*Asia Pacific Journal of Educators and Education, Vol. 30, 107–122, 2015*

# ASSESSING ORAL COMMUNICATION SKILLS IN SCIENCE: A RUBRIC DEVELOPMENT

**Azizi Alias[1, 2*] and Kamisah Osman[3]**

[1]Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[2]Textbook Division, Level 1-3, Block E15, Complex E, 62604 Putrajaya, Malaysia
[3]Faculty of Education, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

*Corresponding author: azizi.alias@moe.gov.my

**Abstract:** This research was conducted to develop an analytical rubric for alternative assessment of scientific oral activities with the inculcation of 21st century skills. The aim of the research is to determine whether the developed analytical rubric is suitable for assessing oral communication skills in science activities by taking into account advice from a panel of 15 experts in science education to validate the rubric in three Delphi rounds and by using 5 experienced teachers as assessors to test the reliability of the rubric using an intra-class correlation coefficient (ICC). The study found that the analytical rubric has a high relevance percentage of 100% and high consensus of 75.0% as well as a high absolute agreement of 85.0% [ICC = .85 (95% CI, .69 to .94)]. This result shows that the analytical rubric developed is appropriate for assessing oral communication skills in school science activities. This effort is essential in realising the assessment of authentic outcomes, concurrent with 21st century needs. However, further research on the validity and reliability of the rubric is necessary.

**Keywords:** 21st century skills, oral communication skills, alternative assessment, analytical rubric

## INTRODUCTION

According to Malaysia's Ministry of Higher Education (2012), the deficiency in communication, especially in English, is the downfall that shows that higher education graduates are lacking in communication skills. Employers often complain about the incompetence of graduates in general skills such as communication, problem-solving, interpersonal skills and the flexibility to overcome problems in various situations (Kaur & Kaur, 2008). Does the current education system line up with the demands of the 21st century, and what student outcomes are needed to fulfil the expectations of employees? Is a 21st century skill assessment implemented in schools to generate future citizens who are able to foster new skills to face the challenges ahead?

Students in the 21st century must be able to develop and to implement new skills in preparing themselves for an ever-changing world (Osman, Abdul Hamid, & Hassan, 2009). The direction of 21st century skills is to foster in students the ability to apply technology through digital literacy, to think critically and creatively and to develop interpersonal and social skills (Tuan Soh, Osman, & Mohamad Arshad, 2012). The question is how to nurture 21st century skills in science education settings in school and to assess communication skills specifically.

Effective communication skills involve teamwork and cooperation, with an emphasis on interaction between two or more individuals to solve problems, to produce new products and to explore the content of a certain subject (NCREL & Metri Group, 2003). Communication skills are an important element in problem-based learning because this type of learning involves many activities that require the students to interact as a team to achieve solutions by information gain in presenting writing or presenting processes (Judge, Osman, & Mohd. Yassin, 2011). What type of assessment can be implemented to expand students' oral communication skills?

Alternative assessment is an approach to determine students' learning and thinking that focuses on higher-order thinking and problem-solving abilities (Burke, 2005). These abilities are crucial in preparing students for their future careers and to develop skills needed in the 21st century. Chapman and King (2012) argued that alternative assessment provides a chance for students to use information or skills in real life situations. Thus, alternative assessment can explicitly expose students to authentic activities, which cannot be accomplished through traditional testing. Does science education in Malaysia promote such practices and support the development of 21st century skills? Banks (2005) suggested ways to implement alternative assessment effectively:

1. Focusing on the "big idea" and not depending merely on facts and certain skills.
2. Involving interaction between educator and student, student and student.
3. Providing multiple approaches to demonstrate the achievement of students.
4. Using a scoring scheme focused on the main idea of a task, not on the easiest way to get good marks.

In alternative assessment, the scoring scheme is important to provide guidance for educators in evaluating students' achievement and students' preparation for completing their task appropriately. The exposure to real life problems can be reflected in alternative activities, and the use of a rubric is the best approach in assessing competencies that are needed in life (Airasian, 2001). How can a rubric

be developed that can foster 21st century skills and provide space for students to nurture their communication skills in school? The objective of this study is to develop an analytical rubric to assess oral communication skills for oral activities such as debate, public speaking and discussion, taking into account the validity and reliability of the rubric. It is hoped that the rubric can be implemented in alternative activities and can facilitate systematic assessment and promote meaningful learning.

## LITERATURE REVIEW

Oral communication refers to the ability to speak verbally with the help of visual aids and non-verbal elements to support delivery of meaning, and it involves speech, presentation, discussion and interpersonal communication aspects (MTD Training, 2012). As stated by Lantz (2004), body language and the selection of voice tone plays an important role in oral communication and has a great impact on listeners and on the information content. Oral communication is essential for scientists because they need oral communication skills to deliver information to society (Rezaei & Lovorn, 2010). As stated by Garvey and Griffith (1972), scientists must be capable in communicating their research performance and ideas with others. Thus, science education must play an essential role in enhancing students' competence in communication (Holbrook & Tannikmae, 2007). Hence, communication skills should be nurtured in preparing students as for future careers as scientists.

