposeful when the results are valid and reliable. However, although, Black and Wiliam (2012) indicated that classroom assessment helps to determine the mastering

level of what students have been taught, it is affected by suitability of the questions for the students being assessed, phrasing of the questions, consistency in test administration, for example, the length of time given for the assessment, directions given to students before the test etc. Furthermore, Johnson (2013) notes that lack of clarity and applicability of assessment criteria leads to classroom assessment unreliability.

In the context of Ghana, Yeboah (2018) observed that teachers have been taught how to evaluate their test items, take necessary steps to improve quality, or accurately set criterion levels for student performance. Despite this training, many senior high school teachers often rely on their own inadequate experiences of constructing poor test items to assess their students. Similarly, studies reveal that some Ghanaian senior high school teachers lack a strong preference for test construction due to inadequate knowledge of how to develop valid and reliable tests based on their teaching and learning domains (Kissi, Baidoo-Anu, Anane & Annan-Brew, 2023; Adiyaa, Osei-Poku & Essel, 2022; Frimpong & Osei, 2021), this was due to teachers' low literacy in using a table of test specification (Asamoah, Shahrill & Abdul Latif, 2023).

Quansah, et al. (2019) reported that SHS teachers in Ghana have a negative attitude towards test construction. Poor attitude of teachers in the planning of test, item writing, item review and assembling of the items was reported. Quansah, et al. concluded that this attitude of teachers influenced the quality of test used for assessing students. It is of essence teachers see test construction as a burden. Test items flaws introduce systematic error of construct-irrelevant variance to assessments (Downing, 2005), thereby reducing the validity evidence for examinations and penalizing some examinees (Räisänen, Tuononen, Postareff, Hailikari & Virtanen, 2016). These problems with the validity and reliability of assessment lead to perceptions that the assessment is unfair. For methods, Quansah et al. employed document analysis, whereas Asamoah et al. utilized a mixed method approach but focused on different aspects of validity and reliability than the conceptualization in this current study.

The current study employs the descriptive method and discusses validity and reliability issues such as relevance, thought processes, congruency, differentiation of assessment types, communication of expectations, and systematic assessment procedures. Thus, this study fills this gap regarding the use of test in the classroom on validity and reliability in the Ghanaian context.

Research Questions

The study was guided by this research question.

1. What are senior high school teachers understanding of validity and reliability in classroom assessment?

LITERATURE REVIEW

Validity and reliability theory

Classroom assessment when effectively conducted yield meaningful decisions. In the classroom, effective decisions are taken when test results are valid. Validity is "the degree to which evidence and theory support the interpretations of test scores entailed by the proposed uses" of a test (AERA, APA, & NCME, 2014, p. 11). The validity of assessment refers to "grade integrity" which is about the extent to which grades correspond with the quality, breadth, and depth of students' academic achievement (Sadler, 2009). Validity of assessment information refers to its meaning and value Brookhart (2003), the accuracy of test-based interpretations of scores (Popham, 2016) and consequences for learning and instruction Moss (2003). According to Nitko (2001), validity refers to the "soundness of one's interpretation and uses of students' assessment results". This means that for teachers in the senior high schools to produce valid results of their students, the students' results must be supported with series of evidence. The issue of validity concerns the interpretations and uses of measurement scores. The interpretations and uses of one's assessment results are said to be valid only when the values implied by them are appropriate.

Categories of validity evidence include content-related evidence, criteria-related evidence, and construct-related evidence. Yeboah (2019) indicated content-related evidence of validity as the extent to which the sample of items, tasks, or questions on a test is representative of the domain of content. Content validity provides senior high school teachers with knowledge to select adequate content areas when assessing their students. For teachers at the senior high schools to construct good test, it is important for teachers to have knowledge about construct validity. Construct validity indicates one's ability to translate or transform a concept, idea, or behaviour into a functioning reality (Trochim, 2006; DeLuca & Klinger, 2010; Moss, 2003). Further, criterion-related evidence of validity has to do with the degree of correspondence between a test measure and one or more external referents (criteria). Criterion-related evidence answers the question, how well the results of an assessment can be used to infer or predict an individual's standing on one or more outcomes. At the senior high schools in Ghana, teachers have acquired knowledge about validity during their training at the education universities. But,

the question is, are senior high school teachers able to use this knowledge during classroom assessment?

Components of validity

The study focused on teachers' conceptualization of validity and reliability in relation to classroom assessment. Components of validity included relevance, thought processes, and congruency. Relevance ensures that assessment measures what it is intended to measure. It is about aligning the content of the assessment with the specific objectives it aims to assess (Messick,1989; Popham, 1997). If assessment is not relevant, the results will not accurately reflect the true abilities of individuals being tested. Level of thought processes deal with how assessments accurately measure the intended cognitive skills acquired by learners at an appropriate level (Wiggins, 1990; Messick,1989). Ensuring that an assessment aligns with these levels helps in validating that it accurately measures the intended cognitive processes. Furthermore, congruency demonstrates how well a measure correlates with other measures of the same construct (Furlong & Oancea, 2005; Kane, 2013). Essentially, it shows that different variables measuring the same concept are related in expected ways.

Methods of estimating validity

The methods of estimating validity are centered on the categories of validity evidence. For instance, several statistical tools have been developed for content validity (Choudrie & Dwivedi, 2005). Popular among them are the Lawshe (1975) method content validity ratio (CVR) and modified kappa statistic (K). To assess criterion evidence, the correlation coefficient of the criteria and predicator is estimated (Nitko, 2014). The coefficient gives an indication as whether there is a relationship between the scores and how well the predictor predicts or relate with the criteria. Another approach to check predictive validity is by the use of the expectancy table. It is a two-way table that allows criteria to be predicted from a score. Principal Component Analysis (PCA) is used for estimating the construct related evidence of validity. For discriminant and convergent validity, a factor analysis can be conducted utilizing principal component analysis (PCA) with varimax rotation method (Koh & Nam, 2005; Wee & Quazi, 2005).

Reliability theory

Educational decisions like any other decisions, are based on data. These data may come from both classroom and standardised test scores, classroom observations, parental reports, and many other sources. In using the data for decision making, one should know something about

the quality of the data. Here, high-quality data should be weighted more deeply in one's decision than poor-quality data. In principle, data should be reliable and the inferences one draws from them should be valid. Reliability is paramount in assessing individuals.

