International Journal of Latest Research in Humanities and Social Science (IJLRHSS)
Volume 07 - Issue 02, 2024
www.ijlrhss.com || PP. 72-80

A Descriptive Study on the Three Dimensions of Language Assessment among ESL Teachers

Charlene J. Belmonte

A Research Paper in IMRD Format
Presented to the Faculty of Philippine Normal University in Partial Fulfillment of Requirements for the Degree of Master of Arts English Language Education

Introduction

Park (2018) argues that a well-written test is an important thing that exam takers will receive from test developers; to motivate them to fully exhibit what they can conceptually process and what they can execute in real-life, realistic, and interactive test content and scenarios are used.

Danuwijaya (2018) found that the first step to better test item quality is to make sure items were created with the test's intended purpose in mind and by an experienced test writer. It's conceivable that Filipino students weren't exposed to the types of questions and objects used on the PISA reading comprehension test, which may explain why they scored so poorly on the test.

According to Parris, Estrada et. al (2016), the 21st century's digital approach to teaching English compared to its traditional equivalent, age uses different instructional strategies, incorporates different technologies, and has different focuses. Teachers in the digital age use ICT tools to develop alternative assessments with the goal of assessing "the student's ability to perform a task based on a learning outcome," as opposed to teachers in the traditional classroom who design English language tests to gauge the amount of knowledge that their students have accumulated. Moreover, they stress that by reducing the anxiety level frequently associated with language production for second language learners, participation becomes inclusive for all levels of English-language learners (ELL) and positively influences their levels of motivation and self-confidence as well as classroom engagement and ultimately language learning. By using online response tools in the classroom, instructors can see how well their students understand concepts and give each student a variety of ways to participate in class. To evaluate student performance on these tasks considering various learning goals, teachers can also use various project-based language learning tasks that encourage student participation. Task-based and project-based learning encourage language acquisition, boost student-to-student communication, and give students the chance to work collaboratively with their peers for extended periods of time while examining and addressing a complicated question, problem, or challenge from the inside out.

As stated by Sutherland (2017), a genuine digital age learning environment also encourages teachers to "think outside the box" when developing and implementing alternative digital age instructional approaches, as well as when coming up with assessment techniques that allow students to show their knowledge and skills in relation to the 4 C's, which stand for critical thinking, communication, collaboration, and creativity (Sutherland, 2017).

Gottlieb (2016) goes even farther to stress the value of taking learners' choices into account even as the exams are being designed. He describes evaluation as a special framework with three components as follows:

Assessment as learning, in which he emphasizes the important role that learners' self-assessment and reflection play in the development of their autonomy and responsibility for their own learning through setting their own goals, keeping track of their progress, and reflecting on it, selecting the desired assessment task projects, and using and maintaining a student portfolio.

Assessment for learning, which is comparable to formative assessment, because it enables teachers to regularly gather data on students' progress to guide instruction and give students the necessary comments and feedback.

Standardized tests, as well as unconventional performance-based or project-based evaluations, are all examples of summative assessments of learning. Instead of relying solely on conventional testing formats that neither resemble the technology-enhanced instructional approaches adopted inside the classroom nor effectively reflect the actual skills and proficiency of the English language learners, alternative assessments that incorporate ICT allow English teachers to evaluate, improve, and redirect their instruction in ways that can respond to the needs of their students.

Assulaimani (2019) added that ICT-based summative and formative assessment methods should be used in conjunction with "other forms of alternative assessments, e.g. students' portfolios and task- and project-based language learning, may render as an even more valid and reliable way of assessing the students' achievement and proficiency in a second language course," the author emphasizes. Innovative English teachers now have the

chance to develop, customize, and validate more cutting-edge and effective tools of language learning assessment thanks to the integration of ICT in new alternative methods of assessment that make it easier to measure students' performance that traditional exams fail to test.

Editors (2019) explained that a student's skills and conceptual comprehension are evaluated through performance tests, which are accurate simulations of real-world situations, and by creating a genuine product, students demonstrate their understanding of a skill. Realistic simulations may be used in authentically created performance assessments as long as they accurately reflect real-world circumstances. They are rated on particular criteria and have clear objectives. With the use of these tests, ELL and non-ELL students alike are given different opportunities to showcase their capabilities in a regular classroom setting. Performance tests frequently require higher-level thinking, and if they are not carried out with clear instructions and expectations, they may overwhelm ELL pupils. The fact that teachers have varying expectations makes it challenging to compare this type of assessment. Performance evaluations could be regarded as judgmental assessments since teachers may evaluate the outcomes based on previous beliefs about each student's work ethic and aptitude. A rubric should be in place to ensure that performance assessments are stored properly. This type of evaluation should be used to gauge a student's proficiency in the vocabulary standard being evaluated.

Grimes-Hillman et. al (2014) stated that writing, reading, and verbal evaluation should all be included in multiple measure examinations in addition to pencil and paper. ELL students frequently fail to combine these three skills when speaking, writing, or reading in the second language. To help ELL students demonstrate their knowledge, all three topics are examined in a multiple measure assessment. Giving ELL pupils multiple-measure tests gives them the chance to demonstrate their knowledge in various ways. When conducting this type of examination, teachers can precisely identify a student's conceptual strengths and weaknesses. Teachers can score multiple factors on multiple measure tests to determine the understanding of ELL pupils.

