E-ISSN: 2706-8927
P-ISSN: 2706-8919
www.allstudyjournal.com
IJAAS 2021; 3(1): 259-265
Received: 04-10-2020
Accepted: 06-12-2020

Sayed Abdul Aziz Ahmady Falah
Chemistry Department, Education Faculty of Parwan University, Afghanistan

Bashir Ahmad Karimi
Chemistry Department, Education Faculty Of Parwan University, Afghanistan

Corresponding Author:
Sayed Abdul Aziz Ahmady Falah
Chemistry Department, Education Faculty of Parwan University, Afghanistan

Developing of valid and reliable stem test item for assessment of middle school

Sayed Abdul Aziz Ahmady Falah and Bashir Ahmad Karimi

Abstract
The fourth industrial revolution is linked to the 21st century. Education 4.0 must be improved or established in order to overcome the fourth industrial revolution. The main feature of Education 4.0 is the integration of technology and the STEM program at school. Standard and regular assessment and evaluation can improve education, including STEM and student learning achievement. The aim of this study was to develop valid and reliable STEM items for middle school students (MSS). The researcher developed 15 subjective and 5 objective questions and randomly selected 30 students from Etihad Private School. The test is validated by a senior lecture, an education school, Teknologi University Malaysia and an experienced teacher from Afghanistan. For both query forms, which have the Alpha value of the Cronbach subjective test item of 0.77 and the reliability factor for the objective test item of 0.8, the results indicate a reasonable reliability. Researchers proposed more study with a larger sample size and more random sampling method subjective questions, as well as the introduction of another criterion of reliability, such as test reliability, inter-rater reliability.

Keywords: STEM, chemistry, test item, validity, reliability

1. Introductions
The fourth industrial revolution, or IR4.0, is applied to the 21st century. Based on the first industrial revolution, we can infer that IR4.0, in all units of the human world, including the educational system, induces the great transition. Yet IR4.0 or the emergence of the next technological revolution has provided great social and economic recreation, and has also seen it as a huge human challenge around the world (Manda & Backhouse, 2017) [16]. Undoubtedly, not only trade, government and citizens are affected by the forthcoming industrial revolution, but also by education and training of 4.0 was created. The important aspect of education 4.0 is technical convergence. In reality, education for 4.0 meets the criteria of the fourth industrial revolution (IR.4.0), in order to develop new technology based of people requirements (Diwan, 2017) [9]. In human lives such as, social events, political problems, economic concerns, studying and learning, and even in our free time activities, innovations are growing rapidly (Aida Arjani Shahroom, 2018) [2].
Nowadays, STEM program is one of the most important teaching strategy related to the education 4.0. The STEM Program was launched by the National Science Foundation (NSF) as the Science, Mathematics, Engineering and Technology (SMET) program. This training initiative was designed to provide all students with critical thinking skills that would enable them to solve creative problems and ultimately make them more marketable. It is seen that every student involved in STEM education, particularly inK-12 would be advantageous if they choose not to pursue post-secondary education or if they did attend college, in particular in the STEM field, they would have even greater advantage. (Butz *et al*., 2004) [3]. The Instructor 4.0 takes account of the growth and promotion of this progress (Karre, Hammer, Kleindienst, & Ramsauer, 2017) [13].
Today, the STEM program is one of the most effective education 4.0 strategy. In the form of the Science, Mathematics, Engineering and Technology (SMET) program the national science foundation (NSF) has initiated the STEM programme. This training initiative was designed to offer all students critical thinking skills which could solve and eventually market innovative problems. It can be seen that any STEM student, particularly in K-12, would be advantageous if he/she did not choose to study post-secondary education or if they attended college, especially STEMs (Butz *et al*., 2004) [3]
Moreover, even college graduates who intend to compete internationally, the STEM offers better training in secondary schools.

That means that each student can master their curriculum program and direct them to learn from the school more innovatively, creatively, problem-solving and critically. STEM substitutes more research and project-oriented methods for conventional lecture-based teaching techniques (Breiner, Harkness, Johnson, & Koehler, 2012). Scott's (2012) revealed that STEM declaration help is technical improvements and offers students master's know-how for their future professions.