Reliability is extremely important because evidence of reliability is necessarily the first step in establishing the scientific acceptance and usefulness of a test. Reliability refers to the consistency or stability of a measurement (Segal & Coolidge, 2018; DeLuca & Klinger, 2010;). The reliability of a procedure or measure is the degree to which the procedure or measure yields consistent results (Benton & Li, 2017). Reliability is the extent to which test scores are not affected by chance factors. It is the extent to which the test taker's score does not depend on (a) the specific day and time of the test (as compared with other possible days and times of testing), (b) the specific questions or problems that were on the edition of the test that the test taker took (as compared with those on other editions), and (c) the specific raters who rated the test taker's responses (if the scoring process involved any judgment) (Livingston, 2018; Moss, 2003;). In achievement measures, reliability refers to the amount of confidence one has that the score the student obtained is his or her actual level of achievement (Brookhart, 2003)

In physical measurement, one can ordinarily obtain very reliable measures Ebel and Frisbie (as cited in Liaquat, Asif, Siraji & Maroof, 2012) which include measures in education and psychology. This is true mainly for three basic reasons:

1. Physical characteristics can usually be measured directly rather than indirectly.
2. The instruments used to obtain the measures are quite precise.
3. The traits or characteristics being measured are relatively stable.

Components of reliability

The components of reliability are differentiation of assessment types, communication of expectations and systematic assessment procedures. Teachers' conception of reliability dictates that they differentiate their classroom assessments. This will mean that teachers will use different assessment methods to monitor student learning to provide continuous feedback. Black and Wiliam (1998) emphasize that differentiation of assessment type can enhance students learning to produce reliable outcomes. Scriven (1967) noted the use of statistical procedures such as Cronbach's alpha and test-retest reliability to produce reliability outcomes. Test results will accurately reflect students' knowledge and skills to make results reliable. Also, teachers' clarity of expectations communicated to students can significantly impact the

reliability of assessment results. Effective communication of expectations often involves the use of detailed rubrics and scoring guides. Andrade (2000) indicates that when students are aware of the criteria on which they will be assessed, they are more likely to perform reliably and meet those expectations. This transparency enhances the reliability of assessments by aligning student efforts with educator expectations. Consistent and constructive feedback is crucial in communicating expectations. Hattie and Timperley (2007) note that timely and specific feedback helps students understand where they stand concerning expectations, thus enhancing their performance reliability. Systematic assessment procedures are essential for ensuring reliability across different assessment types. According to Messick (1989), the reliability of assessments can be improved by standardizing procedures across all phases of the assessment process. This includes training assessors, utilizing consistent assessment tools, and adhering to predetermined scoring methods. Systematic approaches to analyzing and reporting assessment data also play a vital role in reliability. Consistent data analysis techniques, such as item response theory (IRT), can provide insights into the reliability of assessment instruments (Embretson & Reise, 2000).

Methods of Estimating Reliability

The methods used to estimate reliability differ in that they consider different sources of error (Liaquat, Asif, Siraji & Maroof, 2012). Reliability and validity are related. It should be noted that reliability is necessary but not enough evidence of validity. Many different approaches can be used to estimate reliability, but the more common ones used are Test-rest method, Equivalent forms method and Measures of internal consistency. The test-retest method is a measure of stability and considers scores of students over a period of time. The same test is given to a group of students twice within an interval ranging from several minutes to years. The scores on the two administrations are correlated and the result is the estimate of the reliability of the test (Etsey, 2012). Teachers' knowledge of reliability aids them to know the true performance of their students.

In contrast to the test-retest method, the equivalent-forms estimate of reliability is obtained by giving two forms (with equal content, means, and variances) of a test to the same group on the same day and correlating these results. With this procedure, one determines how confident one can generalize a person's score to what he would receive if he took a test composed of similar but different questions. In this case, instead of measuring changes from one time to another, one measures changes due to the specificity of knowledge. Teachers'

knowledge on reliability helps them to measure changes in specificity of knowledge and to construct test items of the same difficulty level and of the same format.

Another method of estimating reliability is the split-half method. Here, a single test is given to the students. The test is then divided into two halves for scoring. The two scores for each student are correlated to obtain the estimate of reliability. The test can be split into two halves in several ways. These include using odd-even numbered items, and first and half-second half. The split-half method aids practicing teachers to understand the true performance of the student in the sense that, the student is assessed twice with one assessment task at the same time.

The Cronbach alpha is another method of estimating internal consistency (Nitko, 2001). Cronbach alpha is the average split-half correlation based on all possible divisions of a test into two parts. This internal consistency is used when test items are scored pass-fail or when more than one point is awarded for a correct response (Salvia & Yesseldyke, 2001). Kuder and Richardson (KR-20 & 21) is more restricted method of estimating a test's reliability which is based on the average correlation between all possible split halves (Nitko, 2001). Under KR-20 and 21, test items are scored dichotomously (that is, items that can be scored only right and wrong). Nitko (2001) indicated that internal consistency is founded on the idea that the consistency with which students respond from one assessment task to the next provide the basis for estimating the reliability coefficient for the total scores. Teachers' knowledge in reliability, generally, aids them to effectively assess their students to see how consistent assessment results are. But the question is, are practicing teaches in senior high schools able to use competencies and skills gained in reliability to assess their students?

Related work on teachers' conceptions of validity and reliability

A good test should be valid and reliable. The validity of a test is the extent to which it measures what it is supposed to measure and nothing else (Furwana, 2019). Shulman (2004) noted that every assessment starts with specifying exactly what is to be assessed. Teachers do this for students to set targets they aim for. Oduro-Okyireh (2008) posited that "the key to effective achievement testing is careful planning" (p. 15). It is during the planning stage that the purpose of the test must be determined. Most senior high school teachers took a one-semester course in assessment in education to acquire knowledge of validity and reliability. These teachers are to transfer the knowledge gained into their classroom assessment practices. Teachers would be

able to assess students well and make sound decisions based on the results students produce (William, 2010).

Teachers' knowledge about reliability and validity would provide them with skills and competencies to develop a table of specification in planning their assessment. Here, teachers determine what to be tested or measured. Etsey (2004) noted that assessment helps teachers to determine the chapters or units of the course content that the test should cover as well as the knowledge, skills or attitudes to be measured. Instructional objectives need to be defined in terms of student behaviours and linked to what has been stressed in class. A test plan made up of a table of specifications should be made. The table of specifications matches the course content with the instructional objectives (Etsey, 2004). Teachers also must be able to articulate clear learning objectives that are congruent with both content and depth of thinking implied by standards and curriculum goals, in such a way that they are attainable and assessable. A competent teacher defines and describes the knowledge and skills students need to learn in clear, attainable, and assessable ways. There must be targets that the students can envision for themselves and achieve. Targets must be assessable so that both students and teachers will know whether and to what degree they have achieved the targets (Stiggin, 2008; Taylor & Nolen, 2005).

Coupled with the above, knowledge about validity and reliability helps practising teachers to have a collection of strategies for communicating to students what achievement of a learning is. It also equips practising teachers to share and communicate the learning objectives with students, parents and colleagues as well. This includes telling, showing, and having students discover for themselves the learning objectives in the content areas they teach (Moss & Brookhart, 2009). Thus, to enhance student learning, the assessment should be aligned with intended learning outcomes (Biggs, 2003). It is also important that the assessment is transparent (Segers, Dochy & Gijbels, 2010) and that students are aware of the assessment criteria for different grades (Prosser, 2014; Sadler, 2005; Handley & Williams, 2011). Clearly set out criteria also help teachers to assess students' answers validly and reliably (Yorke, Bridges & Woolf, 2000).