The latest assessment conducted by Program for International Student Assessment (PISA) showed that 15 years old Filipinos average reading performance, Mathematics, and Science scores is among the lowest in all PISA-participating nations in which the Philippines ranked 76 over 76 countries in 2018 (Education GPS - Philippines - Student Performance (PISA 2018, 2022).

Nonetheless, the evaluation has also indicated that the overall number of learning activities in the Philippines was among the highest, as the Philippines ranked first out of 63 nations for the largest average number of learning activities.

The research gap of this study was focus on bombarded activities given by the teachers, yet the annual report of scores is still in a low rate therefore there is a big possibility that the student is not interactive in the class discussion, and for some instances, teacher doesn’t meet the trends and achieve the standard when it comes to creating language assessments.

Specifically, this study sought to answer the following questions:

1. **What is the profile of the respondents in terms of:**
   1.1. Age
   1.2. Length of Service
   1.3. Educational Attainment
   1.4. Designation
2. What is the level of competency on the three dimension of language assessment of the ESL Teachers in terms of:
   2.1. Skills
   2.2. Knowledge
   2.3. Principles
3. Based on the findings, what are the recommendations that can be formulated?

Methods

This study uses a quantitative descriptive research method to determine the three dimensions of language assessment among ESL Teachers. Descriptive research is a type of study designed to give a picture of the people or events under investigation. This descriptive approach emphasizes the "what" of the study topic above the "why" (Bhat, 2022). Descriptive research is designed to make sure that there is no bias in the collection of data and at the same time to reduce errors in interpreting the data collected. Usually, the researchers use this kind of research to obtain firsthand data from the respondents for them to formulate conclusions and recommendations for the study. Specifically, the researchers use descriptive survey research design in this study.

In this study, purposive sampling technique was used to select the participants.The information was gathered through the help of the respondents who actively answered the survey questionnaire, which are the 33

public English teachers in Junior High School from different schools' division in the Philippines, to determine the efficiency of the three dimensions of language assessment among the ESL Teachers.

The following statistical treatment and tools are used to interpret the data gathered.

Percentage - the study determined the proportion of respondents for each category in the profile.

Formula: Percentage = F/N x 100%

Where: F – frequency

N – Population

Frequency - This is used to determine the number of respondents based on profile.

Weighted Mean - This was used by the researchers to determine the descriptive Study on the Three Dimensions of Language Assessment among ESL Teachers.

Formula: WM = Σfx/N

Where: WM – weighted mean;

N – Sample size;

Σfx – sum of products of the frequency with weights

Results

I. Personal Profile

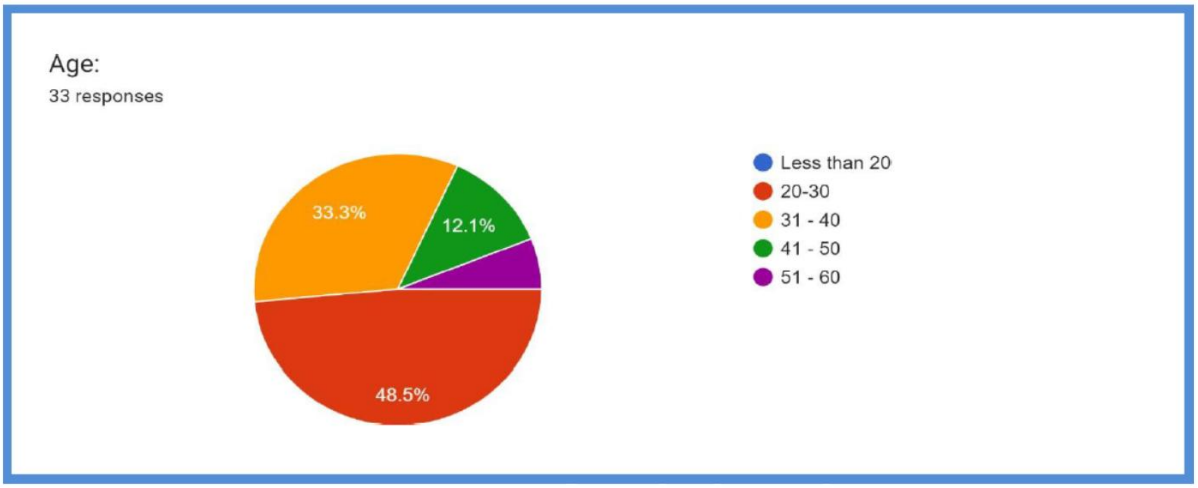

Figure 1 Percentage of respondents based on their ages

This statistic reveals that 48.5 percent of English as Secondary Language (ESL) teachers are between the ages of 20 and 30, 33.3 percent of respondents are between the ages of 31 and 40, and 12.1 percent of respondents are between the ages of 41 and 50. This merely indicates that 6.1% of responders may be between the ages of 51 and 60. Thus, majority of the respondents aged 20-30.

