British Journal of Education
Vol.11, Issue 6, pp. 7-24, 2023
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

Content Validity Index of Basic Education Certificate Examination (BECE) Test Items in Basic Science and Technology (BST)

Dr. Ijeoma Joy Chikezie
Department of Teacher Education, National Institute for Nigerian Languages,
Aba, Abia State

Prof. Eme U. Joseph
Department of Educational Foundations, University of Uyo, Akwa Ibom State, Nigeria

doi: https://doi.org/10.37745/bje.2013/vol11n6724 Published May , 2023

Citation: Chikezie I.J. and Eme U. J. (2023) Content Validity Index of Basic Education Certificate Examination (BECE) Test Items in Basic Science and Technology (BST), British Journal of Education, Vol.11, Issue 6, pp. 7-24,

ABSTRACT: *This study quantitatively provided evidence of content validity of BECE items in BST by estimating content validity index (CVI). Descriptive survey research designs were employed. Purposive and accidental sampling techniques were adopted in selecting 31 BST teachers in Abia State junior secondary schools. Three research questions were addressed. The instrument for data collection was 2021 BECE items in BST comprising four subtests with 30 items each. Item level CVIs, test level CVI and modified kappa statistic were computed. Findings revealed that 95% of the items of BST reported I-CVIs ranging from 0.70 – 1.00 and were accepted as having adequate content validity whereas the remaining 5% of the items, four from PHE and one from Computer Studies had low I-CVIs. The T-CVI yielded a high index of 0.95. Finally, kappa statistic adjusting for chance agreement indicated 0.763 – 1.00 for 95% of the BST items which proved excellent agreement among the experts.*

KEYWORDS: Basic Science and Technology, Item level content validity index, Test level content validity index, Kappa statistic

INTRODUCTION

Basic Education Certificate Examination (BECE) is a national examination administered annually by National Examination Council (NECO) to assess students' achievement in Basic Science and Technology (BST) along with other subject in Nigeria. An assessment tool like BST is administered to candidates at the end of three years upper basic education designated at Junior Secondary three (JS3) to determine the extent they can accomplish certain tasks, demonstrate mastery of skills and knowledge of content area (Adom et al., 2020). Crucial decisions about individuals that sat for examination is the nation are based on the test scores generated. Therefore, the validity of this assessment tool is of essence as it is the major psychometric characteristics of any instrument considered. For BST to be adjudged a good assessment tool it should show evidence of high validity, objectivity, comprehensiveness, ease of use and results can be justified (Adom et al., 2020).

Validity is the extent to which an assessment tool measures what it was designed to measure and how well it does it. Kubai (2019) sees validity as the degree to which an instrument measures what it intended to measure. Further Wojton (2017) defined validity as the extent to which scores produced by assessment instruments reflect what the instrument is intended to measure. Wojton definition supports the assertion that "validity is not the property of an instrument, but the property of the scores achieved by an instrument used for a specific purpose on a special group of respondents" (Zamanzadeh et al., 2015: 166). Validity, therefore, is concerned with the specific use of the test scores and the fair interpretations of those scores. Four basic types of validity include face validity, content validity; construct validity and criterion-related validity. This present study explored content validity index of BST items using as assessment instrument in BECE in Nigeria.

REVIEW OF RELATED LITERATURE

Content Validity

One of the theoretical considerations in testing is the validity of the assessment instrument. Content validity is considered a critical step from the outset of instrument development because it is required as a prior condition for other validities (Aravamudhan & Krisahnaveni, 2015). Content validity estimates the extent to which items of a measurement instrument make fair coverage or are representative of the content area the instrument purports to measure (Newwman et al., 2013). It asks the question for every item included, 'does this item look like it measures the content it intends to measure'? (Newman et al., 2013). Content validity is further conceptualized as "… the degree to which an instrument has an appropriate sample of items for the construct being measured" (Polit & Beck, 2006: 489). The importance of content validity cannot be overemphasized. Apart from being a prior condition for other validities, it provides evidences on the representativeness and clarity of items and helps improve an assessment instrument through recommendations of expert judgment on the representativeness (Zamanzadeh et al., 2014). Despite the indispensible nature of content validity in instrument development, it is not studied in depth probably because researchers do not realize the required complexities involved or the methods of estimating it are not referred to and no single resource has provided sufficient details on content validity (Rattray & Jones, 2007; Beck cited in Zamanzadeh et al., 2014).

