Evaluation Of General Chemistry Final Examination Of Certificate In Agricultural Sciences Students

Ruby Lyn V Gutierrez[1]
Faculty, College of Agriculture, Forestry, and Engineering
Agriculture Department, Quirino State University, Diffun, Quirino, Philippines
blueruby25@yahoo.co.nz

Abstract

Effective assessment and evaluation strategies are vital aids for facilitators to create a positive learning environment and improve student outcomes.

The study aimed to evaluate the General Chemistry Final Examination of Certificate in Agricultural Science Freshmen students in QSU CAFÉ-Agriculture Department, Diffun, Quirino. Bloom's Taxonomy and the Table of Specifications were used to evaluate the final examination's cognitive level of development. The Desirable Characteristics of the test, for its validity, reliability, objectivity, and other conditions such as administrability, scorability, interpretability, economy, utility, fairness, practicability, efficiency, and ethics in the assessment were also used. Item analysis and plausibility of the examination procedure were also done for the Multiple-choice part of the examination. The Descriptive-Normative method and content analysis were further utilized.

The findings showed that the examination is desirable. Moreover, the type of test and items of the examination were more from lower-order thinking skills. The multiple-choice type of exam was moderately difficult and low, and most of the items needed to be revised.

The study recommended pursuing a strong and practical intervention program, basically a curriculum review to establish a student-centered pedagogical approach to address the challenges.

Keywords: *Certificate in Agricultural Sciences, Desirable characteristics of test, evaluation, General Chemistry, item analysis,*

1.Introduction

Evaluation of the learning process is not an easy task; these are words that are not only said but accepted by most educators and facilitators as well. The learning process is now in many ways process. For each way, the two major ways are many crossroads: where the teacher, now the facilitator, and the center of the learning process, the student, meet. There are many struggles and challenges in gauging if learning has transpired. Does the teaching-learning process achieve what was intended? Is there a learning that takes place after all? These are the basic queries, at the same time, objectives that facilitators always address.

There is learning if there is change; a change for the better. Measuring outcomes has many approaches. Evaluation also has many approaches. If the students pass the examination, they learn something. If they have written something on their papers, they have learned something. What about if they failed? Did they not learn anything? That is why examinations are thoroughly thought out and prepared. It must be varied in form to be favorable to all types of students' levels of thinking.

Examinations should serve as an encouragement to do better the next time, to enhance what is already good. But it should not create fear in the students, making it a reason why they fail. How

the learner experiences the final examination might be made more authentic and, in the process, more constructively aligned with stated learning outcomes (Williams and Wong, 2009; Eastwood, 2024).

About 85% of students said that Chemistry subjects are difficult or boring (Gietler, 2022). The level of difficulty may be attributed to the student's background and interests in the subject (Defista and Aznam,2024). Basic chemistry concepts must have been taught in high schools due to the adapted spiral progression approach of teaching as envisioned in the K to 12 Program (Orbe et al,2018), under the Science, Technology, Engineering, and Mathematics (STEM) strand. However, not all entering the college level were within the STEM program. One of which is the ladderized program, Certificate in Agricultural Sciences (CAS). Before the Bachelor of Science in Agriculture course, a student must graduate from the CAS Program (a 2-year course). If the enrollee is not a STEM graduate, he will take bridging courses in Mathematics, Biology, and Chemistry.

Considering the heterogeneity of their high schools and the challenges of the pandemic added to the struggles of teaching sciences, herein is Chemistry and the absorption of knowledge of the students. For such reasons, connections between chemistry fundamental concepts and laboratory skills and values (Montebon,2014) must be fostered in their college learning.

The student's learning is assessed in the final examination to gauge the extent of their learning for improvement, if there is, and to maintain what is already good learning. The results of the assessment are a determinant for enhancement of the student's learning behavior, skills strengthening, and facilitators based on a better appropriate assessment (Biggs, 2001; Biggs et al, 2022).

Assessment decisions could substantially improve student performance, guide teachers in enhancing the teaching-learning process, and assist policymakers in improving the educational system (Santos, 2007; Adom et al, 2020). Thus, thorough planning and designing of the test is highly encouraged.

The above background gave the researcher reasons to embark on the study.

1.1 Statement of the Problem

This study was conducted to evaluate the General Chemistry Final Examination of the Certificate in Agricultural Sciences (CAS) Freshmen students in QSU-Diffun Campus, 1st semester SY 2021-2022.

Specifically, the study sought to answer the following questions.

1.1.1 What is the cognitive level of development of the examination using Bloom's level of Taxonomy?

1.1.2 What are the observed characteristics of the examination using the Desirable Characteristics of the test according to

1.2.1 Validity

2.1.1 Appropriateness of test items;
2.1.2 clarity of direction;
2.1.3 Reading vocabulary and sentence structure
2.1.4 Difficulty of items;
2.1.5 Construction of test item
2.1.6 length of the test;
2.1.7 arrangement of the test items; and
2.1.8 patterns of answers?