Communication is an important element in science education because students need to transfer their findings in the form of charts, diagrams, tables, figures or graphs and to explain them clearly. However, Bell and Carr (2014) argued that students believed that communication skills are of secondary importance compared to written coursework and laboratory reports. Jeon and Park (2013) have listed communication elements that need to be developed by science students in several countries.

According to Ak and Guvendi (2010), education is not an effort to educate students to memorize and repeat what has been done by the previous generation but to foster students to be innovative by developing educational activities that allow each individual to adapt to the changes and uncertainties of the global environment. Thus, education institutions play an important role in producing human capital that is equipped with 21st century skills, but how are educators to assess oral communication skills with the inculcation of 21st century skills?

| Country | Communication Elements (Policy/Curriculum/Learning Objective) | |
|---|---|---|
| Taiwan | i. | Students need to communicate, cooperate and work in harmony with others. |
| | ii. | The emphasis is on building community spirit through communication. |
| Canada | i. | Students should improve experimental competence. |
| | ii. | Students should improve research competence. |
| | iii. | Students should improve technical problem-solving abilities. |
| Australia | i. | Students must have scientific inquiry abilities in questioning and predicting, planning and conducting, processing and analysing data and information, evaluating and communicating. |
| Korea | i. | Learners can explain. |
| | ii. | Learners can present |
| | iii. | Learners can provide the ground |
| Singapore | i. | Students are able to communicate using figures, tables and graphs. |

Oral communication through alternative activities provides students with social and interpersonal skills needed in the 21st century. Communication activities require students to collaborate as a team and to jointly take responsibility for achieving the goal of the activity goal (North Central Regional Education Laboratory, NCREL, 2002). For example, in public speaking, students are required to communicate well in delivering information and to analyse the topic with the help of technology to clarify the meaning of the speech. Communication skills are the top soft skill attributes needed by employers in today's work place (Robles, 2012).

Assessment that involves students in implementing activities to produce a product that demonstrates their learning is called performance-based assessment and is also known as alternative assessment or authentic assessment (Airasian, 2001). Educators who are ambitious have realised that the traditional approach of the "paper-pencil test" is not an absolute indicator to assess or evaluate learning. Furthermore, performance-based assessment or alternative assessment is the best method for exploring students' abilities and understanding in an authentic way with the application of prior knowledge and skills (Banks, 2005). Alternative assessment refers to activities such as experimentation, projects, surveys, check lists, portfolios, observations and interviews (Nasri, Roslan, Sekuan, Abu Bakar, & Puteh, 2010). A generally accepted goal in building students' understanding of

science is to relate the science that is learned in school to students' everyday life experiences (Lay, Khoo, Treagust, & Chandrasegaran, 2013). Thus, alternative assessment is an appropriate approach in science teaching to develop students' understanding of scientific concepts and to transfer those concepts to real life.

Traditional testing is limited to memorizing facts for "paper-pencil" tests, short essays or simple oral tests (Unal, Bodur, & Unal, 2012). However, the existence of other types of assessment does not mean that traditional testing is discarded because alternative assessment has its own limit and constraints. Alternative assessment is considered to be a part of learning and focuses on how the students build their own knowledge rather than evaluating the level of knowledge that they have mastered (Nasri et al., 2010). As stated by Kelvin (2013), alternative assessment has the potential, value and benefits that are significant for current assessment practices. Additionally, alternative assessment is focused on the development of students, not to merely identify their flaws but to provide a space to build their potential in problem solving, higher-level thinking and creativity (Sasmaz-Oren & Ormanci, 2011). However, science educators found that there are some constraints in implementing a problem-solving approach in relation to real life situations, including time, curriculum and students' abilities (Lee, Tan, Coh, Chia, & Chin, 2000). There are three issues that need to be considered in the implementation of alternative assessment: reliability, sampling and time (McMillan, 2011).

Clearly, alternative assessment is an approach that has the ability to render high stakes examinations, which are concerned about the end product not on evaluating students' performance or the effectiveness of teaching. Jaba, Hamzah, Bakar and Mat Rashid (2013) argued that the implementation of assessment for learning enables educators to evaluate students continuously when they demonstrate their knowledge and skills during the instruction, to make learning more exciting and to increase meaningful outcomes. Developing scientific knowledge, creativity and habits of mind among students for questioning and learning about life-world phenomena can be facilitated through involving students in authentic experiences based on inquiry learning such as problem solving and investigations (Kim & Chin, 2010). Erdogan and Tuncer (2009) found that students preferred activities that involved field trips and in which they were encouraged to present their findings. Thus, activities that involve real life situations and channel students' experiences enhance learning and develop their communication skill.