Importantly, good students can detect the learning intentions or objectives without being taught by the teacher. However, for the sake of other students, the teacher needs to understand how to create and use activities, assignments, and assessments that embody the learning intentions at different levels of student understanding (Ruiz-Primo, Furtak, Ayala, Yin & Shavelson, 2010) without confusing the activities with the learning intentions themselves.

According to Hattie and Timperley (2007), assessment provides teachers with the skills for effective and useful feedback on student work. They further elaborated that effective teacher feedback on students' work is usually descriptive. Effective feedback gives students information about their work against the standards. Kluger and DeNisi (1996) indicated that effective feedback is at an intermediate level of generality so that students can identify specific improvements that are needed. Campbell and Collins (2007) found that assessment provides teachers with the knowledge to communicate the interpretations of assessment results. Test construction skills include the competencies needed for developing quality test items based on stipulated principles of test construction (Ali, as cited in Agu, Onyekuba & Anyichie, 2013). These competencies are outlined by Ujah (2001) as: objectivity, communicative, item validation skills and skills for applying appropriate strategies for ascertaining the reliability of test instruments. Silker (2003) asserted that, skill in test item construction enables a teacher to construct test items with precision, appropriateness of language-use, objectivity and good grading scales.

Agu, Onyekuba and Anyichie (2013) pointed out that teachers need not be experts in educational measurement and evaluation to construct valid and reliable tests, but there are some basic test construction skills which every teacher ought to possess to construct quality test items. These skills help teachers to structure test items to elicit clear and concise answers from students; construct test items that will be appropriate for learners of different ages, abilities, and gender and set test items so that students finish within time and do not grow scared of tests (Ali, 1999). Knowledge about validity and reliability provides teachers with skills and competencies about the quality of test items. Teachers can ascertain how test items are closely linked with their ability to provide the kind of information needed regarding students' performances. A well written test item allows the teacher to accurately and consistently measure students' mastery of specific content taught in class (Agu, Onyekuba & Anyichie, 2013). Results of such tests allow teachers to measure, to some degree, how effective their instruction has been. To construct good test items, therefore, classroom teachers should possess the competencies in the following areas as outlined by Chidolue (1999, p. 36): determining the purpose of each testing exercise; stating specific, measurable educational objectives; making good content outline; preparing test plan which will guide item construction; choosing appropriate test item formats; constructing clear, precise and unambiguous items; constructing items that focus the attention of a group of students, often with widely varying background experiences, on a single idea; constructing items with appropriate difficulty and discriminative

indices; developing marking guide suited for the test; performing item analysis of their test items; developing tests that are economical in time and money; giving clear directions on how the test should be administered and taken; reviewing the test to correct any errors made during item construction. Koksal (2004) also noted that in test construction, it is essential that the teacher asks the following questions: Is the task clear? Is there more than one possible correct response (objective items)? Can test-takers arrive at the correct response without having the skill supposedly being tested? Do test-takers have enough time to perform the task(s)? Campbell and Collins (2007) found that understanding how to use assessment for different kinds of decisions was considered an essential assessment skill, according to a consensus of assessment textbooks authors.

Several studies have examined teachers' beliefs, perception and practices of classroom assessment. For instance, in Turkey, a study by Özdemir-Yilmazer and Özkan (2017) investigated tertiary-level English language teachers' perceptions and practices of speaking assessment. The data was collected via repertory grids, semi-structured interviews, and classroom observation. Focus grid analysis was conducted to analyze and interpret the data. The findings of the study reveal that the participants do have individualized beliefs regarding assessment of speaking, which necessitates training for teachers on classroom assessment to raise their awareness regarding commonality of practice in in-class speaking assessment. In another study by DeLuca, et al., (2019), explored the relationship between teacher candidates' approaches to assessment and their mindset towards learning at the end of their teacher education. Data were collected from 396 teacher candidates. Significant trends between mindset and teachers' approaches to assessment were observed, specifically, in regard to teacher candidates' mindset and approaches to assessment purpose and fairness. The result revealed that mindset is a component of teacher assessment literacy and argues that assessment mindset involves the specific linkage between teachers' mindsets toward learning and their approach to assessment. Both DeLuca, et al. (2019) and Özdemir-Yilmazer and Özkan (2017) focused on the mind of the teachers (mindset and perception) arrive at a convergent finding that the teachers have their unique beliefs on which in tend influence theirs practice of classroom. Even though Özdemir-Yilmazer and Özkan (2017) was not specific on the belief being assessed as in the case of DeLuca, et al. (2019), fairness is central to classroom assessment and cannot be omitted (Wallace & Qin, 2021). Similarly, Koloi-Keaikitse (2017) assessed teacher perceived skill in classroom assessment practices. Data were collected from a sample of (N = 691) teachers selected from government primary, junior secondary, and senior

secondary schools in Botswana. Item response theory models were used to identify teacher response on items that measured their self-perceived skill in classroom assessment practices. Results of the study showed that generally teachers felt more skilled in test construction than other practices, such as using classroom assessment results to make informed decisions in their teaching and learning process. In relation to other studies (DeLuca, et al., 2019; Özdemir-Yilmazer & Özkan (2017), skills demonstrated by the teachers could be inferred to a reflection of their beliefs about assessment.

Asamoah et al. (2024) examined the perception of secondary school teachers in Ghana and Brunei on assessment and how their perceptions influence how they practice classroom assessment. A total of 431 secondary school teachers from Ghana (n = 308) and Brunei (n = 123) were engaged and a Semi-structured interviews conducted with six Bruneian and eight Ghanaian teachers to understand how they practiced assessment amid their school assessment climates. Teachers were found to practice assessment and perceive their school assessment climates differently. Teachers reported positive perceptions of the nature of school assessment climates that existed in their schools. However, the examination-oriented climates that prioritised summative assessment compelled teachers to practice assessment against their beliefs. This resulted in academic dishonesty, excessive pressure, and mental health problems among teachers and students. Kissi, et al., (2023) and Oppong Frimpong and Osei (2021), on the other hand, explored the relationship between multiple choice test construction competence and the quality of multiple-choice tests among senior high school teachers in Ghana. Participants responded to self-designed questionnaire developed to assess teachers' multiple-choice items construction competencies, competence content validity, item "options" handling, and test items assembling. The result revealed that the teachers perceived more competence in ensuring content validity, followed by test item assembling, and handling of "options" (that is, alternatives) of the test items. The study also found serious problems with copies of multiple-choice items teachers have constructed for the students. Whilst Asamoah, et al., (2024) focused on the conceptualization of the assessment, Oppong Frimpong and Osei (2021) and Kissi, et al., (2023) focused on what the teachers were practicing as evidence of their conceptualization of assessment.