2.2 Reliability
2.2.1 nature of the test;
2.2.2 Conditions under which the test is administered;
a) physical conditions;
b) psychological factors;
c) distractions and accidents;
d) accuracy in scoring
2.2.3 administrability;
2.2.4 scorability;
2.2.5 interpretability;
2.2.6 economy;
2.2.7 utility; and
2.2.8 Other properties
a) fairness;
b) practicability and efficiency; and
c) ethics in assessment?

1.1.3. What is the level of difficulty, discrimination index, and overall remarks of the Multiple-choice type of examination?

1.2 Conceptual Framework

A conceptual framework was made to guide the researcher in conducting the study. The paradigm shows the input variables, which are the items in the final examination paper. The processes of evaluation using the criteria on the desirable characteristics of tests, item analysis on the cognitive development level, and plausibility of the examinations led to the improvement of the exam paper.

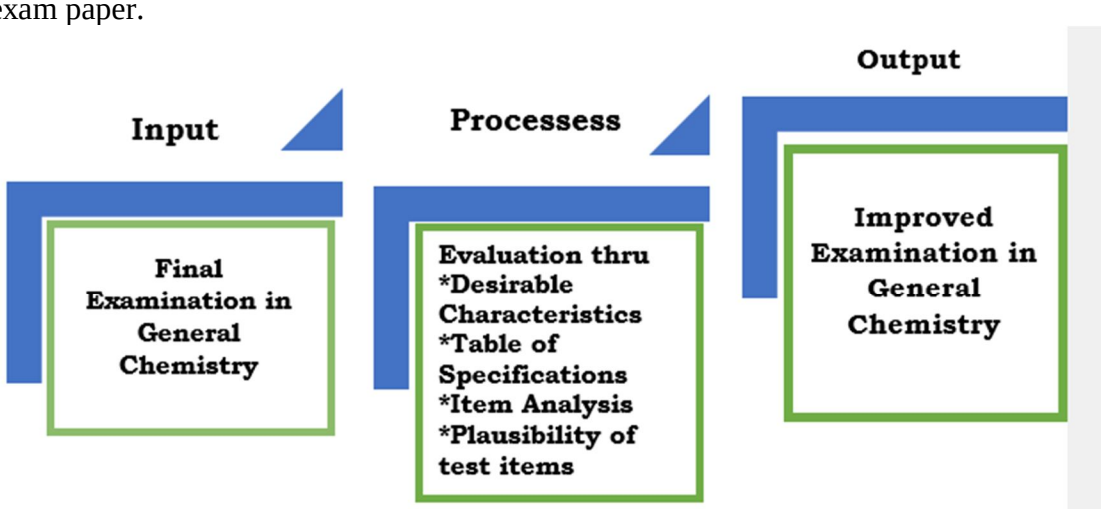

Figure 1 shows the conceptual framework of the study

Figure 1 shows the input, which is the final examination paper, which includes the Type of test I. Multiple choice, II-True or False, III-Problem Solving, IV-Balancing equations, and V-Matching Type from which these variables were evaluated according to the desirable characteristics of the tests. These are based on a) validity through 1) appropriateness of test items; 2) clarity of direction; 3) reading vocabulary and sentence structures; 4) difficulty of items (from easy to difficult); 5) construction of test items; 6) length of the test; 7) arrangement of test items; and 8) patterns of answers; b) reliability thru 1) nature of the test; 2) length of the test; 3) quality of test items/difficulty; 4) objectivity; 5) conditions under which the test is administered: physical

condition, psychological factors, and distractions and accidents; 6) scoring inaccuracy; 7) administrability; 8) scorability; 9) interpretability, economy, utility, fairness, practicability and efficiency and ethics in assessment; its taxonomic label using Bloom's level, item analyzed and determined the item plausibility bringing out the improved exam. Refer to Fig.1.

2.Research Methodology

The study used the Descriptive-normative method of research. It describes the exam and the level of questions made by the researcher for the CAS freshmen for SY 2021-2022.

Content analysis using their examination papers as sources of data.

Item analysis and Bloom's classification were also used to determine the cognitive level.

2.1. **Research Environment**

This study was conducted at the CAFÉ-Agriculture Department, Quirino State University, Diffun Campus.

2.2. **Respondents of the study**

30 CAS Freshmen students were the respondents of the study. The papers of the upper (U) 27% and the lower (L) 27% students through the item analysis method. were utilized.

2.3. Data Gathering Procedure

The final examination was conducted last first semester of SY 2021-2022. The researcher used the exam papers to evaluate the desirable characteristics of the test, plot the items using the Table of Specifications (TOS), and the items, and determine the plausibility.