In addition, the result of the STEM program is in turn connected to the program efficiency level and the educators improve the programme, and enhance it through assessments of the program and its evaluations (Japhet, 2019) [11]. A successful assessment helps us build the education 4.0 teaching approach for training and preparing people to solve IR4.0. There are different ways to assess the usefulness of the evaluation program for various topics in a school. The fundamental way to establish tests and ovulation of the students' learning results is to increase the reliability and validity of trials.

2. Background of the Study

According to Osadebe (2001 and 2012) [19], the schools must have ample relevant, accurate tests to assess students and prepare them for external examinations if they have covered the subject area in the curriculum. Most teachers in their various subject fields were found not to be good for constructing exams. Due to this issue, the researcher created a multi-choice, true and accurate STEM in order to improve the method assessment and improve student learning outcome.

Special expertise is required to construct a valid and accurate test. Literature is the building of the test. A instructor who constructs a test is said to be poorly prepared for a test made by a teacher. A formal test is then named if a true and accurate sample is generated by the expert. Anyway, a teacher can construct a test if correctly guided. The challenge of teachers building inadequate evaluations is a significant problem needing special attention to education. However, teachers should consult an expert before using a study. Experts suggested that the teachers should produce standard instruments for evaluating students and stressed that expert teacher should ask test questions and prepare them. So no systematic and standardized tests could even be made by experts. The development and normalization of test items are different areas for experienced individuals (Osadebe, 2012) [19]. Accouding to Osadebe (2015), the introduction of the substandard stamp in secondary school is one of the biggest problems in Nigeria. They typically use dangerous test items to assess the comprehension and mastery of the students.

An expert in test construction and standardization, the researcher decides to build for teachers a true and effective STEM item test as a field of requirement. The test will only be given to deter violence when appropriate. It should be used to evaluate students' performance after STEM learning, particularly in the fields of content. The main aim of this study was to provide teachers and others with a clear and accurate student success measure in the STEM for the purposes of the quality of the standard high school curriculum in particular Reliability and validity are very important for the development of the test item and should be taken into account. Different studies have shown that implementation of any standard test item is a key problem to the advancement of education systems.

The test format is objective and subjective, multi-choice sample objects with structural issues. The structuring planning of the test includes the identification of the targets for instruction and behaviour, the identification of the test subject areas, the decision on the test format and the setup table. The organized test planning comprises the specification of the objectives to be calculated, the identification of the field of the test material, the decision on the test format and the setup table. The table helped to assess the validity of high content (Osadebe, 2012; Ukwuije & Opara, 2012) [19, 23]. Inspection of goods and acceptance by expert opinion assessed the face validity. The behavioral goals come from the cognitive field of Bloom. Bloom is known as the cognitive method which allows students to read, understand, apply, interpret, summarize and assess higher levels of thought (Yahya, Toukal, & Osman, 2012; Clark, 2010). But the Bloom taxonomy has been revised for informing, identifying, applying, testing, assessing, and constructing. 4 out of 6 were used for the application, study, evaluation and creation of KBSM artifacts in these cognitive classes.

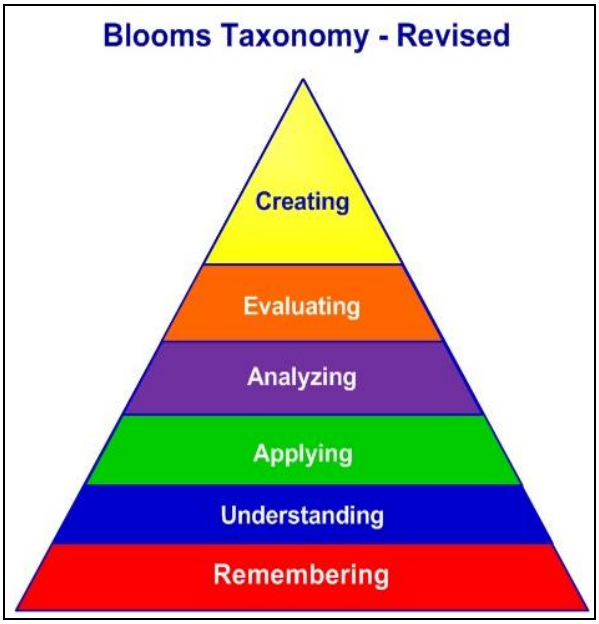

Fig 1: Taxonomy Bloom's Taxonomy Bloom's

The researcher used the entire topic Type 4 in this study since the 4 curriculum enables students to commence their chemistry learning. Students must understand the essential principles they have already mastered and their future in chemistry will become easier to understand. Based on Yunus & Ali (2013), chemistry is important to raise awareness of the chemical environment amongst students, and students must learn Form Four Chemistry curriculum before the test. According to Zwiers *et al*, greater comprehension of the students' information will diversify into a better understanding of content (2014).. A collection of HOTS items in this study are built using the Bloom Taxnonomy and validated by the subject matter expert before it is submitted to the respondent. The things can be measured using the degree of difficulty and the discrimination index. Each object is tested for reliability in order to measure the quality of the products.