Methods

Research Design and Approach

The quantitative descriptive survey design was used for the study. Descriptive survey design seeks to explore and describe events as they occur (Creswell, 2014). Quantitative research is a systematic and objective process of using numerical data from only a selected sub-group of a population to generalise the findings to the population (Cohen, Manion & Morrison, 2007; Leedy & Omrod, 2010; Maree and Pietersen (2016). The choice of the quantitative descriptive survey was appropriate to accurately obtain objective data, statistically analyse data to answer research question, and systematically explore and quantify senior high school teachers' conceptualization of validity and reliability theories in classroom assessments.

Participants

The target population for the study involves all core subject teachers in public Senior High School in the Kumasi Metropolis of the Ashanti Region of Ghana. Currently, there are nineteen (19) accredited public Senior High School in the metropolis with 1045 core subject teachers (Ashanti Regional Education Office, 2024). The accessible population comprised teachers with not less than a year teaching experience in Core Mathematics, English Language, Social Studies, and Integrated Science in five selected schools namely Yaa Asantewaa Senior High School, Adventist Senior High, Asanteman Senior High School, Osei Tutu Senior High School and Islamic Senior High School. Kothari (2004) noted that it is from the accessible population that a sample is selected for a study. The distribution of the population among the five selected SHS is shown in Table 1.

Table 1: Distribution of Population Based on Core Subject Teachers

School	**Teachers**
1. Yaa Asantewaa Senior High School	62
2. Adventist Senior High	45
3. Asanteman Senior High School	43
4. Osei Tutu Senior High School	55
5. Armed Forces Senior High School	53
Total	**258**

The purposive sampling method was used to include 200 core subject teachers with not less than a year of teaching experience for the study. All the 258 teachers were to be considered for the study, but at the time of data collection, only 200 teachers participated in the study. Core

subject teachers were selected for the study because core subjects were common in all schools that participated in the study.

Research instruments

The data collection instrument was an adapted questionnaire. The questionnaire was made up of closed-ended items. The original questionnaire has two parts, namely, (a) validity with three subscales and (b) reliability with three subscales. The subscales under validity were (a) relevance with 3 items, (b) thinking processes with 4 items, and (c) congruency with 3 items. Also, the subscales under reliability were (a) differentiation of assessment type with 4 items, (b) communication of expectations with 2 items, and (c) systematic assessment procedures with 3 items. The adapted questionnaire used for the current study had a total of 30 items. Items under the various subscales were modified, and some items were added to the original items to ensure the validity of the study. Under validity, for relevance, items 3 and 4 were added; level of thinking processes, item 4 was added; congruency, items 1 and 2 were added. Under reliability, for differentiation of assessment type, items 3 were added; communication of expectations, items 3, 4, and 5 were added; and systematic assessment type, items 3 and 4 were added. Items on the questionnaire were multiple-scored on a four-point Likert-type scale as 1-Never, 2-Not Often, 3-Often, 4-Always. Items were pilot tested on 50 core subject teachers in different senior high schools. After the pilot testing, items 1 and 2 under communication of expectations, which originally read "test items clearly represent the learning outcomes and clearly communicate the expectation of assessment," were reworded as "I ensure that test items clearly represent the learning outcomes" and "there are sets of illustrative student works that serve as examples for expectation across all achievement levels" were replaced with "I ensure that clear learning objectives are communicated to students." The Cronbach's coefficient alpha for all items on the questionnaire was 0.85.

Data collection procedures

Informed consent was sought from participants by explaining the purpose of the study to them. Anonymity of respondents was highly considered in the study. This gave the participants the opportunity to have their identity concealed. Neither names nor any identifiable information from respondents were taken. On the issue of confidentiality, effort was made to maintain confidentiality of the responses of the participants. Participants were told that their responses would be kept confidential and that no one known to them would have access to the information provided. The questionnaire was administered personally by the researchers to all the 258

teachers involved in the study in the five selected senior high schools in the Kumasi metropolis. A period of two weeks was used to travel to all the schools to administer the questionnaire. Out of the 258 questionnaires administered, 200 representing 77.5% were retrieved.

Data analysis method

Data were analysed using means and standard deviations to answer the research question. The criterion in Table 1 was calculated by dividing the range (3) by the number of categories (4), giving 0.75. Then the criteria are 1.00-1.74: Never; 1.75-2.49: Not Often, 2.50 -3.24 = Often, and 3.25-4.00 = Always (Boone & Boone, 2012; Sozen & Guven, 2019; Adu Gyamfi & Yeboah, 2022).

Results and Discussion

Research question

What are senior high school teachers understanding of validity and reliability in classroom assessment? The research question sought to find out teachers understanding of validity and reliability in classroom assessment. Means and standard deviations of the responses to the items were used to answer the research question. A mean of means of 2.5 and above indicates positive conceptualization whiles a mean of means of less than 2.5 indicates a negative conceptualization. The result is presented in Table 2.

Table 2: Analysis of Results of Teachers Conceptualization of Validity and Reliability (n=200)

S/N	Item	M	SD	Remarks
	Validity			
	Relevance			
1.	I properly align the content of test with learning outcomes of the curriculum unit.	3.61	.57	Always
2.	I properly align the content of test with content taught throughout the curriculum unit.	3.57	.54	Always
3.	I properly align the content of test in other to motivate student learning	3.48	.64	Always
4.	I properly align the content of test in other to improve student attitude towards a subject	3.47	.67	Always
5.	I properly align the content of test with national standards	3.29	.74	Always
	Mean of means	**3.48**	**.63**	
	Level of thinking processes			
1.	I adequately match the difficulty of test to instructional content taught in the classroom.	3.38	.72	Always
2.	I adequately match the difficulty of test to students' ability level	3.23	.75	Often
3.	I adequately match the difficulty of test to ability represented by students during class instruction.	3.03	.77	Often
4.	I adequately match the difficulty level of the test based on the selected assessment technique	2.99	.80	Often
5.	I adequately distribute the range of difficulty items within test to students	2.85	.91	Often
	Mean of means	**3.09**	**.79**	
	Congruency			
1.	Similar test having the ability to measure expected response(s) from students.	3.04	.84	Often

2.	Test having the ability to be positively correlated with previous test.	3.03	.85	Often
3.	Students bring with them prior experiences that could affect outcome of test.	2.82	.80	Often
4.	Students' outcome of test generally matches teacher's expected outcome.	2.71	.85	Often
5.	There are large groups of students who are unexpectedly overachieving or underachieving on test.	2.71	.80	Often
	Mean of means	**2.86**	**.82**	
	Overall mean	**3.14**		
	Reliability			
	Differentiation of assessment types			
1.	I construct test items to solicit different types of information in order to make judgement of what students know.	3.44	.75	Always
2.	There is enough information to make an accurate judgement about student's knowledge, skills and abilities being assessed.	3.38	.73	Always
3.	I construct test items to reflect the varied performance level of the students	3.28	.80	Always
4.	I use other types of tests to elicit knowledge, skills and abilities of students.	3.26	.70	Always
5.	I am confident that students would respond to test items in the same way when they are tested again.	2.99	.70	Often
	Mean of means	**3.27**	**.73**	
	Communication of expectations			
1.	I ensure that test items clearly represent the learning outcomes	3.63	.62	Always
2.	I ensure that clear learning objectives are communicated to student	3.62	.68	Always
3.	I ensure that feedback on students works are given on time	3.58	.53	Always