2.4.Data gathering Instrument

The data gathering instrument was the examination papers of the students, Upper and Lower, 27% of the class.

2.5. **Statistical Treatment of Data**

For the taxonomic level, frequency and percentage count were used.

The item analysis method used was the U-L Method (Stocklein,1957, in Navalta, 2005; Samosa,2021).

After the item analysis, the following tables were used in interpreting the difficulty index.

Table 1. Difficulty Index

Scale	Equivalent
0.00-0.20	Very Difficult
0.21-0.809	Moderately difficult
0.81-1.0	Very easy

Table 2. Difficulty Index

Range	Equivalent	Remarks
0.41 and above	High	Very acceptable; include all
0.41-0.20	moderate	Acceptable; include all
0.19 and below	Low	Limited acceptability, revise, or discard others

3.Results and Discussion

Table 3. Evaluation of the General Chemistry Final Examination according to desirable Characteristics of the test

Type of Test	Total Items	f	%	Taxonomic Level
Multiple Choice	40	18	45	Knowledge
		11	27	Analysis
		6	15	Application
		4	10	Comprehension
True or false	10	4	40	Knowledge
		3	30	Application
		2	20	Analysis
		1	10	Comprehension
Problem-Solving	1	1		Synthesis
Balancing Equations	6	6		Synthesis
Laboratory Activities				
Matching Type	10	10	100	Knowledge
Identification	20	10	100	Analysis

The table shows that in the Multiple types of exam, out of 40 items, the majority are in the knowledge level with 45% (18), Analysis level 27.5%(11), application level with 15% (6 and 10%(4). This means that the type is within the knowledge level of cognitive development in Bloom's Taxonomy.

For the 10 items, True or False,40% (4) are within the knowledge level, 30% (3) are within the application, 20% (2) are within the Analysis, and 10% (1) are within the comprehension level. It presents again that it is at the knowledge level.

For the Problem-solving type, 100% (6) are within the synthesis level, and the last part of the exam is in the Knowledge, analysis, and application level.

It shows that the exam's level is from low thinking to higher-order thinking skills.

Final exams are designed to assess students' understanding of course material by testing their knowledge and retention of the key concepts covered throughout the course. These exams are comprehensive and include a wide range of topics and concepts covered throughout the semester or academic year.

Question formats, including Multiple choice, short answer questions, essays, and practical components, such as laboratory experiments, performances, or presentations. These practical components allow students to demonstrate their ability to apply what they have learned in a hands-on or real-world setting.

Considering that they were Freshmen, they were in the transition stage of adapting to the differences between high school exams and college exams. This also includes the level of difficulty, format, expectations for student performance, and the grading scale (Amanda, 2024; Adom et al, 2020; Alhashem et al, 2020; Boud, 2018; Boud, 2000).

Both higher-order thinking skills (HOTS) and lower-order thinking skills (LOTS) questions should be generally used for class evaluation. This approach should apply not only to final/summative assessments but also to the formative evaluations of learning. For many chemistry teachers, the mastery of computational, LOTS-type exercises is assumed as 'equivalent' to a conceptual understanding of chemistry. However, conceptual understanding requires complex upper-class thinking. It must then be accepted that instruction and examinations should contribute to the development/acquisition of HOCS [HOTS] by our students at all levels (Tsaparlis,2020; Trotter, 2006).

2. Document analysis / Test paper according to the desirable characteristics of a test
   A. Validity: which considers the following conditions
      2.1.1 Appropriateness of test items
      For Appropriateness, the chronologies of the test types were according to the topics discussed.
      2.1.2 clarity of direction
      For directions, they were clear and reiterated for the student's benefit.
      For the problem-solving part, some students were not sure of where their solutions were written or shown. However, the scores for every item should be presented after the direction.
      2.1.3 Reading vocabulary and sentence structures
      The terms used in the discussion were the terms also used in the test.
      2.1.4 Difficulty of items (from easy to difficult)
      For every test type, there was an easy-to-difficult mode, and they were presented according to the discussion of the lesson.
      2.1.5 Construction of test items.
      The test items were encoded clearly, and there were no clues used in any type of test. For the betterment of the examination, test types in Roman numeral arrangements must be corrected, and spaces where answers would be written must be enough and increased in the space for the next item.
      2.1.6 Length of the test
      The test was scheduled for 1 and a half hours, which is believed to be enough for the whole examination.
      2.1.7 Arrangement of the test items
      Every test type of the examination was from easy to difficult. The first part of the test type was on recall.
      For Problem-solving, solutions must be shown.
      2.1.8 Patterns of Answers. There were no patterns used in any type of test.
   B. Reliability as to the different criteria considered.
      2.2.1 Nature of the test.
      The final examination was for general chemistry students.
      2.2.2 The length of the test was just enough for the whole time of the exam.
      2.2.3 Quality of test items/difficulty.
      The exam was believed to be of good quality because it was
      from very easy to difficult ones.