3. Problem statement

Based on IR4.0, many countries have integrated the STEM program into their education system. In order to develop an education system, including the STEM programme, it is necessary to evaluate and improve the processes and the learning achievements of students on a regular basis. Validity and reliability are considered to be the main criteria for the standard test item. The aim of this study was to develop a valid and reliable STEM test item based on Higher Order Thinking Skills (HOTS) to analyze students' knowledge of application, analysis, evaluation and development.

4. Literature Review

Evaluation and assessment of the teaching process is main factor of developing the education system. If we want to enhance the education system, need to have accepted test items to evaluate the teaching process for purpose of improve learning outcome and developing the education system. In many studies (Moskal and Leydens 2000) [18] (Cook and Beckman 2006) [6]; (Cohen, Manion and Morrison 2011) [5]; emphasized the importance of validity and reliability as criteria of instrument quality has been addressed.

4.1. Validity

The term validity indicates what measurement is intended to be taken by an instrument. The validity of the instrument shall determine how well it complies with the standards laid down in order to measure the quality of the instrument using certain criteria. A low-valid tool means that the objects on the tool do not assess what the tool needs to calculate. In other words, the validity of the tool indicates how well the tool calculates what it wants to measure. Various validities of measurement, including facial validity, validity of the content and validity of the building, exist (Wong, Ong, &Kuek,2012) [25].

The measure of the simplicity of the item function is facial validity. It means that an item is correct on a high side whether its function or structure is clear to the respondents. Any experts should determine whether the instrument meets the requirements for measuring the alleged characteristics (Cohen, Manion and Morrison K 2011) [5].

Content validity reflects the evaluation and presentation of the content property (Liang *et al* (2014) [14]. In the fields of scientific, technical, engineering and mathering components and the efficacy of STEM in this study, the reliability of the content will be guaranteed. Content quality can be assessed by demanding the review of the project by content experts with the components of the stem. We asked the STEM expert on the UTM chemistry lecturer to verify the validity of the stem materials in this report. A list of the contents of the test items indicated in the draft was given to the expertsThe expert was then asked whether he accepted that each item fits appropriately with the contents proposed in the draft and if the way the items are presented tests the intended material (Cohen, Manion and Morrison 2011) [5].

The design quality is the measure of the property. Given its very significant instrument value calculation, Westen & Rosenthal (2003) [24] states that there is no clear formula available for defining the accuracy of a tool as a construction. The most common measurement used by researchers to achieve construction validity metrics is the comparison of a structural measurement with a number of other measurements potentially similar to or separate from a structure known as discrimination (Westen and Rosenthal 2003) [24]. Affiliations are centered, italicized, not bold. Include e-mail addresses if possible.

4.2. Reliability

When scientists evaluate a method that they think is consistent over time, the values they get should be coherent over time as well. The exactness of the analysis is the degree to which this is actually the case. For example, intelligence is sometimes regarded as constant over time. A very intelligent person today will be very intelligent next week. This means that any degree of good intelligence should be about the same for this person next week as it is today. Naturally, a structure that can produce highly arbitrary values over time is not a very good measure of (Miller, 2015) [17].

Most practices require a substantial assessment by an observer or a rater. The degree to which many commentators agree with their opinions is consistency. For instance, if you want to test the social skills of university students, you can catch them by video when they speak to another student they first meet. Then two or more observers should view the videos and evaluate every student's social ability level (Wong, Ong&, Kuek 2012) [25]; (Miller, 2015) [17].

The third form of reliability is internal cohesion, which means that the responses of people in the instrument are correctly measured in multiple items. In general, the underlying structure of all items on such measures should be the same, so that the scores in such testing instruments should be related together. Inner precision can only be calculated by data collection and analysis, similar to the standard of laboratory test tests. Another way is to consider a split-half partnership. It involves splitting products into two sets, such as the first and second half items and uniform and uniform numbers.

