Abstract

Objective: The purpose of this study is to measure an instrument that could assess Early Science and Mathematics Standard in the domain of science and technology. Children’s measurement standards in early years include the performance in thinking skills. The design is a combination of the theoretical framework representing Early Sciences and Mathematics Standard and Zone of Proximal Development (ZPD).

Method/Analysis: We developed a performance assessment standard and scoring set (rubric) for measuring children’s responses. Participants in this study were 30 children of 2-year old and 30 children who were 3 years old. These children were from several child-care centres. 21 items on scientific attitude, scientific skills, investigate the nature of life, pre-number experience, concepts of number, shapes and space, and construction were assessed. Mathematical models such as the Rasch model have provided useful representations of social science problems where it coordinates data with the requirements of a useful definition of measurement.

Findings: The study shows the Many-facet Rasch Measurement (MFRM) an extended version of the Rasch Model techniques which is combined with children’s measurement standards to examine the validity and reliability of scores for the performance rating scale. Additionally, an infit statistic (.76) for 2 years and (1.46) for 3 years for higher responses support the validity of scores. For reliability, the person reliability is good at .99, rater reliability is at .97, domain reliability is good at 1.00 and the item reliability is at .96 for overall scoring measure. Application/Improvement: This paper provides educators and researchers with a useful tool to facilitate measurement in early childhood years. It gives a great recommendation where the instrument is valid and reliable.

Keywords: early science and mathematics, domain of science and technology, performance assessment standard, mathematical model, MFRM.

1. Introduction

Recently, there have been important advances in reform movement in education, accompanied by a new progress on assessment. Emphasis on holistic development of children must be assessed that young people are individuals who demonstrate the ability to exceed the current target instruction that reflects the experience of a good education in the early years.
and had a positive effect on their learning in school. All levels of education have an achievement aligned with standard. Nowadays, latest education system from many states has involved into effort to develop early learning standards\textsuperscript{1,2}.

So that, in the new century, extra attention can be paid to the children’s assessment. Appropriate methods for the standard assessment of children are those that correctly document the broad range of skills that emerge and develop during the early years of life\textsuperscript{3}. For example, performance-based tasks are considered more suitable than traditional, multiple choice, paper-and-pencil items for younger children\textsuperscript{4}. This is because, performance-based tasks allow teachers to understanding children how they learn and actively develop meaning in the higher level context such as analysis, synthesis, and evaluation\textsuperscript{5,6,7,8}.

Furthermore, developmental tasks of children’s responses format on standard instrument, and focus on their learning and development, are always a controversial issue. The researchers point out the score to given which it depend of their level to assess response format as reliable and valid\textsuperscript{9,10,11}. For task development, especially, some items need children to give responses and their performance was guided by focus, content and learning standard and scoring rubrics have been constructed through a development theory for scoring set of response.(refer, portfolio). Outline given from the theoretical framework especially the guidelines for tasks and items development also scores format, but some kind of issues among the expert discussed, about how the scoring for their grading performances is to do\textsuperscript{12}. Without strong and enough base of prior knowledge and performance-based linking to standard, it would be difficult to find tools to measure learning and development children’s.

In view of all of this, it is important to examine performance of children using reliability and validity assessment. The importance of measuring young children appropriately has been an ongoing theme in this text. Contrary to many standardized tests and more formal strategies that have been criticized as in appropriate to the young child’s development, performance assessments can be good tools for evaluating progress in development; because they are designed to measure a child’s performance of a real or designed task or activity relevant to the desired learning, performance observations directly related to the child’s development and achievement\textsuperscript{6}.

In addition, performance assessments are integrally relates to instruction. The performance activity is a natural outcome of ongoing curriculum and instruction and not a separate, unrelated type of experience that is unfamiliar to the child. The terms such close relationship as “blurring the line between curriculum and assessment”\textsuperscript{13}. When using performance-based evaluation, the classroom teacher needs to know how to design appropriate, related assessment tools, interpret assessment results to understand the child’s progress and plan for further instruction, and interpret performance assessment results for parents and administrators\textsuperscript{14}.

In fact, performance assessments objectively used to assess whether children’s activities are meeting the needs of children. Good performance assessments tools help clarify the goals of preschool programs to provide standard curriculum. Progress assessment reflects both individual developmental progress and accomplishment of performance assessment goals\textsuperscript{6,15}. The teacher then has the responsibility to report overall accomplishments in a meaningful way to administrators and parents\textsuperscript{14}.