2.2.4 objectivity.

The exam was measured objectively because there was no personal opinion presented.

2.2.5 Conditions under which the test was administered

Under physical condition. The rooms were ventilated; there was enough lighting; and the seating arrangement had been prepared before the examination. The psychological factors were left to the students who were to take the exam, who they should always ready to take the exam.

For the teacher, he was always ready for any clarifications from the students. Patience and a moderate tone of voice were used so that no disturbance would occur. Words of good luck and a sense of calmness within the room can add to the comfort of students in taking the exam.

Motivations are a great factor too in administering exams (Ryan and Deci, 2000; Ryan et al, 2021)

2.2.6 distractions and accidents.

The students were asked to bring/borrow their calculators to avoid disturbance during exams. If borrowing from another calculator cannot be avoided, it should be cleared and turned off silently before sharing it. This would prevent cheating and minimize noise. Both the students' and teachers' cell phones should be in silent mode and kept in their bags away from them. For the students to avoid cheating and passing answers, the teacher to focus on the administration of the exam.

2.2.7 Scoring Inaccuracy

Scores after each test type should be written so that students can gauge their percentage of passing or score.

2.2.8 administrability

The administration of the exam was smooth because before the students began their exam, some unclear items were already clarified to avoid any distractions.

2.2.9 scorability

The number of points for every item must be stated.

2.2.10 interpretability

Since it was on the final examination, the test papers were checked, scored, and recorded by the teacher.

2.2.11 economy

The test paper was in back-to-back pages, whereas the 2nd page back page was for solutions in the problem-solving part.

2.2.12 utility. The test was used for the final examination.

2.2.13 according to other properties.

The examination paper showed fairness. The items specify what was asked and how it must be done. It does not show any discrimination that the exam is only for fast learners, but for all types. This was shown in the chronology of the difficulty of items.

In terms of practicability and efficiency. Cheating was strictly prohibited. Queries and clarifications were recognized and addressed calmly and with minimal noise to avoid disturbance.

The checked papers were shown to students who wished to see their papers a week after the examination. Only papers were shown to their owners for confidentiality.

This examination nevertheless considered varied types of examination which give chances to different types of students' manner of taking tests (Rivera,2008).

A good test is characterized by validity, reliability, objectivity, practicality, and fairness. These key elements ensure that the test produces meaningful and accurate results, making it a valuable tool in various fields, from education to psychology. Understanding and implementing these characteristics contribute to the development of high-quality assessments that stand the test of scrutiny and provide valuable insights into individual capabilities and characteristics (Alsalihi,2019; Boud, 1995;2006).

The reliability of test scores is the extent to which they are consistent across different occasions of testing, different test editions, or different raters scoring the test taker's responses (Carlson in Nasr,2021). The improved exam would be tested again for the next set of students.

As to the study of Sambell et al (1997), validity and reliability are normally considered to be of key importance, though there are many others, such as feasibility, acceptability, and intelligibility. Through time, the concept of validity has been extended in recent years to include the effects of assessment or testing on the teaching and learning context and the social consequences of using assessment information.

Likewise, according to French et al (2022), based on their reviews of the benefits and drawbacks of high-stakes final examinations in higher education, the balance of empirical evidence poorly justifies the current heavy reliance on high-stakes final examinations in many university subjects. Consequently, low student motivation is associated with a substantial decrease in test performance (Wise and Demars,2005).

These take on assessment of the final examinations must always focus objectively on the learner and the other related factors-curriculum, teacher, and facilities for improvement.

Table 4. Item analysis of the Multiple type of exam

Item No.	Upper 27%	Lower 27%	U	L	Difficulty Indexed	Remarks	Discrimination Index	Remarks	Overall Remarks
1	8	8	1	1	1	Very Easy	0	Low	Revise
2	8	8	1	1	1	Very Easy	0	Low	Revise
3	8	8	1	1	1	Very Easy	0	Low	Revise
4	4	4	0.5	0.5	0.5	Moderately Difficult	0	Low	Revise
5	4	3	0.5	0.375	0.4375	Moderately Difficult	0.1	Low	Revise
6	4	3	0.5	0.375	0.4375	Moderately Difficult	0.1	Low	Revise
7	1	2	0.125	0.25	0.1875	Difficult	0.031	Low	Revise
8	3	4	0.375	0.5	0.4375	Moderately Difficult	0.03	Low	Revise
9	8	8	1	1	1	Very	0	Low	Revise