Based on Haertel (2013), the internal reliability of an item can be determined by a variety of methods, such as Cronbach's alpha, Revelle's beta, McDonald's ɯh, and Sijtsma's, The Cronbach Alpha is the most common measure used, as stated by Tang, Tui and Babenko (2014) [21]. In this analysis, the Cronbach alpha coefficient is used as a measure of the internal consistency of the instrument. The following is the mathematical statement of the Cronbach alpha coefficient (Cronbach L J 1951) [7]:

$$\alpha = \left[\frac{n}{n-1}\right]\left[1 - \frac{\sum_{i=1}^{n} \sigma_i^2}{\sigma_X^2}\right]$$

Where α is a lower-bound estimate of the true reliability, n is the number of items in test X, $\sigma 2X$ is the observed score variance of test X, and $\sigma 2i$ is the variance of item i. The criteria of interpreting an internal consistency reliability coefficient of an instrument are presented in table 2.

Table 2: validity coefficients Interpretations

Validity coefficient values	Interpretation
Above 0.35	Very beneficial (Strongly Valid)
0.21-0.35	Likely to be useful
0.11-0.20	Depends on circumstances
Below 0.11	Unlikely to be useful

5. Method

According to Lodico *et al*. (2010) [15], education research can be defined as a process for acquiring information by collecting data and solving educational difficulties and can then be used to enhance quality and practice in education. In addition, education research has several objectives that aim to improve understanding of the practices and situations that have led to such practices among educators. Based on Chua (2011) [4], the design review in the study is highly dependent on planning to conduct the research in the correct manner. For a study that allows researchers to employ the proper methodology and perfect way to conduct the study, the research design can be described as a process focusing on research study Planning is very necessary. In reality, planning concludes the method of study, sampling, research tools, validity and reliability, research and data analysis. The researcher used this study to assess the reliability and validation of STEM elements using a quantitative approach. Chua (2011) [4] has been used to research and analyze the variables in the phenomenon under study in quantitative approaches. Present study samples were used by selected Ettihad private high school high school middle school students, both for pilot tests and for actual studies. Both for real test and middle school pilot test samples of the same standard.

6. Research Objectives

The main objective of the study was to establish a valid and reliable STEM for students' application. In addition to the external evaluation training, schools must have sufficient objective and accurate evaluations in order to evaluate the students. Most of teachers in their various subject fields were found not to be good and professional for building exams items (Osadebe, (2001 & 2012) [19].

7. Research Question

This study designed to answer the following questions:

a) How is the reliability of STEM items?
b) How is the validity of STEM Item?

8. Selection of Students

The analysis was conducted using 30 selected samples from middle school students of Ettihad private high school, while the Pilot Test was conducted using 10 of these samples. The real research specimens were the same-grade students as the pilot test samples.

9. Procedure of administration

Before students did their testing, the teacher in command provided specific instructions for each portion of the exam. Over one hour, the 16-page examination was completed. The experiment is conducted at the chemical laboratory of the school.

10. Finding and Discussion

10.1 Mean, Median and Standard Deviation

By numerical estimation of the mean and median, the central tendency of the test item is calculated. The average test value received by all 30 respondents is the mean value. This is determined by extracting the cumulative scores received by the 30 students from the total of students. This equation helps us to compare each student's average scores in order to assess if they were well or poorly compared. The average score calculation is based on the following equation. This test item has an average score of 59.6.

$$\mathrm{SD} = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}}$$

Medium are the mid-test scores in order to organize all the test scores and pick the mid-test scores. This means that half test scores and half test scores are above and below the central scores. If two scores occur in the center, all scores will be combined and separated by two. The test scores are grouped in conjunction with the following report. 84, 81, 81, 78, 78, 78, 72, 72, 72, 72, 69, 69, 69, 69, 69, 66, 66, 63, 53, 53, 53, 50, 50, 47, 34, 34, 34, 31, 22, 19.

The above test scores indicate that the middle test scores are 69 and 66. The median values must then be determined by adding the values and divided by two and the median score is 67 for this test.

Standard deviation is generally used for calculating statistical variance. This standard change shows the difference between the mean score in each item. A high standard deviation results from the mean score are tabled across a wide range, whereas a low standard deviation indicates that the test score is close to that of the average score. The standard deviation is calculated with the following equation. This test score norm divide is 18.3.

$$\mathrm{SD} = \sqrt{\frac{\sum |x - \bar{x}|^2}{n}}$$

The average, median and standard variance for this test score is summarized in the following table.