Content validation of a cognitive instrument like BST can be achieved qualitatively and quantitatively by use of table of specification and the judgments of panel of experts respectively. Newman et al. (2013:5) operationally defined table of specification as "a set of procedures that attempts to align a set of items, tasks, or evidence with a set of concepts that are to be assessed". Earlier researchers employed test blue print in establishing content validity of achievement tests by ensuring fair coverage and representativeness of the entire content area and instructional objectives (Newman et al., 2011; Amajuoyi et al., 2013; Osadebe, 2015; Bano et al., 2021; Longjohn et al., 2021). However, Longjohn et al. (2021), validating a basic science achievement test in addition to table of specification, employed DIMPACK software for unidimensionality analysis of the test. The judgments of the panel of experts' approach is widely used in researches in medical, health sciences and psychological tests for establishing the content validity ratio (CVR), content validity index (CVI) for items and entire scale (Polit & Beck, 2006; Polit et al., 2007; Zamanzadeh et al., 2014; Aravamudhan & Krishnaveni, 2015; Zamanzadeh et al., 2015; Vakili & Jahangiri, 2018).

Quantifying of Content Validity

Quantification of content employs the methods of content validity ratio (CVR) which is an item statistic that allows for rejection or retention of specific items in a test based on the panelists judgment on whether the tested knowledge or skill is essential or not essential (Gilbert & Prison, 2016). Content Validity Index (CVI), is the most widely reported measure of quantification of content validity especially, in health sciences (Polit & Beck, 2006; Zamanzadeh et al., 2014). However, a few educational tests employed CVR and CVI in establishing content validity of tests (Ikhsanudin & Subali, 2018; Lau et al., 2019). This study explored quantification of content validity using CVI.

As proposed by Lynn (1986), CVI involves computation of content validity individual items (I-CVI) and computation of content validity of the entire test (T-CVI). Content validity ratio (CVR) provides an item level information based on the judgments of the panels of experts on the essentiality of it rated on three-point scale of essential, useful but not essential, and not essential (Gilbert & Prison, 2016). The rating of the experts is used to calculate the value of the content validity using the formula:

$$CVR = \frac{(ne - (N/2)}{N/2}$$

Where, ne is the number of members of panel of experts indicating an item as essential while N is the total number of panel of experts (Zamanzadeh et al., 2015). To decide on the acceptability or inclusion of each item in the test, the CVR of each item is compared with the values on Lawshe's table. "If the calculated content validity value is equal to or higher than the determined value in Lawshe's table, the item is retained; otherwise, it should be eliminated from the final test" (Vakili & Jahangiri, 2018).

Employing the CVI for establishing content validity involves steps requesting the members of the panel of experts to rate items of the instrument in terms of clarity and relevance to the construct underlying the study (Zamanzadeh et al., 2015). The experts rate on a 4-point scale with response category of not relevant (1-point), somewhat relevant (2-points), quite relevant (3-points), and very relevant (4-points). Content validity Index is estimated in two levels namely, item level content

British Journal of Education
Vol.11, Issue 6, pp. 7-24, 2023
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

validity Index (I-CVI) and test level content validity index (T-CVI). Whereas I-CVI expresses the number of experts judging an item as relevant by rating either 3 or 4, T-CVI gives the proportion of items on an instrument that were rated 3 or 4 by the content experts (Beck & Gable, 2001; Polit et al., 2007).

There are two methods for computing T-CVI, the first method requires universal agreement among members of panel of experts (T-CVI/UA) while the other involves averaging of the I-CVIs (T-CVI/Ave) (Zamanzadeh et al., 2014). Admittedly, values are obtained as T-CVI/UA and T-CVI/Ave, Polit and Beck (2006) contend that instrument developers should state clearly the method used. On the acceptable standard for T-CVI, Davis cited in Polit et al., (2007) proposed that values equal to or greater than 0.80 agreements among judges are acceptable standard for new instruments. Abdollaphour et al., cited in Lau et al., 2019 stated that judgment on each item is as follows; if CVI is greater than 0.79, the item is retained as appropriate, if the CVI is between 0.70 and 0.79, the item needs revision, and if the CVI is less than 0.70, the item is discarded.

Content Validity Index and Chance Agreement

Despite the wide use of CVI for establishing content validity, there are arguments on upholding it as standard of acceptability of items for inclusion because of the chance agreement. Chance agreement is an issue that arises in evaluating indices of inter-rater agreement particularly, when a four-point scale on which the expert responded is dichotomized into relevant and not relevant categories (Wynd et al., 2003). To adjust for the risk of chance agreement, the authors advocated the use of multi-rater Kappa coefficient as it provides information on degree of agreement beyond chance. Content validity indices in general capture various kinds of agreement, including agreement about the low relevance of an item; this is observed to be a problematic measure of CVI. On the other hand, kappa captures agreement on consensus about relevance or non-relevance of an item. Regardless the challenge of chance agreement, CVI is still utilized by researchers and at most supplemented by Kappa statistic because CVI is simple to calculate, easy to understand and provides information on item – level which is used for retention, revision, or elimination of items from the test (Polit & Beck, 2006; Polit et al., 2007). The formula for computing kappa is shown as follows:

$$k = \frac{\text{I-CVI} - \text{Pc}}{1 - \text{Pc}}$$

In this formula, I-CVI is the content validity index for each item; Pc is the probability of chance agreement for each item. It is noteworthy that after adjusting for chance agreement by computing kappa, any item with I-CVI equal or greater than 0.78 is considered excellent and that as the number of judges increases, the probability of chance agreement decreases with convergence of values of I-CVI and kappa coefficient (Baheiraei et al., 2013).