4.	I ensure that the knowledge and skills students need are communicated to them to learn in clear, attainable, and assessable ways	3.49	.64	Always
5.	I ensure that test items clearly communicate the expectation of assessment	3.48	.62	Always
	Mean of means	**3.56**	**.61**	
	Systematic assessment procedures			
1.	I make students aware of test and prepare them for test before they take it.	3.69	.59	Always
2.	I ensure that students are comfortable with test process.	3.59	.58	Always
3.	I ensure that students with disabilities are well catered for during the test process	3.45	.75	Always
4.	I ensure that system for writing new test items are fully functioning	3.38	.71	Always
5.	I ensure that the testing procedure that does not affect the ability of students to demonstrate their knowledge, skills, and abilities.	3.27	.73	Always
	Mean of means	**3.47**	**.67**	
	Overall mean	**3.43**		

NB: M= Mean and SD = Standard Deviation.

On the issue of conceptualization of validity, the results shows that generally, teachers' conceptualization is positive (M= 3.14), specifically, the results from the table show that teachers always conceptualized validity as relevant *(M=3.48, SD=.63)*. It was found that teachers always properly aligned the content of test with learning outcomes of the curriculum unit *(M=3.61, SD=.57)*, properly aligned the content of test with content taught throughout the curriculum unit *(M=3.57, SD=54)*, properly aligned the content of test in other to motivate student learning *(M=3.48, SD=.64)* and properly aligned the content of test in other to improve student attitude towards a subject *(M=3.47, SD=.67)*.

It was observed that teachers often conceptualized validity as level of thinking processes *(M=3.09, SD=.79)*. The study showed that teachers always adequately matched the difficulty of test to instructional content taught in the classroom *(M=3.38, SD.72)*. It was further found that teachers often adequately matched the difficulty of test to students' ability level *(M=3.23, SD=.75)*, adequately matched the difficulty of test to ability represented by students during class instruction *(M=3.03, SD=.77)* and adequately matched the difficulty level of the test based on the selected assessment technique *(M=2.99, SD=.80)*. Teachers often conceptualized validity as congruent *(M=2.86, SD=.82)*. For example, teachers often gave similar tests having the ability to measure expected response(s) from students *(M=3.04, SD=.84)* and gave tests having the ability to be positively correlated with previous test *(M=3.03, SD=.85)*. Teachers often noted that students bring with them prior experiences that could affect outcome of test *(M=2.82, SD=.80)*. It was seen that teachers often realized that students' outcome of test generally matched teacher's expected outcome *(M=2.71, SD=.85)*.

On reliability, the results found that the teachers' conceptualization is positive. It was found that teachers always differentiated assessment types *(M=3.27, SD=.73)*. It was revealed that teachers always constructed test items to solicit different types of information to make judgement of what students know *(M=3.44, SD=.75)*, there was enough information to make an accurate judgement about student's knowledge, skills and abilities being assessed *(M=3.38, SD=.73)* and constructed test items to reflect the varied performance level of the students *(M=3.28, SD=.80)*.

The results further showed that teacher's conceptualization of reliability always reflected in their communication of expectations *(M=3.56, SD=.61)*. It was seen that teachers always ensured that test items clearly represented the learning outcomes *(M=3.63, SD=.62)*, ensured that clear learning objectives are communicated to student *(M=3.62, SD=.68)* and ensured that feedback on students works are given on time *(M=3.58, SD=.53)*. The study found

that teachers had a good conceptualization of reliability by always ensuring systematic assessment procedures *(M=3.47, SD=.67)*. It was observed that teachers always made students aware of test and prepare them for test before they take it *(M=3.69, SD=.59)*, ensured that students are comfortable with test process *(M=3.59, SD=.58)*, ensured that students with disabilities are well catered for during the test process *(M=3.45, SD=.75)* and ensured that system for writing new test items are fully functioning *(M=3.38, SD=.71)*. The results of the study revealed that teachers have positive conceptualization of validity and reliability.

Discussion

The findings of the study revealed that, generally, teachers involved in the study have high conceptualization of validity as used in classroom assessment. The findings showed that teachers always aligned the content of test with learning outcomes of the curriculum unit and the content of test with content taught throughout the curriculum unit. This finding is relevant in the sense that students who would pay particular attention to what is taught based on the curriculum would find test quite moderate and easy to respond to. Students will have the opportunity to learn and know what knowledge and skills they need to acquire to conduct themselves well during testing. Teachers' conceptualization of testing principle which stipulates that test should be anchored on taught curriculum unit would enable students to perform well during testing. The findings lend support to findings of Stiggin (2008); Taylor and Nolen (2005) who reported that teachers conceptualization of validity implies that teachers articulate clear learning objectives that are congruent with both content and depth of thinking implied by standards and curriculum goals, in such a way that they are attainable and assessable. Similarly, Yeboah (2019) indicated that teacher's conceptualization of validity would suggests that sample of items, tasks, or questions on a test is aligned with curriculum unit. Teachers select adequate content areas that are aligned with curriculum units when assessing their students. The findings of Stiggin (2008), Taylor and Nolen (2005) and Yeboah (2019) are all supported by the findings of this study that teachers have positive conceptualization of validity which influences how they practice classroom assessment. In addition, the findings Asamoah, et al (2024) is also supported by the findings of this study. Asamoah, et al (2024) reported that teachers have positive perceptions of the nature of school assessment climates that existed in their schools. Also, Kissi, et al (2023) and Oppong Frimpong and Osei (2021), revealed that the teachers perceived more competence in ensuring content validity, followed by test item assembling, and handling of "options" (that is,

alternatives) of the test items. Quite a number of studies share the same findings as this study that teachers have high conceptualization of validity of assessment.

However, findings from this study are contradictory to findings of Özdemir-Yilmazer and Özkan (2017) and DeLuca, et al (2019). For instance, Özdemir-Yilmazer and Özkan (2017) reveal that the participants do have individualized beliefs regarding assessment and that much training is needed for the teachers. The poor practice of classroom assessment thus emanates from their poor conceptualization of the concept of validity. Also, DeLuca, et al (2019), found a significant trends between mindset and teachers' approaches to assessment were observed, specifically, in regard to teacher candidates' mindset and approaches to assessment purpose and fairness. Similarly, Asamoah, et al (2024) revealed summative assessment compelled teachers to practice assessment against their beliefs. This is indication of poor conceptualization of assessment. The teacher only perceives validity as fairness which suggests a low conceptualization of validity.