Kishbaugh, Cessna, Horst, Leaman, Flanagan, Neufeld and Siderhurst (2012) stated that the application of assessment is not only to grade the product but also to evaluate the aim of teaching, adapting assessment to the students' needs and developing instruction that is valuable for students' learning. The rubric is an

assessment tool that can aid students in determining performance standards and in creating a product based on objective standards (Kohn, 2006). Thus, rubrics are beneficial not only for students but also for educators in enhancing the quality of teaching and learning, especially in authentic activities. There are two categories of rubrics that must be clarified upon implementation. Holistic rubrics require the educator to make an overall scoring upon a student's production without looking at the separate components of the assessment (Nitko, 2001). McMillan (2011) stated that a holistic rubric consist of several criteria that are combined in a single score. However, analytical rubrics are suitable for assessing students' performances based on each specific criterion (Airasian, 2001). Thus, an analytical rubric is appropriate for observing students' specific strengths and weaknesses and also for improving teaching. However, according to Mertler (2001), the use of analytical rubrics is time consuming in terms of developing the rubric and scoring using the rubric.

## METHODOLOGY

The validity of this rubric is determined by implementing the Delphi technique. Theoretically, the Delphi technique is continued until the discussion has reached a consensus or an agreement is identified (Millar, 2001; Valerdi, 2011). The use of the Delphi technique allows creative collaboration and encourages panels to solve problems and issues that may arise in the future (De Loe, 1995). Brookhart (2013) suggested that there are two ways to develop a rubric: *"top-down"* is used when there is a clear curriculum and standards, and *"bottom-up"* uses multiple examples of student production to develop the conceptual framework. In this study, the rubric is developed *"top-down"* by referring to the standard documents (curriculum) and various oral presentation rubrics found in the literature. As the rubric is developed, it will undergo Delphi rounds 1, 2 and 3 to validate the rubric by determining its relevance and reaching a consensus in the panel of experts. The panel of experts will also give suggestions and comments to improve the rubric. The validity of the rubric is determined by analysing the median, the first and third quartile, and the interquartile range from a 5-point Likert scale. Gravetter and Wallnau (2002) confirmed that the median is taken into account because it is less influenced by extreme values. In this study, the researcher has determined that all constructs must demonstrate higher than 70% relevance and high consensus to ensure that the developed rubric has a high validity.

A reliability test using inter-rater reliability was performed using scores from five science teachers who rated one student's performance. The assessors were selected among experienced teachers who are familiar with the scoring of the student outcomes for school-based assessments and have been teaching science for more than 10 years. Unal et al. (2012) argued that recording scores based on

observation and comparing the scores among assessors should yield high credibility and reliability based on evidence. However, Airasian (2001) stated that reliability is not a crucial aspect in assessing authentic activities but that good validity is required. In this study, a clip from a student's presentation video was presented to five science educators, who were selected to assess scientific oral communication skills, and the intra-class correlation coefficient (ICC) was used to determine the reliability of the developed rubric using the same student video presentation. In this study, if the "absolute agreement" reached more than 70%, the rubric is considered to have a good reliability.

## FINDINGS

### Analysis of the Degree of Relevance and Consensus Reached in Delphi Round 1

**Table 1.** The degree of relevance and consensus reached in Delphi Round 1

| Construct | Degree of Relevance (median 4–5 High) | Degree of Consensus (IQR < .5 High) |
|---|---|---|
| Aim | 4 | High |
| Scientific Accuracy | 3 | Fair |
| Supporting Data | 4 | Fair |
| Introduction | 3 | High |
| Content | 4 | High |
| Conclusion | 4 | High |
| Supporting Materials & Appearance | 4 | High |
| Body Language | 4 | High |
| Voice | 3 | High |
| Rhythm | 3 | High |
| Eye Contact | 3 | High |
| Ability to Deliver | 4 | High |
| The Use of Humour or Story | 3 | Fair |
| Language | 3 | High |
| Scientific and Noble Values | 5 | High |
| Percentage | High: 50%<br>Fair: 50% | High: 81.25%<br>Fair: 19.75% |

ICC analysis is a method used to identify how close the assessors' scores are, and the reliability is based on the "absolute agreement" reached (Good, 2012). At the end of the scoring program, the assessor will answer a number of items adapted from (Ross, 2011) to determine the reliability that fits the purpose of the developed rubric.

Table 1 shows that there are eight constructs with a median value equal to or greater than 4, which shows that these eight constructs have a high degree of relevance and the other eight constructs have a fair degree of relevance. Thirteen constructs have an interquartile range less than 0.5 or equal to 0 and can be considered to have a high consensus.