Table 2: Summarized scores of mean, median and standard deviation

	Value
Mean	59.6
Median	67
Standard Deviation	18.3

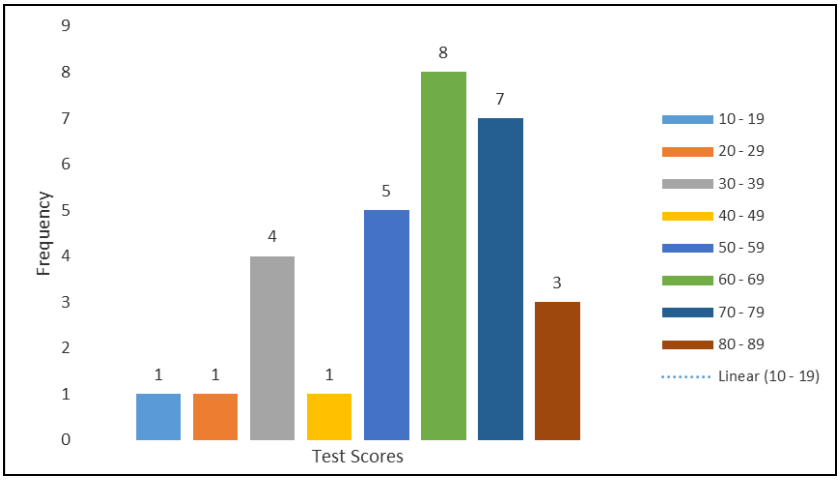

Fig 2: Mean distribution graph

The average student score was 59.6 and is based on the chart above with a standard deviation of 18.9. Students' scores range from 19 to 84 from 100. from the lowest.

10.2 Validation

In this study the questions were divided into two parts to construct the test: 15 questions from multiple choices and 5 questions from trials. All questions are therefore built on the taxonomy flourishing from the implementation, review, evaluation and growth stage to give you a higher level of thought and STEM skills. The first step is therefore to define the item specification table as well as questions for both the MCQ and the essay: This examination is validated by the Senior Lectures, School of Education, University of Teknologi Malaysia. In the study of the item evaluation and STEM with the characteristics of the HOT item, the questions were tested by the validators. See Table 4 and 5.

Table 3: The table of Item Specification for MCQ questions

No	Construct	Topic	Level	STEM
1	Application	Chemical Formulae and Equations	Low	Biology- Chemistry -Mathematics
2	Analysis	Chemical Formulae and Equations	Moderate	Biology -Chemistry
3	Analysis	Periodic Table	Moderate	Technology-Chemistry
4	Application	Periodic Table	Low	Medicine /Chemistry
5	Application	Electrochemistry	Moderate	Engineering- Chemistry
6	Application	Electrochemistry	Low	Engineering -Chemistry
7	Evaluation	Acids and Bases	Moderate	Industry -Chemistry
8	Application	Acids and Bases	Low	Engineering -Chemistry -Mathematics
9	Analysis	Salts	Moderate	Medicine -Chemistry
10	Application	Salts	High	Technology-Medicine -Chemistry
11	Application	Manufacture Substances in Industry	Moderate	Engineering -Chemistry
12	Application	Manufacture Substances in Industry	Low	Engineering -Chemistry
13	Analysis	Structure of Atom	Moderate	Technology -Medicine -Chemistry
14	Analysis	Acids and Bases	Moderate	Engineering -Chemistry Mathematics
15	Analysis	Chemical Bonds	Moderate	Engineering -Chemistry -Mathematics

Table 4: Table of Item Specification for essay questions

No	Topic	Skills												Total		
		Application			Analysis			Evaluating			Creating					
		L	M	H	L	M	H	L	M	H	L	M	H	L	M	H
1	Structure of Atom			1		1									1	1
2	Structure of Atom						1			1						2
3	Structure of Atom & Manufacture Substances in Industry				1				1					1	1	
4	Acids and Bases	1										1		1	1	
5	Structure of Atom		1			1							1		2	1
TOTAL		1	1	1	1	2	1	0	1	1	0	1	1	2	5	4

10.3 Reliability

The reliability of an instrument is characterized as the ability of the instrument to achieve a consistent test result (Kubiszyn & Borich, 2003). The alpha of KR-20 and Cronbach are used for the precision of this instrument. The CR-20 is employed to calculate the reliability of the objective variable and for counting subjective elements the Cronbach alpha. The trust range is shown in the following table.