The findings showed that Ghanaian teachers often conceptualized validity as a level of thinking processes by always adequately matching the difficulty of test to instructional content taught in the classroom, adequately matching the difficulty of test to students' ability level, and adequately matching the difficulty of test to ability represented by students during class instruction. Sadler (2009) noted teachers' understanding of validity in the classroom to reflect a situation where the difficulty of test matches instructional content taught in the classroom and students ability level. Sadler furthered that this would result in "grade integrity" which is about the extent to which grades correspond with the quality, breadth, and depth of students' academic achievement. Supporting the assertion of Sadler (2009), William (2010) opined that teachers conceptualization of validity theory implies teachers matching the difficulty of test to ability represented by students during class instruction. This would help teachers to make sound decisions based on the results obtained by students on a test.

Teachers often conceptualized validity as congruent by often giving similar tests having the ability to measure expected response(s) from students and giving tests that could positively correlate with previous tests. This finding implies that Ghanaian senior high school teachers involved in the study have appreciable knowledge about validity in classroom assessment. Teachers' ability to give similar tests to measure the same expectation from students would suggest they have evidence to predict students' future behaviour. The findings corroborate findings of Furwana (2019) who indicated that teacher's conceptualization of validity suggest that teachers give similar test to measure same response(s) from students. According to

Furwana, validity is essentially about evidence to support an assertion. Teachers often noted that students bring with them prior experiences that could affect outcome of test. Indeed, for teachers at the senior high school level to conceptualize validity as having evidence to support students current or future performance is good. This probably indicate how teachers have understood and ensure validity in classroom assessment.

The findings showed that Ghanian teachers generally have high conceptualization of reliability in classroom assessment. This is reflected in differentiated, communicated, and ensured systematic assessment procedures. Teachers always constructed test items to solicit different types of information to make judgement of what students know and there was enough information to make an accurate judgement about student's knowledge, skills and abilities being assessed. Ruiz-Primo, Furtak, Ayala, Yin and Shavelson (2010) noted that teachers understanding of reliability in classroom assessment implies that teacher differentiated assessment types to obtain appropriate information and teachers have ample information to make judgements about students' academic achievement. Another study that shares similar findings is Koloi-Keaikitse (2017). Koloi-Keaikitse (2017) found that generally teachers felt more skilled in test construction and practices such as using classroom assessment results to make informed decisions in their teaching and learning process. This result implies teacher high conceptualization of reliability as found in this study.

The findings show that Ghanian teachers conceptualized reliability to reflect their communication of expectations. This suggests that teachers always ensured that test items clearly represented the learning outcomes, and that clear learning objectives were communicated to student, and feedback on students works are given on time. The finding is consistent with findings of Hattie and Timperley (2007) who opined that teachers understanding of reliability implies that teachers have the skills for effective and useful feedback on student work. He further elaborated that effective teacher feedback on student work is usually descriptive. Effective feedback gives students information about their work against the standards. Kluger and DeNisi (1996) noted that effective feedback is at an intermediate level of generality so that students can identify specific improvements that are needed. Campbell and Collins (2007), which share similar findings, found that teachers' appreciation of reliability suggest effective communication of students' results.

The study found that Ghanian teachers conceptualised reliability by ensuring systematic assessment procedures. It was observed that teachers always made students aware of tests and

prepared them for tests before they took them. This tends towards reducing the anxiety level of students during examinations. The findings corroborate the findings of Oduro-Okyireh (2008) who noted that "the key to effective achievement testing is careful planning and preparation" (p. 15). Oduro-Okyireh added that students' preparation by teachers during testing is essential for students' academic success. The study showed that teachers ensured that students were comfortable with the test processes. This helps students to acquire skills and competencies to go through test processes smoothly.

Conclusions

The study explored Ghanaian senior high school teachers' conception of validity and reliability in classroom assessment. The teachers generally indicated their understanding of validity as relevant and encompassing thinking processes and as congruent. Teachers' conceptualization of validity helped them to align the content of the test with learning outcomes of the curriculum unit, adequately match the difficulty of test to instructional content taught in the classroom and give similar tests that can measure expected responses from students.

Regarding reliability, teachers generally indicated their understanding of reliability by differentiating assessment types and communicating expectations of the test to learners. The teachers' understanding of reliability enabled them to construct test items to solicit different types of information to make judgments of what students know, ensured that clear learning objectives are communicated to students and ensured that students are comfortable with the test process.

Recommendations

Teachers seem to have a moderate understanding of congruent validity, which has to do with a test's relationship with a known valid and reliable measure of the same construct. Ghanaian senior high school heads and assessment experts should regularly train teachers in congruent validity. This will give teachers more insight into issues of congruent validity. Teachers had relatively low confidence in students' consistent responses to test items on retesting. To boost teacher confidence in students, increased teacher training and student learning are advised.

Limitations

The major limitation of the study was the unenthusiastic teacher responses toward this research work evident in only 77.51% of the 258-questionnaire distributed filled and returned. In the ideal situation, a nationwide study is required. This would have given much confidence to any

generalisations made. Also, not all the subjects taught in the SHS were included in the study. This is because we wanted to concentrate mainly on core subjects, namely, core mathematics, English language, social studies and integrated science. The smaller number of subjects taught at the senior high schools that were used for the study requires caution when generalising the results over all senior high schools in Ghana.

References

Adiyaa, O., Osei-Poku, P., & Essel, H. B. (2022). Validating visual art educators' test construction skills at the senior high school in the Sekyere South District of Ashanti Region, Ghana. *Journal of Educational Science*, *1*(1), 1–8. https://doi.org/10.56388/edu220827

Adu Gyamfi, B., & Yeboah, A. (2022). Readiness of regular education teachers towards inclusive education in Ghana. *American Journal of Educational Research, 10(6), 420-431.doi: 10.12691/education-10-6-8.*

Agu, N. N., Onyekuba, C., & Anyichie, A. C. (2013). Measuring teachers' competencies in constructing classroom-based tests in Nigerian secondary schools: Need for a test construction skill inventory. *Educational Research and Reviews, 8*(8), 431-439.

Ali, A. A. (1999). *Basic research skills in education.* Enugu: Orient Printing and Publishing.

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.

Andrade, H. G. (2000). Student self-assessment: A key to strengthening student motivation and ownership of learning. *Theory Into Practice*, 39(1), 8-13.

Asamoah, D, Shahrill, M., & Abdul-Latif, S. N. (2024). Teachers' perceptions of school assessment climate and realities of assessment practices in two educational contexts. *Frontiers in Education* *9*:1-23 1278187. doi: 10.3389/feduc.2024.1278187

Asamoah, D., Shahrill, M., & Abdul Latif, S. N. (2023). Towards developing the classroom

assessment literacy: Exploring teachers' approaches to assessment across cultures. *Cogent Education*, *10*(2). https://doi.org/10.1080/2331186X.2023.2280301

Benton, S. L., & Li, D. (2017). *IDEA student ratings of instruction and rsvp* (IDEA Paper No. 66). Manhattan, KS: The IDEA Center.