In Delphi Round 1, no construct has low consensus; however, some criteria need to be modified, especially constructs with fair consensus. The panel of experts scored 50% at fair relevance due to some unclear phrases and their questioning of whether the use of the rubric is suitable for low achieving students. Although the rubric has a high consensus of 81.25%, this result should be parallel with the rubric's relevance to claim that the panel of experts have reached complete consensus.

Modifications were made to the criteria with regard to unclear phrases; for example, "has an attempt," which is difficult to be observed, was changed to "fairly accomplish" to differentiate between the excellent and good scores. The construct of "supporting materials must consist of charts, graphs, animation and audio" was changed to "supporting materials consist of charts/graphs/animation/ audio". The panel of experts believed that the burden of preparing the supporting materials must be reduced and should be based on the students' abilities and environment. Although the panel of experts expressed that the number of constructs was too large, they did not state which construct(s) should be eliminated. There was an issue with the construct "supporting materials and appearance" being conflated. How could the performance of a student with good supporting material but a bad appearance be determined, or vice versa? The construct was divided into two different constructs: supporting materials and appearance.

## Analysis of the Degree of Relevance and Consensus Reached in Delphi Round 2

**Table 2.** The degree of relevance and consensus reached in Delphi Round 2

| Construct | Degree of Relevance (median 4–5 High) | Degree of Consensus (IQR < .5 High) |
|---|---|---|
| Aim | 4 | High |
| Scientific Accuracy | 3 | Fair |
| Supporting Data | 4 | Fair |
| Introduction | 4 | Fair |
| Content | 4 | Fair |
| Conclusion | 4 | Fair |
| Supporting Materials | 4 | Fair |
| Appearance | 4 | Fair |
| Body Language | 4 | Fair |
| Voice | 4 | Fair |
| Rhythm | 4 | High |
| Eye Contact | 4 | Fair |
| Ability to Deliver | 4 | Fair |
| The Use of Humour or Story | 4 | Fair |
| Language | 4 | Fair |
| Scientific and Noble Values | 5 | Fair |
| Percentage | High: 93.75%<br>Fair: 6.25% | High: 12.50%<br>Fair: 97.50% |

Table 2 shows that only one construct has fair relevance and 14 constructs have fair consensus for the developed rubric. The rubric shows the high degree of relevance as 93.75% but the degree of high consensus is only 12.50%. The changes after Delphi Round 1 clearly changed the consensus among the panel of experts, and it was necessary to implement Delphi Round 2 to achieve an explicit outcome regarding the degree of relevance and consensus. No criterion was modified in this Delphi round, but some grammatical errors were corrected by the panel of experts.

## Analysis of The Degree of Relevance and Consensus Reached in Delphi Round 3

**Table 3.** The degree of relevance and consensus reached in Delphi Round 3

| Construct | Degree of Relevance (median 4-5 High) | Degree of Consensus (IQR < .5 High) |
|---|---|---|
| Aim | 4 | High |
| Scientific Accuracy | 4 | Fair |
| Supporting Data | 5 | Fair |
| Introduction | 4 | High |
| Content | 5 | High |
| Conclusion | 4 | High |
| Supporting Materials | 4 | High |
| Appearance | 5 | High |
| Body Language | 4 | High |
| Voice | 4 | High |
| Rhythm | 4 | Fair |
| Eye Contact | 5 | High |
| Ability to Deliver | 4 | High |
| The Use of Humour or Story | 4 | Fair |
| Language | 5 | High |
| Scientific and Noble Values | 5 | High |
| Percentage | High: 100% | High: 75% |

Table 3 shows that all constructs reached 100% for high relevance and 75% for high consensus by the 15 experts. Therefore, the developed rubric can be implemented and has a high validity for assessing students' oral communication skills. As stated by Ali (2005), the Delphi technique should be implemented in more than two rounds because the consensus becomes more explicit after Round 3 or higher.

## Scoring by Assessors (Educators)

**Table 4.** Analysis of mean and standard deviation in scores

| Assessor | Mean | S.D. |
|---|---|---|
| 1 | 2.25 | 0.56 |
| 2 | 2.19 | 0.66 |
| 3 | 2.13 | 0.62 |
| 4 | 2.19 | 0.75 |
| 5 | 2.06 | 0.68 |

As shown in Table 4, Assessor 3 scored the highest mean (m = 2.17, s.d. = 0.41) and Assessor 2 scored the lowest mean (m = 1.50, s.d. = 0.84). The mean and standard deviation values are inconsistent, and thus the scoring among assessors is also inconsistent. Correlation analysis is needed to identify the relationship of scoring from assessor to assessor.