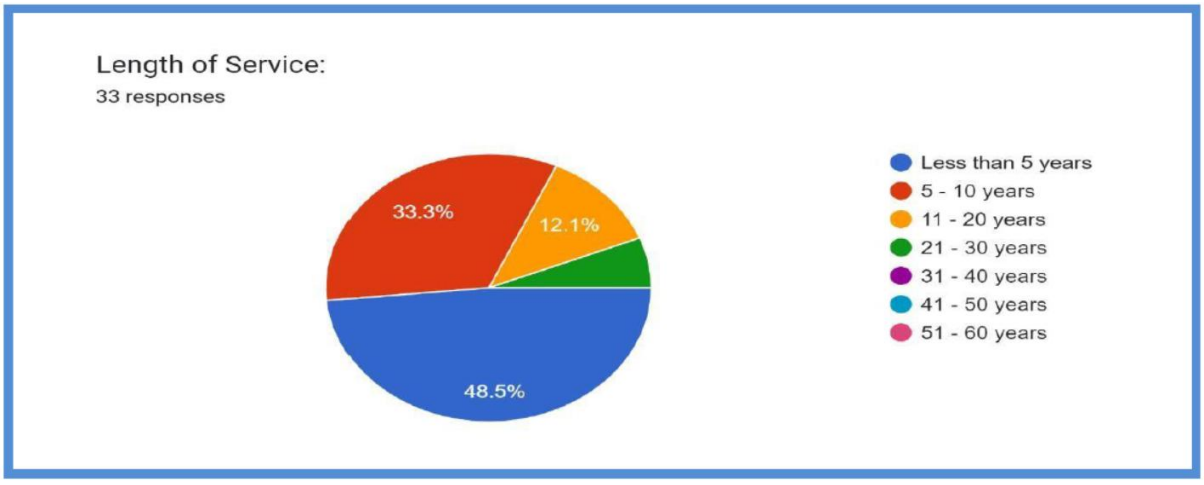

Figure 2 Percentage of respondents based on their length of service

Figure 2 displays that 48.5% of respondents had taught for fewer than five years in a public-school setting. About a third (33.3%) had taught for five to ten years, another 12.1 % for eleven to twenty years, and 6.1 % for twenty-one to thirty years. Hence, majority of the teachers who answered the survey questionnaire have less than five years' experience in teaching and implies that most of the respondents are still neophytes in the field of teaching.

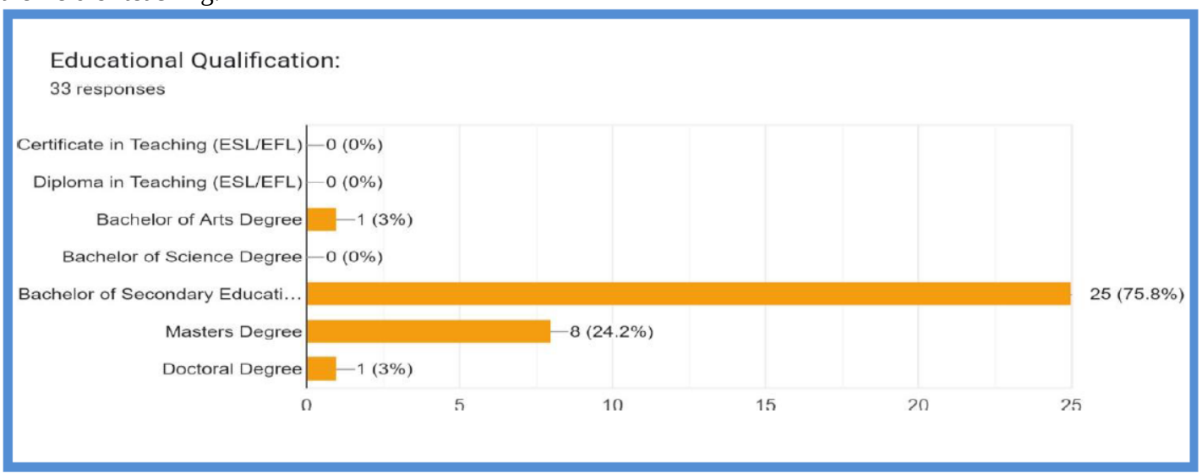

Figure 3 Percentage of respondents based on their educational attainment

Figure 3 indicates that there are 75.8 percent of respondents who have Bachelor of Secondary Education degree, next is master's degree that has 24.2 percent followed by Bachelor of Arts and Doctoral Degree which have 3 percent. This only means that among the ESL teachers who gave responses to the survey questionnaire larger population are graduates of Bachelor of Secondary Education which are still exploring and taking every step slowly in the teaching industry.