Stating the obvious, the use of BST items as an assessment tool for evaluating the ability of Junior Secondary School Students in sciences requires a test with items representative of the content

British Journal of Education
Vol.11, Issue 6, pp. 7-24, 2023
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

domains. Moreover, development of an instrument requires report on the content validation. This study therefore, sought to estimate the I-CVI, T-CVI and kappa statistic for BST test items used for JS3 candidates that sat for BECE in Nigeria in 2021. To achieve this purpose, three research questions were posed to guide the study.

1. What is the I-CVI of BST sub tests items used for JS3 candidates that sat for BECE in Nigeria 2021?
2. What is the T-CVI of BST sub-tests and entire BST used for JS3 candidates that sat for BECE in Nigeria 2021?
3. What is the modified Kappa statistics for BST items used for JS3 candidates that sat for BECE in Nigeria 2021?

METHODOLOGY

Design of the Study

This study adopted descriptive survey research design which requires collection of large volume of data, on a representative sample of a population, that are quantifiable (Kpolovie, 2010). Descriptive survey design was deemed appropriate because a sample of subject specialists were used as judges in the panel of experts.

Participants

The target population of the study comprised the entire Basic Science (BS), Physical and Health Education (PHE), Basic Technology (BT) and Computer Studies (CS) teachers in Abia State Junior Secondary Schools that presented students for the 2021 BECE organized by NECO. Purposive sampling technique was used to select two Junior Secondary schools that participate in BECE while accidental sampling technique was used to draw a total of 31 experts consisting of ten BS teachers, five PHE teachers, ten BT teachers and six CS teachers. These experts were available and were willing to participate in evaluating each of the sub-tests of BST. The criterion for selecting the panel of experts was their subject areas. The recommended number of judges for the chance agreement was five experts (Bahiraei et al., 2013). However, Zamanzadeh et al. (2015) advocated for maximum 10 experts because the probability of chance agreement decreases as number of experts increases.

Instruments

The instruments for data collection were 2021 Basic Education Certificate Examination (BECE) items in BST and BST Syllabus for JSS 1 – 3. BST comprised four sub-tests with 30 items in each of the four sub-tests giving a total of 120 items. The researchers visited the experts and sought their approval to participate in the study, on approval; they were presented with copies of the instrument (BST items), the syllabus, a cover letter explaining the purpose of the study and the guide for scoring the instrument. Panel members were requested to rate the 30 items in their subject areas in terms of clarity and relevance to the content domain on a 4-point rating scale of not relevant, somewhat relevant, quite relevant and highly relevant, weighted 1 point, 2 points, 3 points and 4 points respectively.

Data Analysis

The data collected was analyzed by dichotomizing the ratings of the judges with rating 1 and 2 (not relevant) as one category and rating 3 and 4 (relevant) as another. The I-CVIs were computed by dividing the number of those experts that judged the item as relevant (rating 3 or 4) with the total number of experts. The T-CVI/Ave was calculated by dividing the sum of I-CVIs by the total number of items. This gave the content validity index for the entire BST. The criteria for quantitative values of CVI were as follows: I-CVI below 0.70 was eliminated, 0.70 – 0.78 were recommended for revision and equal or above 0.79 were considered appropriate (Lau et al., 2019). In other words, for a BST sub test, content validity index, I-CVI, has to be equal to 1.00 for 3 to 5 experts and a minimum of 0.78 for 6 to 10 experts. The T-CVI/Ave of equal or greater than 0.80 are accepted as adequate (Njelesani et al., 2019). The modified Kappa coefficient was obtained first by calculating the probability of chance agreement using the formula $Pc = [N!/A!\ (N – A)!] * .5^N$, where Pc is probability of chance agreement, N is number of experts in the panel; A is number of experts rating item 3 or 4 (relevant). Thereafter, modified Kappa coefficient was calculated using the formula K = (I-CVI – Pc) / (1 – Pc) Evaluation criteria for Kappa were as follows: values equal to or higher than 0.74 were considered excellent; values between 0.60 and 0.74 were good and values between 0.40 and 0.59 were considered to have fair content validity index (Zamanzadeh et al. cited in Alonso-Ferres et al., 2022).