### Matrix Analysis of Item Correlation for Oral Communication Rubric

**Table 5.** The correlation matrix between items in scoring using the oral communication rubric

| Assessor | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | – | – | – | – | – |
| 2 | 0.57 | – | – | – | – |
| 3 | 0.84 | 0.60 | – | – | – |
| 4 | 0.65 | 0.51 | 0.52 | – | – |
| 5 | 0.57 | 0.89 | 0.58 | 0.55 | – |

Table 5 shows the scoring correlation between assessors using the oral communication rubric. Assessor 2 and Assessor 5 show the highest correlation of $r = 0.89$, and Assessor 2 and Assessor 4 show the lowest correlation of $r = 0.51$, which shows that all assessors have a high positive correlation and only Assessor 2 and Assessor 4 show a fair positive correlation. Thus the scoring has a fair consistency among assessors.

### Analysis of Intra-Class Correlation (ICC-3)

**Table 6.** Coefficient between classes for absolute agreement between assessors using the oral communication skill rubric

| | Intraclass Correlation | 95% Confidence Interval | |
|---|---|---|---|
| | | Lower Bound | Upper Bound |
| Single Measures | 0.53 | 0.31 | 0.76 |
| Average Measures | 0.85 | 0.69 | 0.94 |

Based on Table 6, the coefficient between assessors shows a high "absolute agreement" of 85% (ICC = 0.85, Cl 0.69 to 0.94). Thus the analytical rubric for oral communication skills has a high reliability of 85%.

## Analysis of the Reliability of the Rubric by Assessors

**Table 7.** Analysis of the reliability of the rubric (adapted from Ross, 2011)

| Item | Statement | Mean | SD |
|---|---|---|---|
| 1. | Does this rubric integrate the main assessment element in the development of all tasks | 4.40 | 0.548 |
| 2. | Does this rubric make the main skills easy to assess | 4.20 | 0.447 |
| 3. | Does this rubric enable the teacher to develop criteria that can be applied in the future | 4.00 | 0.707 |
| 4. | Does this rubric act as a reliable tool to assess the quality of a task | 4.00 | 0.707 |
| 5. | Does this rubric clearly focus on the level required for each criterion | 4.20 | 0.837 |
| 6. | Does this rubric integrate the criteria specifically and beneficially in a real life situation | 3.80 | 0.837 |
| 7. | Does this rubric facilitate making unbiased comparisons in all given tasks | 3.80 | 0.447 |
| 8. | Does this rubric have freedom from bias and stereotypes (gender, age) | 3.60 | 0.548 |
| 9. | Does this rubric help educators to gain criteria for actual performance | 3.80 | 1.095 |
| 10. | Does this rubric have an adequate number of criteria and not too many | 3.50 | 0.605 |
| 11. | Does this rubric consist of criteria that are not too general | 3.60 | 0.548 |

Table 7 shows the analysis of the instrument that was given to the assessors to determine the reliability of the developed rubric, adapted from Ross (2011). The purpose of this instrument is to provide qualitative support, specifically focusing on statements that refer to the reliability of a developed rubric. The statement that the rubric "integrates the main assessment element in the development of all tasks" shows the highest mean score (m = 4.40, s.d. = 0.584), and "the number of criteria being adequate and not too many" shows the lowest mean score (m = 3.50, s.d. = 0.605). This result shows that the assessors believed that the rubric covers all aspect of oral communication skills but has many criteria to be assessed and is time consuming. However, neither experts nor assessors suggested which criteria should be eliminated, and all seem important.

## DISCUSSION

The developed rubric for oral communication skills has 100% relevance and 75% consensus reached by 15 experts in science education. Thus, the rubric has sound validity, agreed upon by the panel of experts. The analysis using the ICC shows that the rubric has a high "absolute agreement" of 85%, can be counted as having good reliability and can be implemented. As Kottner et al. (2011) stated, some researchers have fixed the minimum value of the reliability coefficient at 0.60, 0.70 and 0.80; but, in some cases, the reliability coefficient is 0.90 to 0.95, depending on the requirements of the study. Some panel members stated that the number of criteria in the rubric was too many and needed to be reduced because it took too long to make an assessment. However, the panel members did not describe which criterion should be eliminated, and all criteria seem imperative. In this regard, the assessors scored the lowest mean for the item "the number of criteria is adequate and not too many," and they perceived that there are many criteria in the rubric. Wolf and Stevens (2007) confirmed that three to six criteria are adequate and considered favourable. If the criteria are too numerous, more time is required for assessment, and the assessor will have problems in memorising each criterion. However, if the criteria are too few, it is difficult to precisely determine the students' achievement.