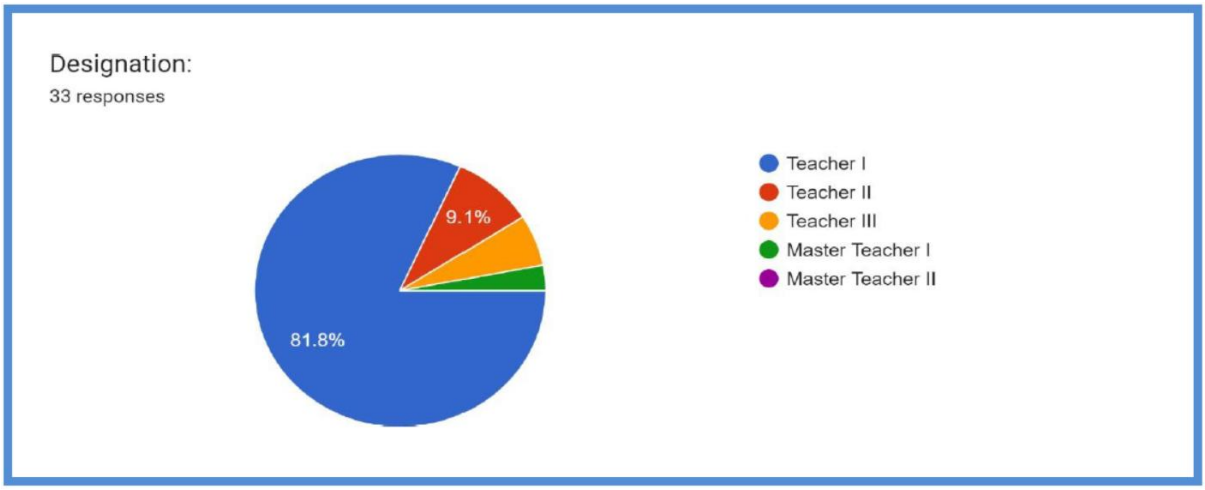

Figure 4 Percentage of respondents based on their position or designation

Figure 4 shows that 81.8 percent of the respondents are designated as Teacher I. On the other hand, 9.1% of the respondents are Teacher II, 6.1% of the respondents are Teacher III and only 3% of the respondents hold Master Teacher I position. Therefore, most of the respondents hold Teacher I positions or Proficient Teachers and only a selected few are considered Highly Proficient Teachers (Teacher III to Master Teacher positions).

II. Language Assessment Literacy Questionnaires

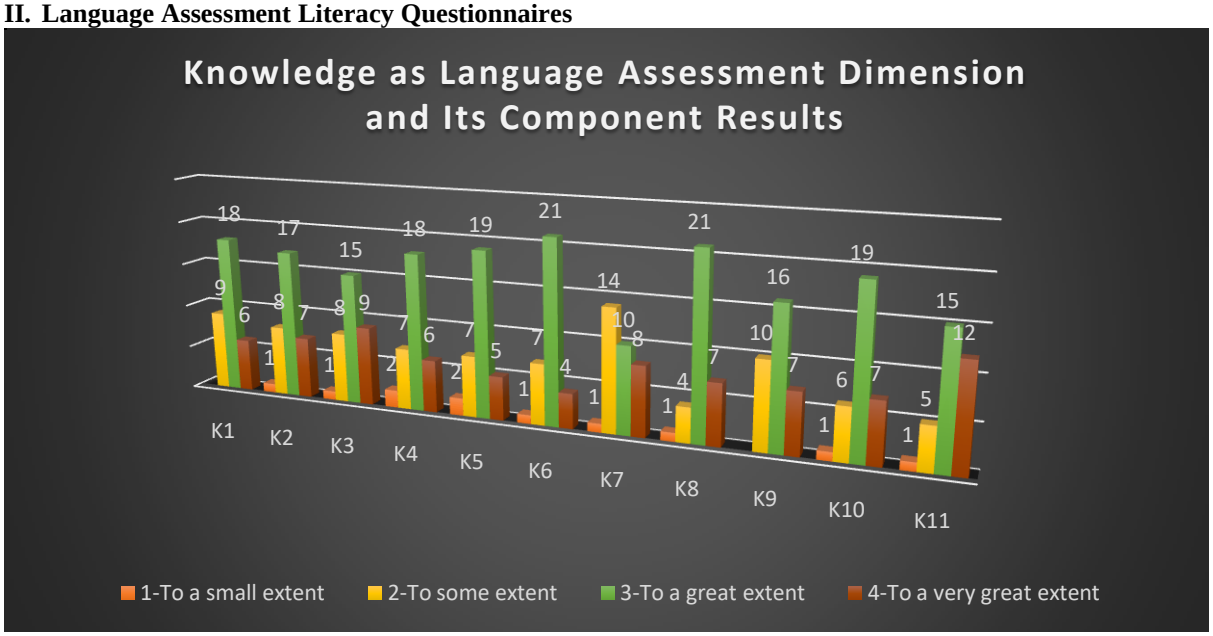

Figure 5 Number of respondents who answered for each knowledge components

No.	Dimensions: Knowledge	Mean Score of 33 Respondents
1	I can compare approaches for language teaching and assessment, e.g., communicative language testing; task-based assessment	2.91
2	I can analyze trends in second language acquisition and their impact on language assessment, e.g., motivation, cross-linguistic influence, and learne	2.91
3	I can interpret reliability, validity, authenticity, and practicality in language assessment and its implications.	2.97
4	I can differentiate concepts related to assessment paradigms: traditional. I versus alternative; norm-referenced and criterion-referenced testing.	2.85
5	I can differentiate major purposes and related decision-making for language testing: placement, achievement, proficiency, etc.	2.82
6	I can explain major steps in developing tests: test purpose, construct definition, content specifications, test specifications, etc.	2.85
7	I can evaluate the kind of washback that assessments can have on learning, teaching, context, curricula, and institutions.	2.76
8	I can contrast assessment methods, with their advantages and disadvantages; tests, portfolios, performance assessment, self-, and peer-assessment	3.03
9	I can articulate the nature, purpose, and design of scoring rubrics; for example, holistic and analytic	2.91
10	I can recognize what feedback implies within a formative assessment paradigm.	2.97
11	I can explain my own beliefs, attitudes, context, and needs for assessment	3.15
	Average of Respondents for Knowledge	2.92 To a great extent