Table 5: Interpretation of reliability coefficient (taken from University of Washington).

Reliability	Interpretation
90 and above	Excellent reliability
80-90	Very good reliability
70-80	Good for a classroom test and a few items should be improved
60-70	Somewhat low. Some item should be improved
50-60	Nod good and suggests need to for revision
50 or below	Questionable reliability. Need to revision

The Kuder Richardson (KR20) is used to calculate the reliability of critical queries. This is the KR-20 formula:

$$r = \frac{K}{K-1}\left[1 - \frac{\sum_{i=1}^{K} piqi}{\sigma 2}\right]$$

Where: K = the number of items, P= proportion pass, Q=proportion fail, σ2 = Variant.The reliability coefficient for this objective test item is 0.8 when determined on the basis of the above formula. It falls in the group of very powerful instruments for classroom studies compared to table 3.4 for calculating the reliability coefficient.

Cronbach's Alpha was calculated using SPSS code to obtain the reliability coefficient for subjective queries. For this test element, the alpha value of Cronbach is 0.77. In the figure below you can see the impact of Cronbach's Alpha.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
.774	.764	11

Fig 2: Result of Cronbach's Alpha

According to Table 7, it falls within the category of great for the calculation of the reliability coefficient, in most areas of the classroom study. Few things are possible to modify. This item is suitable on a table basis, but minor adjustments must be made in order to increase test item reliability.

Reliability testing is very important when designing instruments. Initially, reliability is computed by means of a reliability test equation. That means that the same students should be tested twice on the same instrument in two separate occasions (Webb *et al*. 2006). It doesn't look really acceptable, though. This is because people appear to adapt when the system is used on two occasions in different periods. Maybe students begin to develop new knowledge and insight by reacting to the first test, which makes the response different and better in the second attempt (Taber, 2016)

Cronbach's Alpha is also used as a metric to measure instrument performance Casanoves *et al*., 2015. It was initially developed to assess students know how and comprehension in the field of science education. Various other researchers point out the need to achieve a high degree of Cronbach Alpha while developing methods for evaluating students' knowledge in various scientific concepts (Yang *et al*., 2015). Three other research initiatives have agreed on this test of student expertise in intensity (Howe *et al*., 2014), light subjects (Mumba *et al*., 2015) and student comprehension of chemical solutions (Adadan & Savasci, 2011). The coefficient of reliability of this test item is correct for use in classroom tests, and can be accepted. This is consistent with the research done (Tuan, 2005). According to his research on the production of questionnaires to test the student's motives in science education, he said that the coefficient value between 0.87 and 0.7 was satisfactory. For several other science education jobs with instrument, the value of 0.7 to 0.9 is optimal (Eilam & Reiter, 2014; Heddy & Sinatra, 2013; and Pell & Jarvis, 2001).

11. Conclusion

In conclusion, it is very important to establish an acceptable test item based on the various studies, validity and reliability. STEM is one of the most critical parts in 21 first century education and can be produced through supervised and efficient evaluation. A true and reliable test item must be fine. KR-20 and Alpha of Crombach are considered to be the right way to evaluate the test object.

The following basic conclusions can be made after the review of the effects of the data reached:

a) A senior professor at the University of Teknologi Malaysia and an experienced professor verified the content validity of the SETM paper. The questions examined by the validators are based on the criteria of item analysis and STEM with HOT item features.
b) The subjective and logical questions concerned, all of which were trustworthy. The coefficient of reliability for this object is 0.8 and for the subjective test item, the Cronbach alpha value 0.77.

This test item can be used in any school on the basis of validation and reliability to determine and assess learners' comprehension on four topics in chemistry. oThis result allows middle school science teachers to use these items to test, evaluate and recommend to others to examine subjective items and Crombach's Alpha in objective items using ot KR-20.

12. Limitation and Recommendation

The key threshold for this analysis is random sampling of n = 30. Additional research is required in different schools to perform cluster sampling. The cluster sampling is an excellent way for respondents to be selected when a study is conducted on the basis of different population of adresses, according to Etikan *et al*. (2016). Furthermore, further analysis is needed in this study to improve the product based on other forms of reliability, such as test reliability Inter-Rater Reliability, based on the reliability evaluated on the basis of internal constancy.