RESULTS

Research Question 1: What is the I-CVI of BST sub tests items used for JS3 candidates that sat for BECE in Nigeria 2021

Table 1. Item level content validity index (I-CVI) of BST subtests.

Items	Relevant (3 or 4)	Not Relevant (1 or 2)	I-CVI	Decision
		Basic Science		
1	10	0	1.0	Appropriate
2	10	0	1.0	Appropriate
3	9	1	0.9	Appropriate
4	10	0	1.0	Appropriate
5	10	0	1.0	Appropriate
6	10	0	1.0	Appropriate

British Journal of Education
Vol.11, Issue 6, pp. 7-24, 2023
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

7	10	0	1.0	Appropriate
8	10	0	1.0	Appropriate
9	10	0	1.0	Appropriate
10	10	0	1.0	Appropriate
11	10	0	1.0	Appropriate
12	10	0	1.0	Appropriate
13	10	0	1.0	Appropriate
14	9	1	0.9	Appropriate
15	10	0	1.0	Appropriate
16	10	0	1.0	Appropriate
17	10	0	1.0	Appropriate
18	10	0	1.0	Appropriate
19	10	0	1.0	Appropriate
20	10	0	1.0	Appropriate
21	10	0	1.0	Appropriate
22	10	0	1.0	Appropriate
23	9	0	0.9	Appropriate
24	10	0	1.0	Appropriate
25	10	0	1.0	Appropriate
26	9	1	0.9	Appropriate
27	10	0	1.0	Appropriate
28	9	1	0.9	Appropriate
29	9	1	0.9	Appropriate
30	10	0	1.0	Appropriate
		Physical and Health Education		
1	4	1	0.8	Appropriate

2	4	1	0.8	Appropriate
3	4	1	0.8	Appropriate
4	5	0	1.0	Appropriate
5	3	2	0.6	Eliminate
6	5	0	1.0	Appropriate
7	5	0	1.0	Appropriate
8	5	0	1.0	Appropriate
9	5	0	1.0	Appropriate
10	5	0	1.0	Appropriate
11	4	1	0.8	Appropriate
12	5	0	1.0	Appropriate
13	5	0	1.0	Appropriate
14	5	0	1.0	Appropriate
15	5	0	1.0	Appropriate
16	5	0	1.0	Appropriate
17	5	0	1.0	Appropriate
18	5	0	1.0	Appropriate
19	3	2	0.6	Eliminate
20	3	2	0.6	Eliminate
21	5	0	1.0	Appropriate
22	5	0	1.0	Appropriate
23	5	0	1.0	Appropriate
24	5	0	1.0	Appropriate
25	5	0	1.0	Appropriate
26	5	0	1.0	Appropriate
27	5	0	1.0	Appropriate

28	5	0	1.0	Appropriate
29	3	2	0.6	Eliminate
30	5	0	1.0	Appropriate
		Basic Technology		
1	10	0	1.0	Appropriate
2	10	0	1.0	Appropriate
3	10	0	1.0	Appropriate
4	10	0	1.0	Appropriate
5	10	0	1.0	Appropriate
6	10	0	1.0	Appropriate
7	9	1	0.9	Appropriate
8	9	1	0.9	Appropriate
9	10	0	1.0	Appropriate
10	10	0	1.0	Appropriate
11	10	0	1.0	Appropriate
12	10	0	1.0	Appropriate
13	9	1	0.9	Appropriate
14	10	0	1.0	Appropriate
15	10	0	1.0	Appropriate
16	10	0	1.0	Appropriate
17	10	0	1.0	Appropriate
18	10	0	1.0	Appropriate
19	10	0	1.0	Appropriate
20	10	0	1.0	Appropriate
21	10	0	1.0	Appropriate
22	10	0	1.0	Appropriate

British Journal of Education
Vol.11, Issue 6, pp. 7-24, 2023
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

23	10	0	1.0	Appropriate
24	10	0	1.0	Appropriate
25	10	0	1.0	Appropriate
26	10	0	1.0	Appropriate
27	9	1	0.9	Appropriate
28	10	0	1.0	Appropriate
29	10	0	1.0	Appropriate
30	9	1	0.9	Appropriate
		Computer Science		
1	5	1	0.83	Appropriate
2	5	1	0.83	Appropriate
3	5	1	0.83	Appropriate
4	5	1	0.83	Appropriate
5	5	1	0.83	Appropriate
6	6	0	1.0	Appropriate
7	5	1	.83	Appropriate
8	6	0	1.0	Appropriate
9	5	1	0.83	Appropriate
10	4	2	0.67	Eliminate
11	5	1	0.83	Appropriate
12	6	0	1.0	Appropriate
13	6	0	1.0	Appropriate
14	6	0	1.0	Appropriate
15	6	0	1.0	Appropriate
16	6	0	1.0	Appropriate
17	6	0	1.0	Appropriate