10	6	2	0.75	0.25	0.5	Easy Moderately Difficult	0.25	Moderate	Retain
11	8	7	1	0.875	0.9375	Easy	0.0625	Low	Revise
12	3	8	0.375	0.5	0.4375	Moderately Difficult	0.0625	Low	Revise
13	8	6	1	0.75	0.875	Easy	0.125	Low	Revise
14	7	5	0.75	0.625	0.6875	Moderately Difficult	0.0625	Low	Revise
15	5	5	0.625	0.625	0.625	Moderately Difficult	0	Low	Revise
16	5	6	0.625	0.75	0.6875	Moderately Difficult	-0.0625	Low	Revise
17	3	6	0.375	0.75	0.5625	Moderately Difficult	-0.1875	Low	Revise
18	2	6	0.25	0.75	0.5	Moderately Difficult	-0.25	Low	Revise
19	4	5	0.5	0.625	0.5625	Moderately Difficult	-0.625	Low	Revise
20	5	3	0.625	0.375	0.5	Moderately Difficult	0.125	Low	Revise
21	5	3	0.625	0.375	0.5	Moderately Difficult	0.125	Low	Revise
22	6	3	0.75	0.375	0.5625	Moderately Difficult	0.2	Moderate	Retain
23	6	4	0.75	0.5	0.625	Moderately Difficult	0.125	Low	Revise
24	1	7	0.75	0.875	0.8125	Easy	0.0625	Low	Revise
25	8	3	0.125	0.375	0.25	Difficult	-0.125	Low	Revise
26	8	2	1	0.25	0.625	Moderately Difficult	0.375	Moderate	Retain

27	8	2	1	0.25	0.625	Moderately Difficult	0.375	Moderate	Retain
28	8	7	1	0.875	0.9375	Very Easy	0.0625	Low	Revise
29	8	7	1	0.875	0.9375	Very Easy	0.0625	Low	Revise
30	6	1	0.75	0.125	0.4375	Moderately Difficult	0.3125	Moderate	Retain
31	4	1	0.5	0.125	0.3125	Moderately Difficult	0.1875	Low	Revise
32	5	4	0.625	0.5	0.5625	Moderately Difficult	0.0625	Low	Revise
33	7	1	0.875	0.125	0.5	Moderately Difficult	0.375	Moderate	Retain
34	8	2	1	0.25	0.625	Moderately Difficult	0.125	Low	Revise
35	8	7	1	0.875	0.9375	Very Easy	0.0625	Low	Revise
36	8	3	1	0.375	0.6875	Moderately Difficult	0.3125	Moderate	Retain
37	6	2	0.75	0.25	0.5	Moderately Difficult	0.25	Moderate	Retain
38	6	3	0.75	0.375	0.5625	Moderately Difficult	0.2	Moderate	Retain
39	7	5	0.975	0.625	0.8	Easy	0.125	Low	Revise
40	1	2	0.125	0.25	0.1875	Difficult	-0.0625	Low	Revise

After calculating the difficulty index using the item analysis method, Table 4 shows that there are 26 moderately difficult items; 7 are *very easy,* 4 are *easy,* and 3 are *difficult.* It implies that the majority of the items are *moderately difficult*.

For the discrimination Index, 31 of the items are *low,* and 9 are *moderate.* This would mean that most of the items are considered for *revision.*

The findings showed the same results in the study of Musa et al 2024) that teachers generally found Chemistry concepts to be moderately difficult. However, students perceived these concepts as difficult, with a significant difference in their perceptions. Although exams are difficult,

which results in stress, they can aid in better marks (Shean, 2019; Wulandri et al, 2023) when results are constructively accepted (Van Bergen and Lane, 2014).

A strong recommendation is given by Langitasari et al (2024) in their SiMaYang Learning Approach, whereby students' conceptual understanding of Chemistry needs to be supported by the skill to think at three levels of representation. Chemistry teaching needs to provide a learning environment that involves chemical representations to facilitate students' interconnection skills of three levels of representation and ultimately improve students' conceptual understanding. Many factors, though, affect performance in exams (Wolf and Smith, 1995).

Perceptions of the difficulty of learning in different aspects can be answered through collaborative learning applications and a blended approach to learning. They can be used to design and support assessment activities that increase levels of student engagement with course concepts, their peers, faculty, and external experts, leading to increased student success and satisfaction (Vaughn, 2014), and continuous constructive and curriculum feedback would be effective (Winstone and Caress,2020; Wong et al,2020).