Figure 6 Implementation of Excel formulas for calculating the mean score on respondents' knowledge

Figure 5 shows each 11 statements has a verbal interpretation based on the description of Likert-scale representation:
0-1.49 = to a small extent
1.50 – 2.00 = to some extent
2.10 – 3.49 = to a great extent
3.50 – 4.00 = to a very great extent

K stands for **Knowledge** and **numbers 1-10 have equivalent components that can be seen in Figure 6**. Primarily, K1 with a mean score of 2.91 and a verbal interpretation of to a great extent supported what Parris, et al. (2016) said: that task-based and project-based learning help students learn languages, improve communication between students, and give students the chance to work with their peers for long periods of time while looking at and solving a difficult question, problem, or challenge from the inside out. Next, K2 with a mean score of 2.91 and a verbal interpretation of "to a great extent" backed up the assertion by Belfiore, G. and Lash, D. (2017) that a vocabulary multiple-choice test might ask students to say a word, define it, use it in a phrase or a story, use it in a poem, or look up synonyms and antonyms. A teacher can tell if a student knows

how to use a word by seeing how it is used in different situations. If a student can say the word and give a dictionary definition, but can't explain what it means, the teacher may think that the student still needs help understanding it. The teacher might then plan changes and extra help in future lessons to help the student understand the phrase better. Next, K3 has a mean score of 2.97 and a verbal description of to a great extent. This continues to support Belfiore, G. and Lash, D. (2017) claim that a teacher must be very careful and thorough when creating assessments. The average score for K4 is 2.85, and the verbal description is to a great extent. The average score for K5 is 2.82, and the verbal description is still to a great extent. With a mean score of 2.85 and a verbal description of "to a great extent" for K6, a mean score of 2.76 and a verbal interpretation of "to a great extent" for K7, and a mean score of 3.03 and a verbal interpretation of "to a great extent" for K8, Grimes-Hillman, et al. (2014)'s claim that writing, reading, and verbal evaluation should all be included in multiple-measure exams along with pencil and When speaking, writing, or reading in a second language, English-language learners (ELL) often don't use all three of these skills together. A multiple measure assessment looks at all three topics to help ELL students show what they know. When ELL students take tests with more than one measure, they can show what they know in different ways. When teachers give this kind of test, they can find out exactly where a student's conceptual strengths and weaknesses are. Teachers can find out how much ELL students understand by giving them tests with more than one measure. K9 has a mean score of 2.91 and a verbal description of to a great extent. K10 has a mean score of 2.97 and a verbal description of to a great extent. Finally, the average score for K11 is 3.15, and the verbal interpretation score is to a great extent.

We can conclude that the **highest score for Knowledge** goes to K11 (I can explain my own beliefs, attitudes, context, and needs for assessment) and the **lowest score** for this dimension of language assessment is K7 (I can evaluate the kind of washback that assessments can have on learning, teaching, context, curricula, and institutions.) From this, we could infer that teachers find it easy for them to explain their beliefs, attitudes, context, and needs for assessment since this statement of their ability is subjective and personal. This also means that they are aware of their own values that will help them to realize why they are doing assessments that will be consistent with these values. On the other hand, the researcher thinks that most of the teachers do not fully understand the word *"washback"* in assessment which served as the reason why most of them scored themselves with a very low capacity to evaluate it.