18	6	0	1.0	Appropriate
19	6	0	1.0	Appropriate
20	5	1	0.83	Appropriate
21	5	1	0.83	Appropriate
22	5	1	0.83	Appropriate
23	5	1	0.83	Appropriate
24	6	0	1.0	Appropriate
25	6	0	1.0	Appropriate
26	6	0	1.0	Appropriate
27	6	0	1.0	Appropriate
28	5	1	0.83	Appropriate
29	6	0	1.0	Appropriate
30	6	0	1.0	Appropriate

Table 1 presented the I-CVI for items in each subtest. Based on the criteria for accepting an item as having adequate CVI, all the 30 items of BS were retained as having appropriate I-CVIs with six items yielding I-CVI of 0.90 and 24 items with I-CVI of 1.0. PHE has 26 items with I-CVIs between 0.80 and 1.0 which were accepted as having appropriate and adequate I-CVIs. Four items (5, 19, 20, and 29) with I-CVI of 0.60 were eliminated because they have low I-CVIs. BT with 30 items returned I-CVI = 1.0 for 25 items and I-CVI = 0.90 for 5 items. The 30 items were all adjudged appropriate with adequate I-CVIs. CS reported 16 items with I-CVI of 1.0 and 13 items with I-CVI of 0.83. These items were recommended as appropriate and were retained. Item 10 with I-CVI of 0.67 was eliminated as the I-CVI was considered unacceptably low.

Research Question 2: What is the T-CVI of BST sub-tests and entire BST used for JS3 candidates that sat for BECE in Nigeria 2021?

Table 2.Test Level content validity index (T-CVI) for subtests and entire BST

Indices	BS	PHE	BT	CS	BST
Number items	30	30	30	30	-
Subtest items rated relevant	24	22	25	16	-
Total I-CVI	29.4	27.6	29.5	27.5	-
T-CVI/Ave	.98	.92	.98	.92	.95

Table 2 showed the calculation of T-CVI for each sub-test. BS yielded T-CVI /Ave (0.95) and the index was adequate. For PHE subtest level content validity index was very adequate, T-CVI / Ave (0.92) method. For BT, content validity was adequate, T-CVI / Ave (0.98). Finally, test level content validity Index for CS was adequate, T-CVI /Ave (0.92). The average of subtest level T-CVIs gave the content validity index of the entire BST as T-CVI /Ave (0.95).

Research Question 3: What is the modified Kappa statistics for BST items used for JS3 candidates that sat for BECE in Nigeria 2021?

Table 3. Modified kappa statistics for BST subtest items

Number of Experts	Items	Experts Rating 3 or 4	I-CVI	*Pc*	*K**	Evaluation
			Basic Science			
10	1	10	1.00	.000977	1.00	Excellent
10	2	10	1.00	.000977	1.00	Excellent
10	3	9	.90	.009766	.899	Excellent
10	4	10	1.00	.000977	1.00	Excellent
10	5	10	1.00	.000977	1.00	Excellent
10	6	10	1.00	.000977	1.00	Excellent
10	7	10	1.00	.000977	1.00	Excellent
10	8	10	1.00	.000977	1.00	Excellent
10	9	10	1.00	.000977	1.00	Excellent
10	10	10	1.00	.000977	1.00	Excellent
10	11	10	1.00	.000977	1.00	Excellent
10	12	10	1.00	.000977	1.00	Excellent
10	13	10	1.00	.000977	1.00	Excellent
10	14	9	.90	.009766	.899	Excellent
10	15	10	1.00	.000977	1.00	Excellent
10	16	10	1.00	.000977	1.00	Excellent
10	17	10	1.00	.000977	1.00	Excellent
10	18	10	1.00	.000977	1.00	Excellent
10	19	10	1.00	.000977	1.00	Excellent
10	20	10	1.00	.000977	1.00	Excellent
10	21	10	1.00	.000977	1.00	Excellent
10	22	10	1.00	.000977	1.00	Excellent
10	23	9	.90	.009766	.899	Excellent
10	24	10	1.00	.000977	1.00	Excellent
10	25	10	1.00	.000977	1.00	Excellent
10	26	9	.90	.009766	.899	Excellent
10	27	10	1.00	.000977	1.00	Excellent
10	28	9	.90	.009766	.899	Excellent
10	29	9	.90	.009766	.899	Excellent
10	30	10	1.00	.000977	1.00	Excellent
			Physical and Health Education			
5	1	4	0.8	.156250	.763	Excellent
5	2	4	0.8	.156250	.763	Excellent
5	3	4	0.8	.156250	.763	Excellent