Table 5. Plausibility of the items in the Final Examination

Item No.	U/L	Answers A	B	C	D
1	U	8			
	L	8			
2	U		8		
	L		8		
3	U	8			
	L	8			
4	U			4	4
	L			4	4
5	U			4	4
	L			4	4
6	U			4	4
	L	1		3	4
7	U	2	3	2	1
	L		3	2	2
8	U	3	3		3
	L	4		1	3
9	U			8	
	L			8	
10	U	6	2		
	L	2	4	2	
11	U		8		
	L	1	7		
12	U		5		3
	L		2	2	4
13	U		8		
	L		6	1	1
14	U	7	1		
	L	5	3		
15	U	2	5	1	

Item No.	U/L	Answers A	B	C	D
21	U	5	3		
	L	4	3	1	
22	U	8			
	L	2	6		
23	U	2	6		
	L	4	4		
24	U	2	6		
	L	1	7		
25	U	1	7		
	L	3	5		
26	U		8		
	L	6	2		
27	U	8			
	L	2	6		
28	U	8			
	L	1	7		
29	U	8			
	L	7	1		
30	U		2		6
	L	3	4		1
31	U	3	1		4
	L	3	4		1
32	U	5	1	1	1
	L	5			3
33	U		1	7	
	L	6	1	1	
34	U	8			
	L	2		6	
35	U		8		

	L	1	5	1	1		L		7	1	
16	U	2		5	1	36	U		8		
	L		1	7			L	3	3		2
17	U	3	5			37	U	6		2	
	L	6	2				L	1	3	3	
18	U	6	2			38	U	1		6	1
	L	2	6				L		1	4	2
19	U		4	4		39	U	7		1	
	L		2	5	1		L	5	2		
20	U	3			5	40	U		4	3	1
	L	1	1	2	4		L	1	1	3	2

*U-Upper 27%
L-Lower 27%

From the table above, from items numbers 1-3 (refer to Appendix C), both the Upper(U) and lower (L) 27% of the class got the correct answer.

For items 4-6, only half of the U and L got the correct answer.

In items 7 and 8, the student's answers vary. Only 3 got the correct answer for no. 7 for U but almost half of the L got the correct answer.

For item no. 9, both U and L got the correct answer. For no. 10, most of them got the correct answer, only 2 got the correct answer for U and 4 L.

For no. 11, all of U got the correct answer but only 7 from U. For no.12, only 3 of U got the correct answer but most of them got the wrong answer. However, for L half got the correct answer.

For no. 13, all U got the correct answer but only 6 got it from the L.

For no. 14, 8 of U got the correct answer and 7 of L got it right.

For no. 15, most of both U and L got the correct answer.

For no. 17, 3 of U only got the correct answer and for L 6 got it right.

For no. 18, only 2 of U got the correct answer, and 6 for L.

For no. 19, 50% of U got the right answer and 5 for L.

For no. 20, 5 of U got it, and 4 for L.

For no. 21, the majority of U and L got the correct answer and for no. 22 all of U got the correct answer and 6 of L got the wrong ones. This means that the item was mistaken by the L.

For no. 23, 6 of U got it right but half of L got the correct answer.

For no. 24, both U and L got it correct.

For no. 25, most of them got the wrong answer. 7 of U and 5 L. This implies that most of them were not able to analyze the item correctly.

For no. 26, all of U got the correct answer but only 2 got the correct one. This means that the item is a difficult one for L.

For no. 27, all of U but only 2 of L got the correct answer.

For no. 28, no one got the correct answer for U but 7 of L got it correct. This item is very easy maybe for the L, but not for U.

For no. 29, all of U got it correctly but only 7 got it from L.

For no. 30, 6 of U got the correct answer but only 4 of L got it.

For no. 31, only half of U got the correct answer but only 1 from L got it. This shows that this item is difficult for L.

For no. 32, 5 of U got it correctly both for U and L.

For no. 33, 7 of U got it but only 1 got it from L.

for no. 34, all of U got it correctly but only 2 from L got it.

For no. 35, both U and L got it correctly.

For nos. 36-40, this was on the analysis of graphs based on the gas laws. All of U got it correctly but only 3 of L got the correct answer.

For no. 37, 6 of U got it but only 3 for L.

For no. 38, 6 of U got it correctly.

For no. 38, 6 of U got it correctly but only 4 from L.

For no. 39, most of them got the correct answer but only 4 of them got it, and from no. 40, most of them did not get the correct answer.

The above findings show that there were items which are easily understood by both the U and L. There were also items that the U could analyze readily, but not for L. Likewise, some items were easy for L but not for U.

This was a piece of evidence that individual students have differences in the cognitive, psychomotor, and affective levels of understanding as shown in their solving various items. Taking into consideration too that they came from different strands in their senior high schools.

The students' low comprehension of chemistry concepts can be attributed to interest in the subject, comprehension level, listening and problem-solving skills, interpretation of phenomena, and the connection of the concepts in daily life (Wulandri et al, 2023); these are only some of the various reasons (Defista and Aznam, 2024).