Biggs, J. (2003). Teaching for quality learning at university. In What the student does (2nd ed.). Buckingham: *The Society for Research into Higher Education*. Open University Press.

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

Black, P., & Wiliam, D. (2012). *The reliability of assessments. In Gardner, J. (ed.) Assessment and Learning. (2nd ed.) London: SAGE Publications, 243–63.*

Boone, H. N., & Boone, D.A. (2012) Analyzing likert data. *The Journal of Extension, 50*, 1-5. https://joe.org/joe/2012april/tt2.php

Brookhart, S. M. (2023). Developing measurement theory for classroom assessment purposes and uses. *Educational Measurement: Issues and Practice, 22*(4), 5–12.

Campbell, C., & Collins, V. L. (2007). Identifying essential topics in general and special education assessment textbooks. *Educational Measurement: Issues and Practices, 26* (1), 9-18.

Chidolue, M. (1999). Introduction to tests and measurements. *Unpublished mimeograph on measurement and evaluation*. Awka, Nnamdi Azikiwe University.

Cohen, L., Manion, K., & Morrison, L. (2007). *Research methods in education*, (6th ed.). New Yorke: Routeledge Taylor & Francis Group.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches.* Sage Publications.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods*

research. . Sage Publications.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory.* USA: Lengage Learning

Darling-Hammond, L., Herman, J., Pellegrino, J., & Abedi, J. (2013). *Criteria for high-quality assessment*. Stanford, CA: Stanford Center for Opportunity Policy in Education

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419-438. https://doi.org/10.1080/0969594X.2010.516643

DeLuca, C., Coombs, A., & LaPointe-McEwan, D. (2019). Assessment mindset: Exploring the relationship between teacher mindset and approaches to classroom assessment. Studies in Educational Evaluation, 61, 159-169

Dingel, M., & Wei, W. (2014). Influences on peer evaluation in a group project: An exploration of leadership, demographics and course performance. *Assessment & Evaluation in Higher Education, 39*(6), pp. 729-742

Downing, M. S. (2005). The effects of violating standard item writing principles on tests and students: The consequences of using flawed test items on achievement examinations in medical education. *Advances in Health Sciences Education*, *10*(2):133-43, DOI:10.1007/s10459-004-4019-5

Earle, S. (2020). Balancing the demands of validity and reliability in practice: Case study of a changing system of primary science summative assessment. *London Review of Education, 18*(2): 221–235. https://doi.org/10.14324/LRE.18.2.06

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Psychology Press.

Etsey, Y. K. A. (2004). *Educational measurement and evaluation.* Lecture notes on EPS 203. Unpublished document, University of Cape Coast, Ghana.

Etsey, Y. K. A. (2012). *Assessment in education*. University of Cape Coast: Unpublished.

Fischer, J., Bearman, M., Boud, D., & Tai, J. (2023). How does assessment drive learning? A focus on students' development of evaluative judgement. *Assessment & Evaluation in Higher Education*, *49*(2), 233–245. https://doi.org/10.1080/02602938.2023.2206986

Fraenkel, J. R. & Wallen, N. E. (2009). *How to design and evaluate research in education.* Ninth ed. New York: McGraw-Hill

Frimpong, S., & Osei, D. (2021). Knowledge and practice of assessment: Empirical evidence from early childhood teachers in the Kumasi Metropolis, Ghana. *European Journal of Humanities and Social Sciences*, *1*(1), 10–18. https://doi.org/10.24018/ejsocial2021.1.1.5

Furlong, J., & Oancea, A. (2005). Assessing quality in applied and practice-based educational research: A framework for evaluation. Research Papers in Education, 20(4), 457-471.

Furwana, D. (2019). Validity and reliability of teacher-made english summative test at second grade of Vocational High School Palopo. *Journal lof Language and Literature 13*(2).

Girolamo, N., & Mans, C. (2019). Research Study Design. *Fowler's Zoo and Wild Animal Medicine Current Therapy, Volume 9*, 59–62. Retrieved on 20th September, 2021 from https://doi.org/10.1016/B978-0-323-55228-8.00011-4

Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education, 36*(1), 95-108. http://dx.doi.org/10.1080/02602930903201669

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 7*(7), 81-112.

Johnson, S. (2013). On the reliability of high-stakes teacher assessment. *Research Papers in Education, 28*(1), 91-105. DOI:10.1080/02671522.2012.754229

Kane, M. T. (2013). Validating measures of construct validity. Psychological Bulletin, 139(3), 737-765.

Kane, T. M., & Saskia, W. (2020). *Perspectives on the validity of classroom assessments. In book*: Classroom Assessment and Educational Measurement (pp.11-26), DOI:

10.4324/9780429507533-2.

Kelly, M. P., Feistman, R., Dodge, E., St. Rose, A., & Littenberg-Tobias, J. (2020). Exploring the dimensionality of self-perceived performance assessment literacy (PAL). Educational Assessment, Evaluation and Accountability, 32(4), 499 517.https://doi.org/10.1007/s11092-020-09343-7 *(PDF) A Content Validity Study for Vocational Teachers' Assessment Literacy Instrument (VoTAL)*. Available from: https://www.researchgate.net/publication/355088004_A_Content_Validity_Stud_y for Vocational_Teachers%27_Assessment_Literacy_Instrument_VoTAL [accessed Oct 02 2024].

Kissi, P., Baidoo-Anu, D., Anane, E., & Annan-Brew, R. K. (2023). Teachers' test construction competencies in examination oriented educational system: Exploring teachers' multiple-choice test construction competence. *Frontiers in Education*, 8, 1154592. https://doi.org/10.3389/feduc.2023.1154592

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 11*(9), 254-284.

Koksal, D. (2004). Assessing teachers' testing skills in Elt and enhancing their professional development through distance learning on the net. *Turkish Online J. Distance Educ., 5*(1), 122-131.

Koloi-Keaikitse, S. (2017). Assessment of teacher perceived skill in classroom assessment practices using IRT Models. *Cogent Education*, *4*(1). https://doi.org/10.1080/2331186X.2017.1281202

Kothari, C. R. (2004). *Research methodology: Methods and teaching.* New Delhi, India: New Age International Limited Publishers.

Leedy, D. P., & Ormrod, E. J. (2010). *Research: Planning and design* (9th ed.) Merril, Upper Saddle River: Pearson Education Inc.

Liaquat, H, Asif, J. M., Siraji, J., & Maroof, K. (2012). Development and Standardization of Intelligence Test for Children. *International Journal of Learning & Development,*

2(5), 190-202.

Livingston, A. S. (2018). *Test reliability basic concepts*. Princeton, New Jersey. Educational Service.