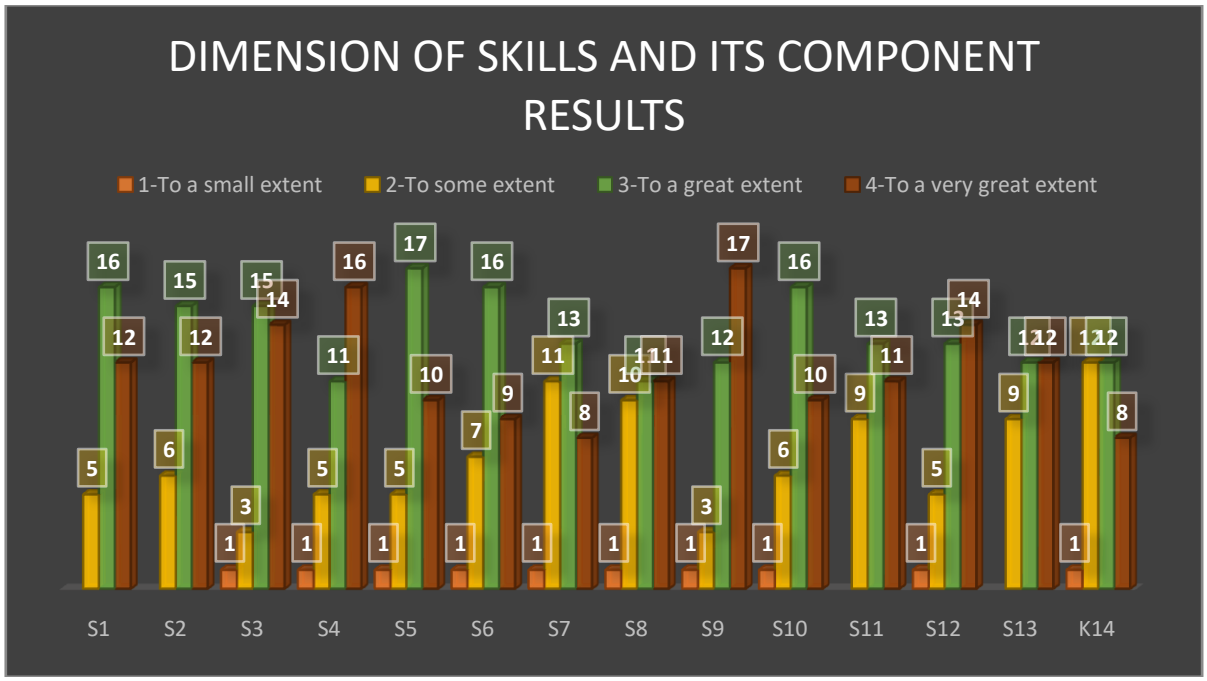

Figure 7 Number of respondents who answered for each skills' components

No.	Dimensions: Skills	Mean Score of 33 Respondents
1	I can align curriculum objectives, instruction, and assessment.	3.21
2	I can plan, implement, monitor, record, and report student language development.	3.18
3	I can communicate and provide feedback on students' assessment performance (norm- and criterion-referenced) to parents and students.	3.27
4	I can collect formal data (e.g., through tests) and informal data (while observing in class) on students' language development.	3.27
5	I can improve instruction based on assessment results and feedback.	3.09
6	I can utilize multiple methods and alternative means of assessment to make decisions based on substantive information.	3.00
7	I can construct test specifications (or blueprints) to design parallel forms of a test.	2.85
8	I can design assessments that are reliable, authentic, fair, ethical, practical, and interactive.	2.97
9	I can write selected-response items such as multiple-choice, true-false, and matching.	3.36
10	I can improve test items after item analysis, focusing on items that are either too difficult, too easy, or unclear.	3.06
11	I can design constructed-response items (for speaking and writing), along with rubrics for assessment.	3.06
12	I can infer students' strengths and weaknesses based on data.	3.21
13	I can interpret data related to test design, such as item difficulty and item discrimination.	3.09
14	I can run operations on Excel; for example, descriptive statistics and reliability correlations	2.82
	Average Score of Respondents for Skills	3.10 To a great extent

Figure 8 Implementation of Excel formulas for calculating the mean score on respondents' skills

Figure 7 unveils each 14 statements with verbal interpretation based on the description of Likert-scale representation. **S stands for Skills and numbers 1-10 have equivalent components that can be seen in Figure 8.**S1 with an average score of 3.21, or to a great extent when translated verbally. Following S1 with a mean score of 3.18 and a verbal interpretation of "to a great extent," we get S2. To a great extent is the verbal interpretation of both S3 and S4, with a mean score of 3.27. An S5 average score of 3.09 indicates that the skill is exercised to a great extent. With an average of 3.00 and a verbal description of "to a great degree" for S6, 2.85 and a verbal interpretation of "to a great extent" for S7, and 2.97 and a verbal interpretation of "to a great extent" for S8. The average rating for S9 is 3.36, and it's described as done to a great extent. The mean score for S10 and S11 is 3.06, and both have a verbal description of to a great extent. The average teachers got for S12 is 3.21 and are able to describe their experience as "to the great extent" verbally. Last but not least, the mean score for S13 is 3.09, and the mean score for verbal description of S14 is to a great extent, whereas the mean score for S14 is 2.82.

With this, S9 *(I can write selected-response items such as multiple-choice, true-false, and matching)* got the highest score for the skills in language assessment of 33 English as Second Language (ESL) teachers have. To the contrary, S14 *(I can run operations on Excel; for example, descriptive statistics and reliability correlations)* is the skill that teachers think there is still need for improvement.