5	4	5	1	.031250	1.00	Excellent
5	5	3	0.6	.078125	.566	Fair
5	6	5	1	.031250	1.00	Excellent
5	7	5	1	.031250	1.00	Excellent
5	8	5	1	.031250	1.00	Excellent
5	9	5	1	.031250	1.00	Excellent
5	10	5	1	.031250	1.00	Excellent
5	11	4	0.8	.156250	.763	Excellent
5	12	5	1	.031250	1.00	Excellent
5	13	5	1	.031250	1.00	Excellent
5	14	5	1	.031250	1.00	Excellent
5	15	5	1	.031250	1.00	Excellent
5	16	5	1	.031250	1.00	Excellent
5	17	5	1	.031250	1.00	Excellent
5	18	5	1	.031250	1.00	Excellent
5	19	3	0.6	.078125	.566	Fair
5	20	3	0.6	.078125	.566	Fair
5	21	5	1	.031250	1.00	Excellent
5	22	5	1	.031250	1.00	Excellent
5	23	5	1	.031250	1.00	Excellent
5	24	5	1	.031250	1.00	Excellent
5	25	5	1	.031250	1.00	Excellent
5	26	5	1	.031250	1.00	Excellent
5	27	5	1	.031250	1.00	Excellent
5	28	5	1	.031250	1.00	Excellent
5	29	3	0.6	.078125	.566	Fair
5	30	5	1	.031250	1.00	Excellent
			Basic Technology			
10	1	10	1	.000977	1.00	Excellent
10	2	10	1	.000977	1.00	Excellent
10	3	10	1	.000977	1.00	Excellent
10	4	10	1	.000977	1.00	Excellent
10	5	10	1	.000977	1.00	Excellent
10	6	10	1	.000977	1.00	Excellent
10	7	9	0.9	.009766	.899	Excellent
10	8	9	0.9	.009766	.899	Excellent
10	9	10	1	.000977	1.00	Excellent
10	10	10	1	.000977	1.00	Excellent
10	11	10	1	.000977	1.00	Excellent
10	12	10	1	.000977	1.00	Excellent
10	13	9	0.9	.009766	.899	Excellent
10	14	10	1	.009766	.899	Excellent
10	15	10	1	.000977	1.00	Excellent
10	16	10	1	.000977	1.00	Excellent
10	17	10	1	.000977	1.00	Excellent
10	18	10	1	.000977	1.00	Excellent
10	19	10	1	.000977	1.00	Excellent
10	20	10	1	.000977	1.00	Excellent

British Journal of Education
Vol.11, Issue 6, pp. 7-24, 2023
Online ISSN: 2054-636X (Online)
Print ISSN: 2054-6351(Print)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development-UK

10	21	10	1	.000977	1.00	Excellent
10	22	10	1	.000977	1.00	Excellent
10	23	10	1	.000977	1.00	Excellent
10	24	10	1	.000977	1.00	Excellent
10	25	10	1	.000977	1.00	Excellent
10	26	10	1	.000977	1.00	Excellent
10	27	9	0.9	.009766	.899	Excellent
10	28	10	1	.000977	1.00	Excellent
10	29	10	1	.000977	1.00	Excellent
10	30	9	0.9	.009766	.899	Excellent
			Computer Studies			
6	1	5	0.83	..093750	.812	Excellent
6	2	5	0.83	.093750	.812	Excellent
6	3	5	0.83	..093750	.812	Excellent
6	4	5	0.83	..093750	.812	Excellent
6	5	5	0.83	..093750	.812	Excellent
6	6	6	1	.015625	1.00	Excellent
6	7	5	0.83	..093750	.812	Excellent
6	8	6	1	.015625	1.00	Excellent
6	9	5	0.83	..093750	.812	Excellent
6	10	4	0.67	.234375	.569	Fair
6	11	5	0.83	..093750	.812	Excellent
6	12	6	1	.015625	1.00	Excellent
6	13	6	1	.015625	1.00	Excellent
6	14	6	1	.015625	1.00	Excellent
6	15	6	1	.015625	1.00	Excellent
6	16	6	1	.015625	1.00	Excellent
6	17	6	1	.015625	1.00	Excellent
6	18	6	1	.015625	1.00	Excellent
6	19	6	1	.015625	1.00	Excellent
6	20	5	0.83	..093750	.812	Excellent
6	21	5	0.83	..093750	.812	Excellent
6	22	5	0.83	..093750	.812	Excellent
6	23	5	0.83	..093750	.812	Excellent
6	24	6	1	.015625	1.00	Excellent
6	25	6	1	.015625	1.00	Excellent
6	26	6	1	.015625	1.00	Excellent
6	27	6	1	.015625	1.00	Excellent
6	28	5	0.83	..093750	.812	Excellent
6	29	6	1	.015625	1.00	Excellent
6	30	6	1	.015625	1.00	Excellent