Oral communication activities allow students to demonstrate their talents and diverse intelligences such as linguistic, bodily-kinesthetic, musical and interpersonal. Baum, Viens and Slatin (2005) argued that debate and drama have the ability to develop linguistic and bodily-kinesthetic intelligence. Such activities require students to speak and to use proper language and posture to ensure the transfer of meaning during the play, which also includes the use of music and singing. Furthermore, a drama production requires students to use their communication and interpersonal skills to ensure that the objective of the play is shown to the audience. Each criterion for assessing oral communication skills is listed in the developed rubric. In building the analytical rubric, certain criteria were adopted from several rubrics in the literature. In oral communication activities, it is crucial to have a bold objective in giving an argument so that the whole idea can be understood and this objective can be used as a stepping stone to a meaningful presentation. A presentation must define an aim, subject, information and data that are consistent with the theme (Jarvis & Cain, 2003). Discussion with the teacher is essential in preparing and planning the text to ensure that the writing is on track with regard to the criteria in the rubric.

As stated by Lauer and Hendrix (2009), scientific presentations must consist of the concept, facts, statistics, scientific phenomenon or story, and the references. The references demonstrate that the facts and/or statistics are reliable and guide the audience in performing further exploration. In a presentation, the introduction

plays an important role in clarifying the whole idea, providing a clear objective and motivation factor to engage the audience with the presentation. According to Jarvis and Cain (2003), a good introduction must contain a statement of objective, be attractive to the audience and be focused.

The content must consist of the topic that will be presented and must be well researched by the presenter, focused and relevant. The main idea should be summarized and also easily understood by the audience (Jarvis & Cain, 2003). In oral communication skills, personality and appearance play an important role in convincing the audience of the content of the speech and that it is concurrent with the theme. Presenters must be calm, confident and well-dressed during the presentation (Lantz, 2004). The individual who makes the presentation must be confident and have the ability to enhance the audiences' knowledge and understanding of the topic significantly using humour, analogy or a story that is suitable to the topic, as well as scenarios or motivation (Lantz, 2004). Coon (2012) suggested that the language used must be descriptive, appropriate for the audience, with sound grammar and sentence structure and with only minor mistakes. Researchers have suggested that the rubric can be modified by teachers according to their students' needs and environment. Communication skills play an essential role in science process skills. Higher order concepts and the nature of science require students to have the ability to communicate findings in the form of charts and graphs and with the use of technology (Kishbaugh et al. (2012). By possessing communication skills, students benefit from the ability to communicate, specifically in the field of science, through research papers, posters, brochures and oral communication, all of which are needed in their future careers.

There are some limitations in implementing authentic activities, and sacrifices must be made in the aspect of reliability. Akbulut and Akbulut (2011) stated that alternative assessment is not a prime choice made by teachers because teachers believe that alternative assessment has low reliability and is not beneficial to the students. This is a challenge for education institutions in implementing alternative assessment to measure competencies based on teachers' observations, which is questionable in the degree of validity and reliability. The aim of assessment is to help students, not to only focus on the reliability of the assessment; however, the assessment must be valid (Airasian, 2001). Thus, criteria in an assessment tool that is used to measure competencies must be understandable and explicit.

Several factors influence the implementation of alternative assessment: the expertise of the assessor, possibilities of bias and the competency of the assessor in implementing activities (Good, 2012). Wolf and Stevens (2007) argued that the complication in assessing achievement for each criteria in a rubric can be

reduced by using words that are clear and easy such as the ability to list, to draw, to discuss, to explain, to compare, to criticize, to predict, etc. In this study, some adjustments were made to clarify some criteria regarding behaviour that is difficult to observe such as "makes an attempt" and "is trying to," which were changed to "has the ability" and "is capable of".

A well-develop rubric is capable of assessing achievement with high validity and reliability and enhancing the quality of instruction. The use of a rubric does not necessarily make assessment easier, but a good rubric enables teachers to demonstrate their professional knowledge and to use their knowledge specifically for students with diverse personalities and learning difficulties (Wolf & Stevens, 2007). As stated by Bresciani et al. (2009), the use of a rubric does not eliminate the variation among assessors, but a well-developed rubric is capable of identifying problems and providing practice in making assessments to increase its reliability.

Scoring and the reliability of a task are influenced by several factors, including the objective of the task, items, scoring, difficulty level of the task, homogeneity of assessors, time provided, number of tasks and the domain that must be covered (Bresciani et al., 2009). Training and professional discussion on how to use and justify each criterion are crucial to increase an assessment's reliability and to ensure similar scoring among assessors.