13. References

1. Abdelrazeq A, Jassen D, Tummel C, Rechirt A. Teacher 4.0: Requirements of the teacher of the future in context of the fourth industrial revolution, in ICERI 2016;8:54-65
2. Aida Aryani SN. Industrial Revolution 4.0 and Education. International Journal of Acacomic Research in Bussiness and Soucial scienc 2018;8(9):314-319. doi:10.6007/IJARBSS/v8-i9/4593
3. Butz WP, Kelly TK, Adamson DM, Bloom GA, Fossum D, Gross ME. Will the scientific and technology workforce meet the requirements of the federal government? Pittsburgh, PA: RAND. Dick, S. (1980). The birth of nasa. Retrieved A 2004.
4. Chua. Kaedah Penyelidikan (Edisi Kedua). McGraw Hill Sdn. Bhd 2011.
5. Cohen L, Manion L, Morrison K Research Methods in Education (7th ed) (London: Routledge) 2011.
6. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: Theory and application The American Journal of Medicine 2006, 119-166.
7. Cronbach LJ. Coefficient Alpha and The Internal Structure of Tests Psychometrika 1951;16:297
8. Dillman DA, Smyth JD, Christian LM. Internet, Phone, Mail, and Mixed-mode Surveys: The Tailored Design Method (4th ed) (Hoboken, NJ: John Wiley and Sons) 2014.
9. Diwan P. Is Education 4.0 an imperative for success of 4th Industrial Revolution? 2017. Accessed from https://medium.com/@pdiwan/is-education-4-0-an-im-perative-for-success-of-4th-industrial-revolu- tion-50c31451e8a
10. Etkina E. Millikan award lecture: Students of physics-Listeners, observers, or collaborative participants in physics scientific practices? American Journal of Physics 2015;83:669
11. Japhet E, Lawrence. Designing a Unit Assessment Using Constructive Alignment. International Journal of Teacher Education and Professional Development 2019;**2(**1):30-51.
    Online publication date: 1-Jan-2019
12. Haertel EH. Reliability and Validity of Inferences about Teachers based on Student Test Scores (Princeton, NJ: Educational Testing Service) 2013
13. Karre H, Hammer M, Kleindienst M, Ramsauer C. "Transition towards an Industry 4.0 state of the LeanLab at Graz University of Technology". Procedia Manufacturing 2017;9:206-213.
14. Liang Y, Laua PW, Huang YW, Maddison R, Baranowski T. Validity and reliability of questionnaires measuring physical activity self-efficacy, enjoyment, social support among Hong Kong Chinese children. Prev Med Rep 2014;1:48-52.
15. Lodico MG, Spaulding DT, Voegtle KH. Methods in educational research: From theory to practice John Wiley and Sons 2010, 28.
16. Manda MI, Backhouse J. Digital transformation for inclusive growth in South Africa. Challenges and opportunities in the 4th industrial revolution. 2nd African Conference on Information Science and Technology, Cape Town, South Africa 2017.
17. Miller MJ. Graduate Research Methods 2015. Available from: http://www.michaeljmillerphd.com/res500.../reliability_and_validity.pdf. [Last accessed on 2015 Oct 10].
18. Moskal BM, Leydens JA. Scoring rubric development: Validity and reliability Practical Assessment, Research and Evaluation 2000;7:71.
19. Osadebe PU. Procedures for Construction, Validation and Standardization of Test. A Seminar Paper Presented at Delta State University, Abraka 2012.
20. Rothman KJ, Greenland S, Lash TL. Modern Epidemiology. Philadelphia, USA: Lippincott William and Wilkins. 2008, 128-47.
21. Tang W, Cui Y, Babenko O. Internal consistency: Do we really know what it is and how to assess it Journal of Psychology and Behavioral Science 2014;2:205
22. Sanders M. STEM, STEM education, STEMmania. The Technology Teacher 2009;68(4):20-26.
23. Ukwuije RPI, Opera MI. Test and Measurement for Teachers. Port Harcourt: Chadik 2012.
24. Westen D, Rosenthal R. Quantifying construct validity: Two simple measures Journal of Personality and Social Psychology 2003;84:608
25. Wong KL, Ong SF, Kuek. Constructing a survey questionnaire to collect data on service quality of business academics. Eur J Soc Sci 2012;29:209-21.