Table 3 presented modified kappa statistic for adjusted each I-CVI for chance agreement. For BS and BT subtests, all the items were considered for excellent agreement because they returned modified kappa coefficient of 0.899 – 1.0. For PHE subtest, 26 and 4 out of 30 items yielded kappa coefficient of 1.0 and .763 respectively. These were considered excellent agreement. However, 4 items (5, 19, 20, 29) yielded kappa coefficient of .566. This is a low coefficient. Kappa coefficient of 1.00 and 0.812 were obtained by 16 and 13 items respectively on CS subtest. Only item 10 returned a low kappa coefficient of 0.569 for CS subtest.

DISCUSSION

The study examined the content validity index of BST with four subtests. The first research question sought answers to the content validity index of each item in the four subtests. In the four subtests, items with I-CVI with high relevant score (≥ 0.79) were accepted as appropriate and all the items were retained. These constituted more than 95% of the items in the four subtests. The criteria used was the submission of Lau et al. (2019); Polit et al. (2007) who suggested that items with I-CVI ≥ 0.79 for three or more experts could be considered evidence of good content validity. Four items from PHE subtest and one item from CS subtest were judged as having inadequate content validity indices because they reported I-CVIs less than 0.70.

The second research question presented the test level CVIs for the subtests and for the entire BST. The T-CVI /Ave method was used for the subtests and the average of subtests T-CVIs gave test level CVI for the entire BST (Lau, et al., 2019). All the subtests yielded adequate index of .92 and above. Applying the acceptability standard of 0.80, content validity of the 120 items of BST was adequate with T-CVI / Ave (0.95) (Polit et al., 2007).

Finally, the third research question transformed the values of I-CVIs into values of modified Kappa statistic to adjust for chance agreement. For items of BST, about 95% obtained Kappa coefficient of 0.763 – 1.00 which is prove of excellent agreement among experts (Cicchetti & Sparrow in Zamanzadeh, 2015). On the other hand, four items from PHE subtest yielded coefficient of 0.566 and one item from CS a coefficient of 0.569. Both 0.566 and 0.569 are low coefficients and evidence of lack of consensus among the experts (Polit et al., 2007).

IMPLICATION TO RESEARCH AND PRACTICE

The findings of the study have the following implications:

- The results of this study suggest that test developers need to apply content validity index for establishing content validity of instruments since psychometric properties of any measuring instrument is of the essence.
- These findings further hold implications for test developers who are expected verify test-level content validity index (T-CVI) alongside item level content validity index (I-CVI).
- A key area emerging from the results is that unlike CVI, kappa captures agreement on consensus about relevance or non-relevance of an item hence, adjusting for chance agreement which an issue of concern in inter-rater agreement.

CONCLUSIONS

Based on the findings, it was concluded that the item-level content validity index of 95% of the BST subtest items were adequate. The remaining 5% from PHE and CS subtests are to be discarded. The test-level content validity index of BST using the averaging method gives more robust content validity index. The T-CVI /Ave is preferred not only because it expresses item-level information through the averaging feature, it also takes care of the risk of chance agreement. The study also concluded that modified Kappa index of 0.763 – 1.00 for 115 items out of 120 items of BST were substantial and showed evidence of excellent agreement among experts. Therefore, BST is a valid instrument for assessing students at Junior Secondary School level.

FUTURE RESEARCH

The current study explored content validity index as a quantitative method to content validity evidence of BST items, it would be interesting to replicate the study for other subjects examined by BECE and also apply a qualitative method of using table of specification. The content validity study that employs a mixed method approach can produce a good test with appropriate psychometric properties. In addition to validity evidence, reliability evidence based on testlets effect (BST has four subtests) analyzed using generalizability theory should also be obtained.