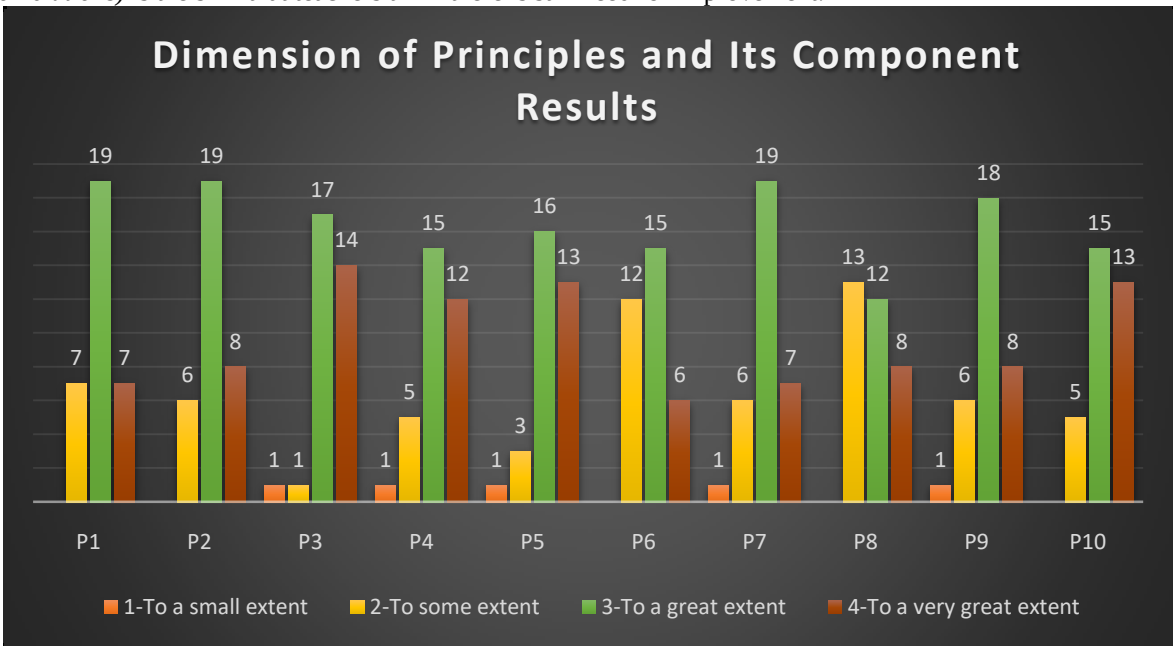

Figure 9 Number of respondents who answered for each principles' components

No.	Dimensions: Principles	Mean Score of 33 Respondents
1	I can clearly inform the inferences and decisions that derive from scores in assessments.	3.00
2	I can use assessment results for feedback to influence language learning, not other construct-irrelevant sources (e.g., personal bias towards a student).	3.06
3	I can treat all students, or users of language assessment, with respect.	3.33
4	I can use tests, test processes, and test scores ethically.	3.15
5	I can provide assessment practices that are fair and non-discriminatory.	3.24
6	I can critique the impact and power standardized tests can have and has a stance towards them.	2.82
7	I can observe guidelines for ethics used at the institution regarding language assessment.	2.97
8	I can criticize external tests based on their quality and impact.	2.85
9	I can implement transparent language assessment practices; informs students of what, how, and why of assessment.	3.00
10	I can implement democratic language assessment practices, by giving students opportunities to share their voices about assessment.	3.24
	Average Score of Respondents for Principles	**3.07** To a great extent

Figure 10 Implementation of Excel formulas for calculating the mean score on respondents' principles

Figure illustrates each 10 statements with verbal interpretation based on the description of Likert-scale representation. **P stands for Principles and numbers 1-10 has equivalent components that can be seen in Figure 10.**All the components for principles also got a mean score that is equivalent "to a great extent". This means that all of 33 ESL teachers do adhere to these principles whenever they are to create a language assessment.

Consequently, the 33 ESL teacher-respondents scored themselves the highest in P3 (I can treat all students, or users of language assessment, with respect.) and the lowest in P6 (I can critique the impact and power standardized tests can have and have a stance towards them.) We could deduce that ESL teachers strongly believe in the diversity of our learners and as a result we must take note of these differences – the possible advantages and disadvantages the students will encounter as they perform the assessment. Conversely, they suppose they have the least control on the weight and significance of standardized exams, as it is generally accepted that a class's poor performance on a standardized test is a reflection of the teacher's teaching style.

Discussion

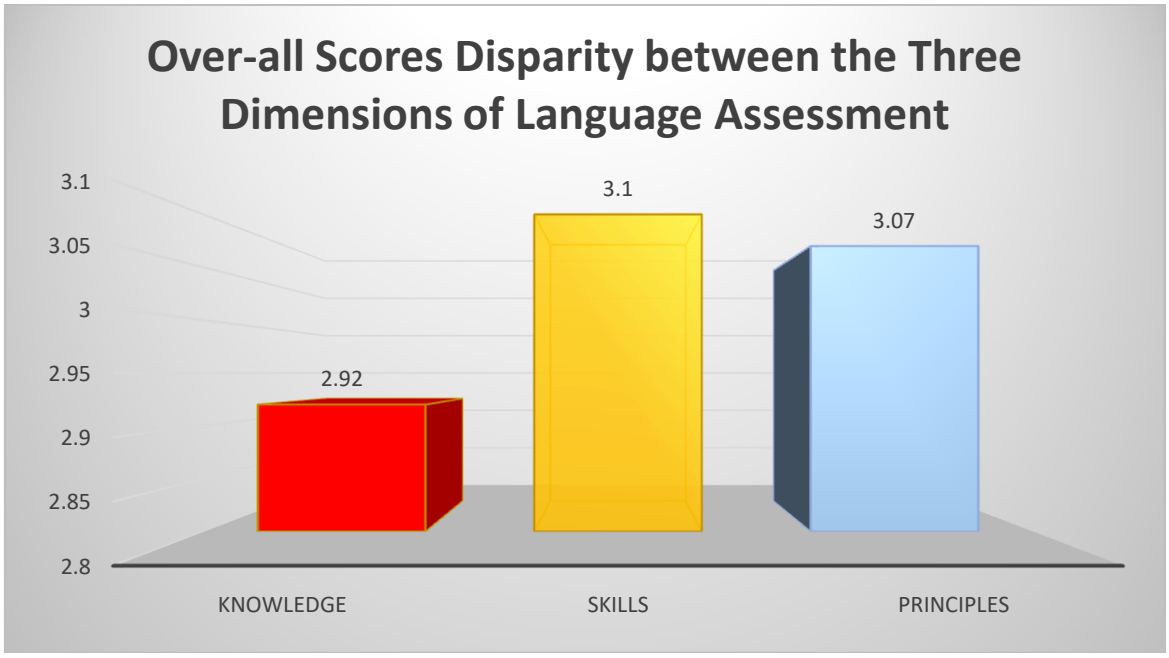

Figure 11 Over-all Disparity between the Three Dimensions of Language Assessment

Figure 11 depicts the gap between the three dimensions of language assessment as Skills placed first with a mean score of 3.10 percent, followed by Principles with a mean score of 3.07%, and Knowledge positioned