Therefore, in this research, we are trying to identify accurate levelled against the use the performance assessment as part of task development for learning and development standard. We also introduce a time-efficient, effective learning and development standard instrument and clear guidelines for scoring set which represent the children’s abilities to respond. These feedback are depended to the psychometric setting like reliability and validity. Participant in this study are two and three year old children who are asked to perform for every response-
item of the domain of science and technology. Responses are evaluated for how well children can do all items and solve the performance-based standard and implement the scoring set. The research questions in this study included:

1. Can an explanatory performance task, based in theories of cognitive, be developed for two and three year old children?

2. Can a scoring set be reliable and valid aligned with theory be developed to measure scores of the performance task?

   2a. to what degree is there support for the reliability of scores and to what degree can assignment of scores be performance responses to yield high inter-rater agreement?

   2b. to what degree is there support for the validity of performance assessment scores based on cognitive theory?

2. **Theoretical framework: combining theories of learning and development**

   In our view, children are dynamic individuals. Their learning involves the dynamic interaction especially in their cognition. So, while we learn about children’s cognitive development, we cannot forget that cognitive development is influenced by other domains of development as well. Children possess a variety of experiences, environments and their background, which influences how they think. Thinking serves an adaptive purpose as it enables children to devise plans for attaining goals. However, attaining goals also requires the ability to take action. Without this ability, thinking would be pointless. Children development through several stages to reach the maximum level. Assessments should be carried out to find out what level the child is at because of increasingly difficult tasks and less guidance given. Through the assessment, children can learn to do their assignments or otherwise, identify certain tasks which can only be done with the help of the teacher, or without, and tracking tasks whether easy or otherwise. From the theories of Piaget and Vygotsky (ZPD), transformation of the thinking level of children will be effectively implemented at task performance. Besides the theory linked, to the performance assessment which operates interactively between elements of children, the instrument, task performance, rating scale structure and rater. From the interaction, we look deeply it as essential to design a performance assessments standard and develop a scoring level that would give us to measure children individually.

3. **Domains of Learning and Development Standard**

   Early learning and development standard, also known as “early learning guidelines”\(^ {16}\) including a statements that specify an expectation for achievement. They reflect what children know and are able to do at a period of time during early childhood. From the observation, the standard will achieve and all the information will collected, so that it will be used to make comparison in evaluation, judging effect, quality, value or quantity\(^ {9}\). Clearly the terminology of standard are reflected in clustering of dimension and domains that illustrate holistic child learning and development. The standard also provides information on reasonable levels of behavior or other accomplishments for children at a given time during early childhood. It strongly results from many research evidence by Kagan & Britto about early learning and development standard which are considerate the cultural, linguistic, and socioeconomic different backgrounds. In their part of systematic process must looking into content indicator
and then age validation process to ensure that learning and development standard reflects the true norm of children in country.

3.1 Early Science and Early Mathematics Standard
Early science provides opportunities for children to explore their environment using simple equipment and sensory organs. In the process of this exploration, scientific attitude positive as inquisitive years, systematically and thoroughly develop. During the exploration process, the basic skills of science process skills such as observing, measuring using non-standard units, comparing and classifying will also be developed. Exploration conducted on the nature of life, (e.g. plants, animals, natural materials (e.g. submerged objects and arise, water, magnets) and the physical environment (e.g., rainbow, weather). Early math provides initial experiences of mathematics to children. Pre-number experience includes concept, numbers, operation simple numbers, and the money, the concept of time, shape and space.

3.2 Measurement of Early Science and Early Mathematics Standard Instrument
Instruments compiled based on three important components in the format of Focus, Content Standards and the Standards of Learning. Focus, Standard Content and Learning Standards are the areas referred to when teachers plan and implement teaching and learning. The focus is the field of study that would be developed in the child. Each focus is expressed through some Content Standards. Content standards are specific statements about what children must know and have knowledge to do in a period. It is written, covering the knowledge, skills and values. All content standards elaborate on some Learning Standards. Learning Standards is a setting of criteria or indicators in the form of behavioral objectives (behavioral) that can be measured, to ensure control of any Content Standards. ST describe for Science and Technology. Refer Figure 1. Early Science and Early Mathematics Standard.