## CONCLUSION

The aim of study was to develop an analytical rubric to assess scientific oral communication skills through alternative assessment with the inculcation of 21st century skills. The validity and reliability of the rubric were obtained using three rounds of the Delphi technique and an inter-rater reliability test. The rubric covered most of the communication elements that are needed in helping students develop their oral communication skills in science education. Furthermore, the developed rubric is expected to serve as an assessment tool for teachers, helping them to diversify their instructional approach and to nurture 21st century skills, preparing their students for their future careers. Because this study has identified some constraints in implementing alternative assessment and issues regarding the validity and reliability of the rubric, it is essential for educators to modify the rubric according to their students' abilities and resource vitality. Professional discourse among educators on how to assess authentic activities using rubrics will intensify competency in assessment and increase the validity and reliability of the rubric. Moreover, students need to be exposed to the rubric to ensure that they can plan to properly meet the standards assigned by the educators. In future research, it is recommended that the rubric include evidence from students with

different abilities and that more educators be involved in determining the reliability of the rubric.

## REFERENCES


Airasian, P. W. (2001). *Classroom assessment: Concept & application* (4th ed.). New York: McGraw-Hill Higher Education.

Ak, E. & Guvendi, M. (2010). Assessment of the degree to which primary school teachers use alternative assessment and evaluation methods. *Procedia Social and Behavioral Sciences, 2*, 5599–5604.

Akbulut, O. E. & Akbulut, K. (2011). Science and technology candidates' opinion regarding alternative assessment. *Procedia Social and Behavioural Sciences*, *15*, 3531–3535.

Ali, A. K. (2005). Using the Delphi Technique to search for empirical measures of local planning agency power. *The Qualitative Report*, *10*(4), 718–744.

Banks, S. R. (2005). *Classroom assessment: Issues & practices*. Boston, MA: Pearson Education Inc.

Baum, S., Viens, J. & Slatin, B. (2005). Multiple Intelligences in the elementary classrooms. New York: Teachers College Press.

Bell, M. & Carr, P. (2014). Building communication skills for science students in video conference tutorials. *International Journal of Innovation in Science and Mathematics Education*, *22*(4), 65–78.

Bresciani, M. J., Oakleaf, M., Kolkhorst, F., Barlow, C. N., Duncan, K. & Hickmott, J. (2009). Examining design and inter-rater reliability of a rubric measuring research quality across multiple disciplines. *Practical Assessment Research and Evaluation*, *14*(12), Retrieved from pareonline.net/pdf/v14n12.pdf

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading*. Virginia: Association for Supervision and Curriculum Development (ASCD).

Burke, K. (2005). *How to assess authentic learning* (4th ed.). California: Corwin Press.

Chapman, C. & King, R. (2012). *Differentiated assessment strategies: One tool doesn't fit all* (2nd ed.). California: Corwin Press.

Coon, R. M. (2012). *A compilation of rubrics to be used in chemistry to emphasize argumentative writing in science classroom*. Education & Human Development Master's Thesis, State University of New York.

De Loe, R. C. (1995). Exploring complex policy questions using the policy Delphi: A multi-round interactive survey method. *Applied Geography*, *15*(1), 53–68.

Erdogan, M. & Tuncer, G. (2009). Evaluation of a course: Education and awareness of sustainability. *International Journal of Environmental & Science Education*, *4*(2), 133–146.

Garvey, W. D., & Griffith, B. C. (1972). Communication and information processing within scientific disciplines: Empirical findings for psychology. *Information Storage and Retrieval, 8,* 123–126.

Good, J. (2012). Crossing the measurement and writing assessment divide: The practical mplications of inter-rater reliability in faculty development. *The WAC Journal, 23,* 19–30.

Gravetter, F. J. & Wallnau, L. B. (2002). *Essentials of statistics for the behavioural science* (4th ed.). Australia: Wadswoth & Thomson Learning.

Holbrook, J. & Tannikmae, M. (2007). The nature of science education for enhancing scientific literacy. *International Journal of Science Education, 29,* 1347–1362.

Jaba, S., Hamzah, R., Bakar, A. R. & Mat Rashid, A. (2013). Acceptance towards school based assessment among agricultural integrated living skills teachers: Challenges in implementing a holistic assessment. *Journal of Technical Education and Training*, *5*(1), 44–51.

Jarvis, L. & Cain, J. (2003). Diversifying assessment: Presentations in undergraduate history of science. *PRS-LTSN Journal*, *2,* 50–72.

Jeon, S. & Park, J. (2013). Analysis on communication skills in science education. *Advanced Science and Technology Letters*, *36,* 18–21.

Judge, S. K., Osman, K., & Mohd. Yassin, S. F. (2011). Cultivating communication in PBL with ICT. *Procedia & Behavioral Sciences, 15,* 1546–1550.

Kaur, G. & Kaur, S. (2008). Malaysian graduates' employability skills. *Unitar E-Journal*, *4*(1)*,* 14–44.

Kelvin, T. H. K. (2013). Variation in teachers' conceptions of alternative assessment in Singapore primary schools. *Educational Research for Policy and Practice*, *12*(1), 21–41.