Maree, K., & Pietersen, J. (2016). Surveys and the use of questionnaires. In K. Maree (Ed.). *First steps in research* (pp. 173-190). Pretoria: Van Schaik Publishers.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-104). American Council on Education.

Mitchell, M. L., & Jolley, J. M. (2010). *Research design explained*. New York: Wadsworth Cengage Learning.

Moss, C. M., & Brookhart, S. M. (2009). *Advancing formative assessment. In every classroom*. Alexandria, VA: ASCD.

Moss, P. A. (2003). Reconceptualizing validity for classroom assessment. *Educational Measurement: Issues and Practice, 22*(4), 13–25.

Nitko, A. J. (2001). *Educational tests and measurements* (3rd ed.). Prentice-Hall, Inc. Upper Saddle River, New Jersey.

Oduro-Okyireh, G. (2008). *Testing practices of senior secondary school teachers in the Ashanti Region of Ghana*. Unpublished master's thesis. University of Cape Coast, Cape Coast, Ghana.

Oppong Frimpong, S. & Osei, D. (2021). Knowledge and practice of assessment: Empirical evidence from early childhood teachers in the Kumasi Metropolis, Ghana. EJ-SOCIAL, *European Journal of Humanities and Social Sciences, 1*(1), 10-18.doi: http://dx.doi.org/10.24018/ejsocial.2021.1.1.5

Özdemir-Yilmazer, M., & Özkan, Y. (2017) Speaking assessment perceptions and practices of English teachers at tertiary level in the Turkish Context. *Language Learning in Higher Education*, 7(2), 371-391.

Popham, W. J. (1997). Classroom assessment: What teachers need to know. Allyn & Bacon.

Popham, W. J. (2016). *Classroom assessment: What teachers need to know*. Pearson Education

Prosser, M. (2014). Perceptions of assessment standards and student learning. In C. Kreber, C.

Anderson, N. Entwistle, & J. McArthur (Eds.), *Advances and innovations in university assessment and feedback (pp. 114-128)*. Edinburgh: Edinburgh University Press. http://dx.doi.org/10.3366/edinburgh/9780748694549.003.0007

Quansah, F., Amoako, I., & Ankomah, F. (2019). Teachers' test construction skills in senior high schools in Ghana: Document Analysis. *International Journal of Assessment Tools in Education, 6*(1), 1–8 DOI: 10.21449/ijate.481164

Räisänen, M., Tuononen, T., Postareff, L., Hailikari, T., & Virtanen, V. (2016). Students' and teacher's experiences of the validity and reliability of assessment in a bioscience course. *Higher Education Studies, 6*(4), 181-189.

Ruiz-Primo, M. A., Furtak, E. M., Ayala, C., Yin, Y., & Shavelson, R. J. (2010). Formative assessment, motivation, and science learning. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 139-158). New York: Routledge.

Sadler, R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30, 175-194. http://dx.doi.org/10.1080/0260293042000264262

Sadler, R. (2009). Grade integrity and the representation of academic achievement. *Studies in Higher Education, 34*, 807-826.http://dx.doi.org/10.1080/03075070802706553

Salvia, J., & Ysseldyke, J. E. (2001). *Assessment in special and remedial education*. Boston: Houghton Mifflin.

Scriven, M. (1967). The methodology of evaluation. In R. E. Stake (Ed.), *Curriculum evaluation* (pp. 39-83). Rand McNally.

Segal, L. D., & Coolidge, L. F. (2018). *The SAGE encyclopedia of lifespan human development*: Reliability. Thousand Oaks, SAGE Publications, Inc.

Segers, M., Dochy, F., & Gijbels, D. (2010). Impact of assessment on students' learning strategies and implications for judging assessment quality. In P. Peterson, E. Baker, & B. McGraw (Eds.), *International Encyclopedia of Education (3rd ed.).* Oxford: Elsevier.http://dx.doi.org/10.1016/B978-0-08-044894-7.01625-0.

Shulman, L. S. (2004). *The wisdom of practice: Essays on teaching, learning, and learning to teach*. San Francisco: Jossey-Bass.

Silker, R. T. (2003). *Teachers and tests*. London: Basil Blackwell.

Sözen, E., & Güven, U. (2019). The effect of online assessments on students' attitudes towards undergraduate-level geography courses. International Education Studies, 12(10), 1-8

Stiggins, R. J. (2008). *Student-involved assessment for learning* (5th ed.). Upper Saddle River, NJ: Pearson Merrill Prentice Hall.

Stiggins, R. J., & Chappuis, J. (2012). *An introduction to student-involved assessment FOR Learning* (6th ed.). Boston: Pearson.

Taylor, C. S., & Nolen, S. B. (2005). *Classroom assessment: Supporting teaching and learning in real classroom*. Upper Saddle River, NJ: Pearson.

Tontus, O. H. (2020). *Concept of assessment and evaluation*. Glossary of Assessment of and Evaluation in Higher Education, 11-17.

Trochim, W. M. K. (2006). Introduction to Validity. *Social Research Methods*, retrieved from www.socialresearchmethods.net/kb/introval.php, September 9, 2016.

Ujah, E. U. (2001). *Development and validation of an introductory technology achievement test*. Unpublished M.Ed. thesis, University of Nigeria, Nsukka.

Wallace, M. P., & Qin, C. Y. (2021). Language classroom assessment fairness: Perceptions from students. LEARN Journal: Language Education and Acquisition Research Network, 14(1), 492-521.

Watson, E. (2023). *Defining assessment. Centre for teaching and learning*. University of Alberta

Wiggins, G. P. (1990). The case for authentic assessment. *Practical Assessment, Research, and Evaluation, 2*(2).

Wiliam, D. (2013). Assessment: The bridge between teaching and learning. *Voices from the Middle 21(2):15-20*

William, D. (2010). An integrative summary of the research literature and implications for a new theory of formative assessment. In H. L. Andrade & G. J. Cizek (Eds.), *Handbook of formative assessment* (pp. 18-40). New York: Routledge.

Wulandari, (2023). *Analysis of teacher assessments of student achievement in using online learning media at the tenth grade students of Ma Alraisiyah Mataram in academic year 2020/2021*. Thesis presented to the faculty of Teacher Training and Education Muhammadiyah University of Mataram.

Yeboah, A. (2018). Competencies and skills in the assessment course: The perspective of the graduate teacher in Ghana. *The International Journal of Humanities and Social Studies, 6*(7), 221-227.

Yeboah, A., Gyamfi A., & Isaac, N. S. (2019). Relevance of assessment course: a follow-up study of graduate teachers in Ghana. *Acad. J. Educ. Res. 7*(9), 299-307.

Yorke, M., Bridges, P., & Woolf, H. (2000). *Mark distributions and marking practices in UK higher education. Some challenging issues*. Active learning in higher education, 1, 7-27.http://dx.doi.org/10.1177/1469787400001001002