Figure 1. Early Science and Early Mathematics Standard

4. ZPD and Scoring Set
Vygotsky define concept zone of proximal development (ZPD) as the area of children can develop better more than her or his peer and cause of it they can be more competent either adult or peer. This concept is about level ability each proximal development that children show their performance itself. Rather it is the difference between what the child can accomplish independently and what he or she can achieve in conjunction with another more competent person. The zone is thus created in the course of a social interaction. Moreover, Vygotsky also agrees about the concept of dynamic and interactive process.

“Learning is not development, however, properly organized learning results in mental development and sets in motion a variety of developmental processes that would be impossible apart from learning. Thus learning is a necessary part and universal aspects of the process of developing culturally organized, specifically human, psychological function”

Viewed from the perspective of rater cognition research, an important role is to be assigned to performance features as perceived by raters, and their mapping to the scoring criteria and the rating scale categories as usually spelled out in the scoring rubric. Figure 2 describes scoring set as reference.
5. Methodology

Performance is a very broad concept of assessment and also has a child and the different domains and abstract definitions. We will use specialized abstract concepts in order to create this item that assess children's behavior. Components and variables can be interpreted, based on theories of cognitive and socio-cultural activities as well as research literature. An assessment based performance brings a very important development in the assessment. He describes the nature of performance assessment that has changed the world of education because it is really to identify and measure the actual abilities of pupils. This assessment is more suitable and will be able to assess the ability of children to build complex thinking, how to identify problems, plan for the specific and focus on learning outcomes. Assessment as a better approach than a pencil and paper test for assessing the extent of the children's skills in the real world.

The history of performance-based assessment described in detail in the article "A Short History of Performance Assessment by Madaus & O'Dwyer, it is explained that the assessment is based on performance that is a new and an authentic approach to assess pupils in certain skills. However, it has long been used in the context of education, especially in certain areas or performance to be assessed is the reference to the objectives of the curriculum. The components involved in the assessment are based on performance (based on the direction of the arrow), children who show performance based on the instruments used, previously assessed based on the objectives set 22. Rating is in relation to the form of scoring to assess performance. Interactions that exist here are some rating with children.

5.1 Participant
Sixty children from two to three year in private nursery were involved in this study.

5.2 Raters
Two raters or examiners in this study are used to obtain reliable scores. Raters should have understand the rating scales of items and must have real experiences as well as content knowledge of the items and responds that to be measured 23. The raters were trained about the aim of this research, and can complete their scoring as responsibility to this study. Training Raters and Rating Scale Performance assessments describe as constructed-response items. Such items require children to show a response, and this is different from the traditional format that need choose answer given. To analyzing the score, it was depend to the rater’s ability how the data can interpreted and achieve the objectives. In keeping with rater cognition perspective of performance assessment, the process of assessing examinee performance can be described as a complex and indirect one: Children respond to measure’s items or tasks for each development domain or construct and raters look up into the response of performance and in the same time evaluate the level of their ability.

In this study, we set a specific objective to administer assessment on children for the reaction of all components and to be able to identify children's abilities, and the influence of rater effect, the domain difficulty and hardship items were analyzed using the Many-Facet Rasch Measurement Model, which is also the variable in Figure 3 as methodology framework:
This part of the framework have the facets and this facets have an effect to the children. The most important is, or should be, the proficiency of children who represent the instrument (construct) being measured (e.g. domain Early Science and Mathematics). Besides that, for measurement structure, an error will be find in the scoring set such as using the rating scale, or scales. That is, some categories of have different meaning between raters, within raters over time, between performance, in construct or between children ability \(^{13, 24}\). Other factors that are basically no one can be labeled as the cause to be as big factors or systematic measurement error in the rating are shown in the assessment. These include rater severity and scoring domain, item difficulty. For Example, raters may have different background demographic and it will effect their interpretation of the scoring using of likert scale; that is, some raters may actually perceive two adjacent categories, or the performance levels implied by these categories, to be much closer together than raters do.

Moreover, to illustrate that the route from children performance to an observed score or rating is represented by an arrow for children’s ability, rater effect, domain difficulty and item difficulty. In this research, this framework was implement and scoring set refers to the developmental rubrics.

6. Result and Discussion
To address both parts of research question two, us MFRM analyses was used. In addition, for this question, we used MFRM analysis. In addition, for question 2b, we analyzed the validity of Early Science and Mathematics domain, item and children’s scores.