Kim, M & Chin, C. (2010). Pre-service teachers' views on practical work with inquiry orientation in textbook-oriented science classrooms. *International Journal of Environmental & Science Education*, *6*(1), 23–37.

Kishbaugh, T. L. S., Cessna, S., Horst, S. J., Leaman, L., Flanagan, T., Neufeld, D. G. & Siderhurst, M. (2012). Measuring beyond content: A rubric bank for assessing skills in authentic research assignments in the sciences. *The Royal Society of Chemistry*. DOI: 10.1039/c2rp00023g.

Kohn, A. (2006). The trouble with rubrics. *English Journal*, *95*(4). Retrieved from http://www.alfiekohn.org/article/trouble-rubrics/

Kottner, J., Audige, L., Brorson, S., Donner, A., Gajewski, B.J., Hrobjartsson, A., Roberts, C., Shoukri, M. & Streiner, D. L. (2011). Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *Journal of Clinical Epidemiology*, *64,* 96–106.

Lantz, H. B. (2004). *Rubrics for assessing student achievement in science grades K-12.* California: Corwin Press.

Lauer, T., & Hendrix, J. (2009). A model for quantifying students learning via repeated writing assignments and discussion. *International Journal of Teaching and Learning in Higher Education, 20*(3), 425–437.

Lay, Y. F., Khoo, C. H., Treagust, D. F., & Chandrasegaran, A. L. (2013). Assessing secondary school students' understanding of the relevance of energy in the daily lives. *International Journal of Environment & Science Education, 8*(1), 199–215.

Lee, K. W., Tan, L. L., Coh, N. K., Chia, L. S., & Chin, C. (2000). Science teachers and problem solving in elementary schools in Singapore. *Research in Science and Technological Education, 18*(1), 113–126.

McMillan, J. H. (2011). *Classroom assessment: Principles and practice for effective standards-based instruction* (5th ed.). Boston, MA: Ally & Bacon.

Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation, 7*(25), Retrieved 11 November 2015 from http://PAREonline.net/getvn.asp?v=7&n=25

Millar, G. (2001). The development of indicators for sustainable tourism: Results of a Delphi survey of tourism researchers. *Tourism Management, 22,* 351–362.

Ministry of Higher Education Malaysia. (2012). *The national graduate employability blueprint 2012–2017*. Putrajaya: Author.

MTD Training. (2012). *Effective communication skills.* United Kingdom: Venture Publishing.

Nasri, N., Roslan, S. N., Sekuan, M. I., Abu Bakar, K., & Puteh, N. P. (2010). Teachers' perception on alternative assessment. *Procedia Social and Behavioral Sciences, 7,* 37–42.

NCREL & Metri Group. (2003). enGauge 21st century s kills: Literacy in the digital age. Retrieved from www.ncrel.org/engague

Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Upper Saddle River, NJ: Merrill.

North Central Regional Education Laboratory, NCREL. (2002). enGauge 21st century skills: Digital literacies for a digital age. Retrieved 20 January 2012, from http://www.ncrel/og/ engauge/skills/skills.htm

Osman, K., Abdul Hamid, S. H., & Hassan, A. (2009). Standard setting: Inserting domain of the 21st century thinking into the existing curriculum in Malaysia. *Procedia Social and Behavioral Sciences, 1,* 2573–2577.

Rezaei, A. R., & Lovorn, M. (2010). Reliability and validity of rubrics for assessment through writing: Assessing writing. *Practical Assessment, Research & Evaluation, 15,* 18–39.

Robles, M. M. (2012). Executive perceptions of the top 10 soft skills needed in today's workplace. *Business Communication Quarterly, 75*(4), 453–465.

Ross, R. G. (2011). Analysis and validation of a rubric to assess oral presentation skills in university contexts. *Electronic Journal of Research in Educational Psychology, 9*(25), 1043–1062.

Sasmaz-Oren, F., & Ormanci, U. (2011). Teacher candidate levels of familiarity with the methods, techniques and tools composing the alternative assessment approaches. *Procedia Social and Behavioral Sciences, 15,* 3476–3483.

Tuan Soh, T. M., Osman, K., & Mohamad Arshad, N. (2012). M-21CSI: A validated 21st century skills instrument for secondary science students. *Asian Social Science, 8*(16), 38–44.

Unal, Z., Bodur, Y., & Unal, A. (2012). Choosing or designing the perfect webquest for your learners using a reliable rubric. *Contemporary Issues in Technology and Teacher Education, 12*(2), 209–231.

Valerdi, R. (2011). *Convergence of expert opinion via the wideband Delphi method: An application in cost estimation models.* Massachusetts: The International Council of Systems Engineering (INCOSE).

Wolf, K., & Stevens, E. (2007). The role of rubrics in advancing and assessing student learning. *The Journal of Effective Teaching, 7*(1), 3–14.