Given the above reasons, several studies offer various effective means of solving the challenges. One of which is structured inquiry-based learning significantly enhances students' conceptual understanding and academic performance in Chemistry (Emborgo et al,2024); Technology-assisted chemistry instruction (Woldemariam,2024; Agrahari et al,2013); web-based chemistry learning (Frailich et al,2007), and many others. All of them are on the path of achieving deepened chemistry learning, but above all, these are the main objectives to understand the learner and provide open, agreed, and constructive feedback for everybody’s perusal and betterment.

3.Conclusions And Future Works

The following were the conclusions:

1. The types of tests and items of the examination were more from knowledge level rather than higher-order thinking skills.
2. The examination is desirable according to the criteria of desirable characteristics of the test. However, the construction of test items, clarity of the directions, and accurate scoring should be given attention.
3. The multiple-choice type of exam was moderately difficult and low, and most of the items were revised.

With the conclusions, the following are recommended:

1. Facilitators of learning should increase the number of items for higher-order thinking skills. Since the students are at their college level being molded as professionals.
2. The improved examination paper should be administered again for validity and reliability of the exam to the next group of students (considering pre-test post-test and re-evaluation).
3. The above discussions illustrate that this study is an avenue for more research with other variables and to pursue a strong and practical intervention program: curriculum review to establish a student-centered pedagogical approach by the department.

4.Acknowledgment

The researcher expresses her deepest appreciation to Quirino State University, the CAS Students, and their FAMILIES for all the support. To the TRUEST GOD, be the GLORY.

5. **Conflict of Interest:** The authors have no conflict of interest to disclose.

6. **Ethical Issues:** None.

7.References

Adom D,Mensah JA, and Dake DA.(2020).Test, Measurement, and Evaluation: Understanding and use of the concepts in education. International Journal of Evaluation and Research in Education (IJERE) Vol. 9, No. 1, March 2020, pp. 109~119 ISSN: 2252-8822, DOI: 10.11591/ijere.v9i1.20457

Agrahari, A., & Singh, S. (2013). The impact of Information and Communication Technology (ICT) on achievement of students in chemistry at secondary level of CBSE and UP Board in India. International Journal of Science and Research, 2(8), 126-129.

Alsalihi,HDA.(2019).Characteristics of a Good Test. https://www.researchgate.net/publication/336891350_Characteristics_of_A_Good_Test http://dx.doi.org/10.13140/RG.2.2.21716.76169. *https://alchemlearning.com/what-are-the-main-characteristics-of-a-good-test/*

Alhashem, F & Agha, N. (2020). Analysis Based on the Three Objective Educational Domains for Final Summative Secondary Examinations of Science Subject (Chemistry, Physics, and Biology). Education Research International. 2020. 10.1155/2020/8886126.

Amanda.2024.https://educatetree.com/why-are-final-exams-important/

Biggs, J. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. Higher Education, 41(3), 221–238. https://doi.org/10.1023/A:1004181331049

Biggs, J. B., Tang, C. S., & Kennedy, G. (2022). Teaching for quality learning at university (5th ed.). Open University Press.

Boud, D. (Ed.). (2018). Developing evaluative judgment in higher education. Routledge.

Boud, D. (2000). Sustainable assessment: Rethinking assessment for the learning society. Studies in Continuing Education, 22(2), 151–167. https://doi.org/10.1080/713695728

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. Assessment and Evaluation in Higher Education, 31(4), 399–413. https://doi.org/10.1080/02602930600679050

Boud, D. (1995). Enhancing learning through self-assessment (1st ed.). Routledge. https://doi.org/10.4324/9781315041520

Defista, C & Aznam, N. (2024). Chemistry-Focused Conceptual Understanding Research Trends: A Systematic Review.Bahasa Inggris. Jurnal Penelitian Pendidikan IPA. 10. 180-187. 10.29303/jppipa.v10i4.6534

Eastwood,B.(2024). Importance of Student Assessment and evaluation.https://www.planitteachers.ai/articles/the-importance-of-student-assessment-and-evaluation

Emborgo,IHB et al.(2024). Improving Students' Conceptual Understanding of Chemistry Through Structured Inquiry based Learning. https://doi.org/10.5281/zenodo.11480193

Frailich, M., Kesner, M., & Hofstein, A. (2007). The influence of web based chemistry learning on students' perceptions, attitudes, and achievements. Research in Science & Technological Education, 25(2), 179–197.

French, S., Dickerson, A. & Mulder, R.A. (2022).A review of the benefits and drawbacks of high-stakes final examinations in higher education. High Educ 88, 893–918 (2024). https://doi.org/10.1007/s10734-023-01148-z\

Gietler,S.(2022). We need to talk about Chemistry. https://edu.rsc.org/opinion/we-need-to-talk-about-chemistry/4014985.article

Langitasari,I,Aisyah RS,Parmandhana RN and Nursaadah E.2024.Enhancing Students' Conceptual Understanding of Chemistry in SiMaYang Learning Environment. Researchgate.net/publication/380132960_Enhancing_Students'_

Musa, H & Onu, D (2024). Assessment of Difficult Chemistry Concepts among Teachers and Students in Colleges of Education in North West, Nigeria. 4. 17-30. 10.5281/zenodo.11074008.