Many-facet Rasch Model: Reliability
MFRM was used as a statistical framework for analyzing the data from the assessment process of learning and child development. Making decisions based on raw data collected from the score performance cannot describe the difficulty of a domain, the firmness of each rater, and the dimensions of a rating scale. How can the abilities of children be determined accurately, if firmness rater is difficult to determine the domain of children, and of rater tenderness to easily determine the domain scores of children? "How fair and meaningful measurements can be obtained from ordinal rating scores is uncertain" MFRM has determined what features of the data must be available to support a meaningful measurement\(^{25}\). When the data does not fit the model, meaning a measurement of unclear or questionable and requires further research.

It is a requirement when doing the actual observations, that the scores which are given by the rater show children become more organized to construct. The higher the score, the greater the number shows that latent areas are measured. MFRM will be able to demonstrate the value of the index to detect situations where the requirements are not met if the data is according to the qualitative methods. Fit statistics can indicate and identify data that works or does not work as it should. MFRM indicators provide information about pulling in the measurement, whether learning or developmental domains, too difficult or easy, whether severe or lenient rater or consistency in rating all these elements can be answered using the model.

Rasch measurement model provides information for researchers to measure the parallel constructs with qualitative observations. By using performance -based assessment, such as task this is the most accurate model and it is needed. Using the rater to make the observations and subjective measurement of work children show qualitative evaluation elements which can
be converted into a quantitative score. If the objective is to define and measure the ability of the child and the potential in him, great emphasis should be placed on the need to see the interaction between the rater with the difficulty of domain against children. MFRM quantitative comparison does not depend on who the child is assessed by. The rater. According to the software who does the rating, does not compare children but depends on the responds individually and the data is analyzed unidimensionality for each child, each rater and each item.

Based on work by reported the acceptable range of infit and outfit values as 0-5-1.5 26,27. If a value is <0.5, then that indicates there is little variation in the data. Essentially, there is a lot of redundancy in the data pattern. If the value is >1.5, then that indicates the data represents an extreme or aberrant case as in the instance of a child who does not get easy items correct but answers the more difficult ones accurately. Additionally, an infit statistic (.76) for 2 years and (1.46) for 3 years for higher responses support the validity of scores. For reliability, the person’s reliability is good at .99, rater reliability is at .97, domain reliability good at 1.00 and the item reliability is at .96 for overall scoring measure.

Table 1 gives the overall summary fit statistics for the person (children), rater and domain facets of the model from the Facets computer program. There are several things to note in this summary table. First of all, the person facet is not centered (M=.94, SD=.65). The raters facet is centered with a mean of zero (SD=.07), the domain facets is also centered at zero (SD. 55) and the item facets is also centered at zero (SD. 41)The convention in the Many Facet model is to center all of the facets except the one that represents the object of measurement. The summary statistics (Infit and Outfit), are close to the expected values of 1.00 with a standard deviation of .20 indicating fairly good model-d the rating scale structure for these data was examined.

Table 1 Overall summary fit statistic

6.1 Children Facet

Table 2 shows the measurement reports on children who are assessed as Early Science and Mathematics standard, RMSE, segregation index and the chi-square test. All sizes of children of 0.04 (SE = 0.06 for children 11) measures the ability of children to the lowest and the highest, 2.23 (SE = 0.09, 39 children). RMSE to measure children is 0.07. Separation index was 9.81 shows these children can be separated or divided into different capabilities. This was also confirmed by the chi-square is significant where measurement of children is the value of $x^2 = 5759.5, p <0.05, df = 59$. Analysis shows that the MFRM is used to evaluate the performance of a child by an appraiser who is able to separate the child to a different level. This analysis also shows that the difference in the ability of children is statistically significant. This shows that the results of the use of standards, children may discern at least 9 different levels. This is appropriate, as suggested by Linacre (2002), a good instrument will be able to separate students with at least two different levels. Items with values greater MNSQs infit and clothing items are considered 'misfitting'. These items should be eliminated from the analysis because each is 'sound' and does not contribute significantly to the measurement construct intended.