last with a mean score of 2.92%. Therefore, many ESL teachers trust that they have the skills to create an effective and efficient language assessment. They are also familiar with the different language assessment principles that serve as a guide in creating assessment tools. Unfortunately, they are not familiar with the terms, concepts, and approaches to language assessment. The respondents must have not felt confident that the assessment instruments that they have made adhere to the 7 Principles of Language Assessment – Reliability, Validity, Wash back, Practicality, Transparency, Authenticity, and Purpose. Moreover, a connection to the lowest scoring component of Knowledge (K7) can be linked to ESL teachers lowest scoring Principles component (P6); ESL educators have trouble appraising the wash back that assessments can have on learning, teaching, etc., therefore they struggle to remark on the influence of standardized examinations and modify accordingly.

As a result of this study, the researcher has concluded that a descriptive study of three dimension of language assessment have clearly define the characteristics that the teachers should possess for today's setup of education system, despite of constant change of education system, teachers should possess a life-long learning experiences for their student or learners to develop skills and capabilities using the different dimensions of language assessment. Assessing students with the integration of technological advancement will also ensure student progress as they have the most accessible and wide variety of resources that can be used as a reference for them to be able to understand the lesson critically and skilfully. All instructors should be invited to assessment training so that they may be refreshed on the latest methods for creating and discriminating test items, as well as the most significant terminology linked to assessment.

References:

[1]. Belfiore, G., & Lash, D. (2017, October). *Assessment design for broader, deeper competencies.* Next Generation Learning Challenges. https://s3.amazonaws.com/nglc/resource-files/MyWays_12Assessment.pdf

[2]. Bhat, A. (2022, June 13). *Descriptive Research: Definition, Characteristics, Methods, Examples and Advantages*. QuestionPro. https://www.questionpro.com/blog/descriptive-research/

[3]. Brockhart, S. (2016, December 12). *Performance assessment: What is it and why use it?* Medium. https://medium.com/inspired-ideas-prek-12/performance-assessment-what-is-it-and-why-use-it-1394712c5d3

[4]. Brown, H. D. (2004). *Language Assessment*. Longman.

[5]. Carlson, D., & Albright, J. (2012). *Composing a care for the self: A critical history of writing*

[6]. Cheng, L., & Fox, J. (2013). Review of doctoral research in language assessment in Canada

[7]. Danuwijaya, A. A. (2018). Item Analysis of Reading Comprehension Test for Post-Graduate Students. *English Review: Journal of English Education*, *7*(1), 29. https://doi.org/10.25134/erjee.v7i1.1493

[8]. Fox, J. (2016). *Using Portfolios for Assessment/Alternative Assessment*. SpringerLink. https://link.springer.com/referenceworkentry/10.1007/978-3-319-02326-7_9-1?error=cookies_not_supported&code=db54517a-1695-417c-87e1-ca9050782706

[9]. Ghaith, G. (2001, November 30). *ERIC - EJ671645 - Using Cooperative Learning To Facilitate Alternative Assessment., Forum, 2002*. ERIC - Institute of Education Sciences. https://eric.ed.gov/?id=EJ671645

[10]. Gibson, C. (2016). Bridging English language learner achievement gaps through effective vocabulary development strategies. *English Language Teacher, 9*(9), 134-138. https://doi.org/10.5539/elt.v9n9p134

[11]. Great Schools Partnership (2014, April 29). *Formative assessment.* The Glossary of Education Reform. https://www.edglossary.org/

[12]. Griffin, P., McGaw, B., & Care, E. (Eds.). (2012). Assessment and Teaching of 21st Century Skills. *Assessment and Teaching of 21st Century Skills*. https://doi.org/10.1007/978-94-007-2324-5

[13]. Kofou, I. (2018, January 29). *Alternative assessment modes: Reading and movie diaries*. ACADEMIA. https://www.academia.edu/35787744/Alternative_assessment_modes_Reading_and_movie_diaries S0261444813000244.

[14]. Singh, C. K. S., Muhammad, M. M., Mostafa, N. A, Yunus, M. M., Noordin, N., & Darmi, R. (2022). Exploring ESL teachers' Alternative Assessment Strategies and Practices in the Classroom. *Journal of Language and Linguistic Studies, 18*(1), 411-426. Doi: 10.52462/jlls.191

[15]. Smith, M., & Stansfield, C. W. (2017). Testing Aptitude for Second Language Learning. *Language Testing and Assessment*, 77–90. https://doi.org/10.1007/978-3-319-02261-1_5

[16]. *Take the Test sample Questions from OE CD's PISA Assessments*. (2009). OECD publishing.

[17]. تركي السليماني, ت. ر. ك. ي ا. ل. س ل. ي م ا. ن. ي (2021). طرقبديلة لتقييم اللغة في العصر الرقمي *Journal of King Abdulaziz University Arts and Humanities, 29*(1), 597–609. https://doi.org/10.4197/art.29-1.20