REFERENCES

Adom, D., Adu-Mensah, J. & Dake, D. A. (2020). Test, measurement and evaluation: Understanding and use of the concept in education. *International Journal of Evaluation and Research in Education*, 9(1), 109 – 119, https://www.researchgate.net

Alonso-Ferres, M., Serrano-Montilla, C., Valor-Segura, I. & Exposito, F. (2022). The Spanish adaptation of the relationship power inventory. *Current Psychology*. https://doi.org/10.1007/s12144-022-03072-4

Amajuoyi, I. J., Joseph, E. U. & Udoh, N. A. (2013). Content validity of May/June West African Senior School Certificate examination (WASSCE) questions in Chemistry. *Journal of Education and Practice*, 4 (7), 15 21. https://www.researchgate.net

Aravamudhan, N. R. & Krishnaveni, R. (2015). Establishing and reporting content validity evidence of new training and development capacity building scale. *Management*, 20 (1), 131 – 158. https://www.researchgate.net

Baheiraei, A., Khoori, E. […] & Weiler, R. M. (2013). Psychometric properties of the adolescent health concern inventory: The Persian Versia, *Iranian Journal of Psychiatry*, 8 (1), 28 – 36. https://www.researchgate.net

Bano, H., Khan, M. S. & Khan, S. I. (2021). Development of valid and reliable mathematics achievement test. *Elementary Education Online*, 20 (2), 1642 – 1658. www.ilkogretim-online.org

Beck, C. T. & Gable, R. K. (2001). Ensuring content validity; An illustration of the process. *Journal of Nursing Measurement*, 9 (2), 201 – 215. https://www.researchgate.net

Gilbert, G. E. & Prion, S. (2016). Making sense of measurement: Lawshe's content validity index. *Clinical simulation in Nursing*, 12 (2), 530 – 531. https://www.semanticscholar.org

Ikhsanudin, I. & Subali, B. (2018). Content validity analysis of first semester formative subject for senior high school. *Journal of Physics Conference Series*. 1097:012039. https://www.researchgate.net

Kpolovie, P. J. (2010). *Advanced research method*. Springfield Publishers Ltd.

Kubai, E. (2019). Reliability and validity of research instruments. *Critical Analysis of Policies on Special Education in Kenya*. www.researchgale.net.

Lau, X. C., Wong, L. Wong, I. E.., […], & Poh, B. K. (2019). Development and validation of a physical education module for overweight adolescents: CERGAS programe international. *Journal of Environmental Research and Public Health*, 16, 1506. www.mdpi.com/journal/ijerph

Longjohn, I. Njigwum, A. S. & Ibeh, C. P. (2021). Validation of basic science achievement test using R. software. *International Journal of Innovative Research and Development, 10 (9), 77 – 87*. https://internationaljournalcorner.com

Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35, 382 – 385. https://journals.lww.com

Njelesani, J., Davis, J. A. & Pontes, T. (2019). Occupational repertoire development measure-parent (ordm-p): Face validity, comprehensiveness, and internal consistency. *British Journal of Occupational Therapy*, 83(5),. https://journals.sagepub.com

Newman, I., Lim, J. & Pineda, F. (2013). Content validity using mixed methods approach: Its application and development through the use of a table of specifications methodology. Journal of Mixed Methods Research, 7(3):243-260. https://www.researchgate.net

Osadebe, P. U. (2015). Construction of valid and reliable test for assessment of students. *Journal of Education and Practice*, 6 (1), 51 – 56. https://files.eric.ed.gov

Polit, D. F. & Beck, C. T. (2006). The content validity index: Are you sure you know what is being reported? Critique and recommendations; *Research in Nursing and Health*, 29, 489 – 497. https://pubmed.ncbi.nlm.nih.gov

Polit, D. F., Beck, C. T. & Owen, S. V. (2007). Focus on research methods: Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing and Health*, 30, 459 – 467. https://file.qums.ac.ir

Rattray, J. & Jones, M. C. (2007). Essential elements of questionnaire design and development. *Journal of Clinical nursing*, 16 (2), 234 – 243. https://pubmed.ncbi.nlm.nih.gov

Vakili, M. M. & Jahangiri, N. (2018). Content validity and reliability of the measurement tools in educational, behavioural, and health sciences research. *Journal of Medical Education Development* 10 (28), 106 – 119. https://zuns.ac.ir/edujournal

Wojton, H. (2017). *Foundations of psychological measurement*. Institute for Defense Analysis. https://www.jstor.org/stable/resp22739 Wynd, C. A. Schmidt, B. & Schaefer, M. A. (2003). Two quantitative approaches for estimating content validity. *Western Journal of Nursing Research*, 25 (5), 508 – 518. https://pubmed.ncbi.nlm.nih.gov

Zamanzadeh, V., Ghahramanian, A., Rassaili, M., Abboszedeh, A., Alavi-Majd, H. & Nikanfer, A. (2015). Design and implementation content validity study. Development of an instrument for measuring patient-centered communication. *Journal of Caring Science*, 4 (2), 165 – 178. www.ncbi.nlm.nih.gov

Zamanzadeh, V., Rassouli, M., Abbaszedeh, A., Alavi-Majd, Nikamfar, A. & Ghahramanian, A. (2014). Details of content validity and objectifying it in instrument development. *Nursing Practice Today*, 1(3), 163 – 171. https://npt.tums.ac.ir