6.2 Rater

Table 2 shows the overall size rater assessments are available in various sizes rater firmness, RMSE, index isolation and chi-square. Rater various sizes firmness is between -0.07 (SE = 0.01, Rater 1) the size of the rater lowest and the highest 0.07 (SE = 0.01, Rater 2). RMSE is
0.01 to domain measurements. Index is 6.02 show separation can be separated or divided into different levels of difficulty. This was also confirmed by the chi-square where significant value measurement of \( x^2 = 74.5, p < 0.05, df = 1 \). Based on the analysis, it was found that the analysis done on the coherence of the assessment showed that each assessor have different emphases and different firmness proved by tests confirm significant chi-square that each appraiser has assessed different firmness. Assessing the emphasis varying potential or possibility of the existence of problems in measurement when using raw scores. It has been shown, which shows a significant change between the scores given by different appraisal of the performance of the same has been reported in several studies. To reduce the pressure difference between the appraisal, training courses, sessions can be conducted. However, the assessment cannot be trained to achieve the same severity \(^{28, 29}\), showed that the appraisal pressure\(^{30}\) which are different even though they have gone through a training course. Another alternative other than training that can be used by MFRM. MFRM to estimate the child's ability to independently measure items such as difficulty and severity of assessment. This model provides a measurement of "value-free", "free items" and "independent student" which means that the resulting size is not dependent on the sample or item or appraiser as long as they have a fit that matches the size Rasch model\(^ {31}\). In other words, the variable with the index should be less than 0.50 in this study means that the judges (faculty) scored high-skilled students, and students of low-skilled low (ie, variables over fit), while variables with index greater than 1.50 means that the judges scored the low-skilled students and less proficient students is high (ie, misfit change)\(^ {33}\).

6.3 Domain

Table 2 shows the overall domain measurement reports are available in various sizes, RMSE, index isolation and chi-square. Domain measurement is -0.25 (SE = 0.02), the size of 0.02 to domain measurements. 23:30 segregation index domain design can be separated or divided into different levels of difficulty. This was also confirmed by the chi-square where important measure of the domain \( x^2 = 3035.1, p < 0.05, df = 5 \).

6.4 Item

Table 2 shows the overall domain measurement reports are available in various sizes, RMSE, index isolation and chi-square. Domain measurement is -0.66 (SE = 0.06), the size of 0.00 for measurement domain. The index shows the separation is 5.19 domains can be separated or divided into different levels of difficulty. This was also confirmed by the chi-square where the item is important measure \( x^2 = 1.9.2186, p < 0.05, df = 60 \).

Table 2. Measurement Table

7. Conclusion

Study focused the measurement properties of a recently developed, brief measure of Standard Assessment. Many-Facet Rasch Model (MFRM) can analyzing the multiple level of data. The most important, the data fit to the model. It gave every types of data more reliable and valid for analyses. Measurement in early childhood was utilized, providing a flexible approach with a number of strengths. It is a requirement when doing actual observations, scores of children performances given by the rater, become well organized to construct or latent characteristics (abilities of children). The higher the score, the greater the amount of latent trait measured. MFRM is able to demonstrate the value of the index is to detect situations in which data need
not be met if the following qualitative methods. Fit statistics can indicate and identify data that works or does not work as it should. Indicators FRM provides information about dragging in the measurement, whether the domain of learning or development too difficult or easy, whether the rater expressly or otherwise, the criteria adopted by rater in a consistent state, whether the children are given a high score because rater generous or domain was pleased all these questions can be answered while using this model. Performance-based assessment is particularly relevant in the field of children when they are given the opportunity to demonstrate their abilities, then the observed performance is more natural and authentic or real. This assessment was performed using a variety of strategies and these strategies are aligned with the objectives of the study were assessed to be able to show the performance difference between the abilities of children, showing items or domain more difficult and influence by the rater's observation. This assessment in line with the standard curriculum in children and accounts for the diversity of children's development through selected samples were mixed.

8. Acknowledgment

Special appreciation to the Institute of Research Management & Monitoring (IPPP), University of Malaya Kuala Lumpur in allowing and giving grant for us to conduct this study. Also, to Ministry of Higher Education gave the scholarship to us similarly, the cooperation of respondents and teachers from all kindergartens involved.