Montebon,DRT.(2014).K12 Science Program in the Philippines: Student Perception on its implementation. https://www.ijern.com/journal/2014/December-2014/15.pdf

Navalta, JD.(2005). Teaching Guide in Educ 512 Advanced Test and Measurements. Graduate School. NVSU, Bayombong, Nueva Vizcaya.

Nasr,Y.2021.7 Key Characteristics of a good test in education in 10 minutes. https://qorrectassess.com/en/blog/characteristics-of-good-test-in-education/

Orbe JR,Espinosa AE and Datukan JT.(2018).Teaching Chemistry in a Spiral Progression Approach: Lessons from Science Teachers in the Philippines. *Australian Journal of Teacher Education*.http://ro.ecu.au/ajte/vol43/iss4/2

Rivera,DV.(2008). Evaluation of the types of tests Science Teachers made at ELC-Bambang, Nueva Vizcaya.

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, 25(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Ryan, T., French, S., & Kennedy, G. (2021). Beyond the Iron Triangle: Improving the quality of teaching and learning at scale. *Studies in Higher Education,* 46(7), *1383–1394. https://doi.org/10.1080/03075079.2019.1679763*

Santos,RG.2007.Advanced Methods in Educational and Evaluation: Assessment of Learning 2. http://opac.urs.edu.ph/cgi-bin/koha/

Sambell, K., McDowell, L., & Brown, S. (1997). "But is it fair?": An exploratory study of student perceptions of the consequential validity of assessment. Studies in Educational Evaluation, 23(4), 349–371. https://doi.org/10.1016/S0191-491X(97)86215-3

Samosa, R. (2021). Item Analysis Using the U-L Index Method. https://www.researchgate.net/publication/353750829_Item_Analysis_U

Shean, M. (2019). Don't calm down! Exam stress may not be fun, but it can help you get better marks. The Conversation. https://theconversation.com/dont-calm-down-exam-stress-may-not-be-fun-but-it-can-help-you-get-better-marks-124517.

Tsaparlis, G. (2020). Higher and lower-order thinking skills: the case of chemistry revisited. *Journal of Baltic Science Education,* 19(3), 467-483. https://doi.org/10.33225/jbse/20.19.467

Trotter, E. (2006). Student perceptions of continuous summative assessment. Assessment and Evaluation in Higher Education, 31(5), 505–521. https://doi.org/10.1080/02602930600679506

Van Bergen, P., & Lane, R. (2014). Exams might be stressful, but they improve learning. *The Conversation*. https://theconversation.com/exams-might-be-stressful-but-they-improve-learning-35614.

Vaughan, N. (2014). Student engagement and blended learning: Making the assessment connection. *Education Sciences*, 4(4), 247–264. https://doi.org/10.3390/educsci4040247

Williams, J. B., & Wong, A. (2009). The efficacy of final examinations: A comparative study of closed-book, invigilated exams and open-book, open-web exams. *British Journal of Educational Technology*, *40(*2), 227–236. https://doi.org/10.1111/j.1467-8535.2008.00929.x

Winstone, N. E., & Carless, D. (2020). Designing effective feedback processes in higher education: A learning-focused approach. Routledge, Taylor & Francis Group.

Wise, S. L., & DeMars, C. E. (2005). Low examinee effort in low-stakes assessment: Problems and potential solutions. *Educational Assessment, 10(1),* 1–17. https://doi.org/10.1207/s15326977ea1001_1

Wolf, L. F., & Smith, J. K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, 8(3), 227–242. https://doi.org/10.1207/s15324818ame0803_3

Woldemariam, D.Y., Ayele, H.S., Kedanemariam, D.A. et al.(2024).. Effects of technology-assisted chemistry instruction on students' achievement, attitude, and retention capacity: *A systematic review. Educ Inf Technol* **29**, 13763–13785 https://doi.org/10.1007/s10639-023-12411-2

Wong, H. M., Kwek, D., & Tan, K. (2020). Changing assessments and the examination culture in Singapore: A review and analysis of Singapore's assessment policies. *Asia Pacific Journal of Education,* 40(4), 433–457. https://doi.org/10.1080/02188791.2020.1838886

Wulandari, D & Roza, D & Rangkuti, M & Tanjung, Y & Ramadhani, I. (2023). The level of understanding of thermodynamic concepts for physics and chemistry undergraduate students. *Jurnal pendidikan fisika.* 12. 1. 10.24114/jpf.v12i1.42330