9. References


### Tables and Figures

<table>
<thead>
<tr>
<th>Focus</th>
<th>Content Standard</th>
<th>Learning Standard</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST 1.0</td>
<td>ST 1.1 Demonstrate</td>
<td>ST 1.1.1 Demonstrate a systematic, cooperative, responsible, with guidance</td>
</tr>
<tr>
<td>Scientific</td>
<td>scientific attitude</td>
<td></td>
</tr>
<tr>
<td>Attitude</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ST 2.0</td>
<td>ST 2.1. Make</td>
<td>ST 2.1.4 Watching and told verbally about</td>
</tr>
<tr>
<td>ST</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scientific Skills, observations of objects and phenomena in the environment.</td>
<td>the observations that have been made.</td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>ST 2.3 Compare and classify objects</strong></td>
<td><strong>ST 2.3.1 Comparing and collecting objects by one characteristic:</strong> a color-b-form</td>
<td></td>
</tr>
<tr>
<td><strong>ST 2.3.4 Collecting objects have two characteristics in common:</strong> a-Color and shape b-Size and shape (e.g. green triangle, small round)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ST 3.0 Investigate the Nature of Life,</strong></td>
<td><strong>ST 3.2 Explore the environment using sensory organ</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ST 3.2.1 Using these senses in a guided exploration of the environment</strong> a-sight b-hearing c-touch d-sniff</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ST 3.2.4 Collecting objects have two characteristics in common:</strong> a-Color and shape b-Size and shape (e.g. green triangle, small round)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ST 6.0 Pre-number Experience,</strong></td>
<td><strong>ST 6.2 Comparing the quantity of objects</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ST 6.2.1 Comparing the two groups of objects to express the same, more or less</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>ST 6.4 Recognize recurring patterns and build pattern</strong></td>
<td><strong>ST 6.4.2 Duplicate the pattern of recurrent use of objects such as blocks, beads.</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ST 7.0 Concepts of Number,</strong></td>
<td><strong>ST 7.1 Understanding the Numbers 1-10</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ST 7.1.1 In the name of the number 1 to 10 (rote counting)</strong></td>
<td><strong>ST 7.1.3 Counts objects 1-10</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ST 11.0 Shape and Space,</strong></td>
<td><strong>ST 11.1 Knowing the position of objects in space</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ST 11.1.1 State the position of an object in space (front, back, top, bottom, inside and out)</strong></td>
<td><strong>ST 11.1.3 Placing objects in the direct</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ST 12.0 Construction</strong></td>
<td><strong>ST 12.1 Build a bridge between two objects</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ST 12.1.1 Build a bridge between two objects / structures and give reasons why the link is selected.</strong></td>
<td><strong>ST 12.3 Make a variety of building your own creativity</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ST 12.3.1 Make a variety of building their own creativity and communicating proxies with coaching</strong></td>
<td><strong>ST 12.3.2 Make use of a variety of different building materials according to their own creativity and communicating construction (eg materials: blocks, sand)</strong></td>
<td></td>
</tr>
<tr>
<td><strong>ST 12.3.3 Explore ways to build a strong and sturdy structure</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Early Science and Early Mathematics Standard

<table>
<thead>
<tr>
<th>Score</th>
<th>Description situation</th>
<th>Standard Criteria</th>
</tr>
</thead>
</table>
Table 1 Overall summary fit statistic

<table>
<thead>
<tr>
<th>Measure</th>
<th>Persons</th>
<th>Raters</th>
<th>Domains</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>0.94</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>SD</td>
<td>0.65</td>
<td>0.07</td>
<td>0.55</td>
<td>0.41</td>
</tr>
<tr>
<td>N</td>
<td>60</td>
<td>2</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>Facets</td>
<td>Measure</td>
<td>SE</td>
<td>RMSE</td>
<td>Separation Index</td>
</tr>
<tr>
<td>------------</td>
<td>-----------</td>
<td>--------</td>
<td>---------</td>
<td>------------------</td>
</tr>
<tr>
<td>Children</td>
<td>0.04-2.23</td>
<td>0.07</td>
<td>0.07</td>
<td>9.81</td>
</tr>
<tr>
<td>Raters</td>
<td>-0.07-0.07</td>
<td>0.01</td>
<td>0.01</td>
<td>6.02</td>
</tr>
<tr>
<td>Domain</td>
<td>-0.84-0.96</td>
<td>0.02</td>
<td>0.02</td>
<td>23.30</td>
</tr>
<tr>
<td>Item</td>
<td>-0.64-1.39</td>
<td>0.08</td>
<td>0.08</td>
<td>5.19</td>
</tr>
</tbody>
</table>

Table 2. Measurement Table