CHAPTER II
CHAPTER II
REVIEW OF RELATED LITERATURE

“Assessment performances are day-to-day activities that can also be authentic and engaging demonstrations of students’ abilities to grapple with the central challenges of a discipline in real life contexts” (Kulieke, Bakker, Collins, Fennimore, Fine, Herman, Jones, Raack, & Tinzmann, 1990, p.2).

2.1 INTRODUCTION

The objectives of this study are to investigate the effect of traditional assessment and performance assessment on meta-cognitive skills and academic achievement of pre university chemistry students.

This chapter draws on the areas of educational research in meta-cognition, assessment, performance assessment and learning. In this section, the role of performance assessment in each area and all the perspectives to support the hypothesis will be examined. Then the role of performance assessment that can serve learning by increasing meta-cognitive skills and thereby improving performance will be investigated.

2.2 METACOGNITION

According to Lai (2011), Meta-cognition is a multidimensional set of skills that involve “thinking about thinking.” Meta-cognition entails two components: meta-cognitive knowledge and meta-cognitive regulation. Meta-cognitive knowledge includes knowledge about oneself as a learner and about the factors that might impact performance (declarative), knowledge about strategies (procedural), and knowledge about when and why to use strategies (conditional). Meta-cognitive regulation is the monitoring of one’s cognition and includes planning activities, monitoring or awareness of comprehension and task performance, and evaluation of the efficacy of monitoring processes and strategies. Insights experienced while monitoring and regulating cognition play a role in the development and refinement of meta-cognitive knowledge. In turn, cognitive knowledge appears to facilitate the ability to regulate cognition.

Akturk & Sahin (2011), viewed meta-cognition as a structure that refer as fuzzy by many scholars and has very diverse meanings. Much research has been
conducted for more than 30 years in order to access the inner side of this structure, which is really hard to grasp.

Leader (2008), considered meta-cognition as an umbrella term encompassing the structures that relate to individuals’ thinking processes and information.

Buoncristiani (2008), reported that meta-cognition is the conscious application of an individual’s thinking to their own thought processes with the specific intention of *understanding, monitoring, evaluating and regulating* those processes.

Meta-cognition is thinking about one's thinking. In order to be meta-cognitive, students must know how and know the need to think about their thinking. Teachers can assist students to acquire a set of strategies, define goals, and monitor their progress (Darling-Hammond, et al., 2008). Providing opportunities for students to reflect on what and how they learn creates an environment where students take responsibility for their learning and become more of a partner with their teacher in engaging in meaningful learning experiences.

A review of the literature reveals that the most widely accepted definition of meta-cognition is knowledge that includes (a) awareness of one’s personal abilities (declarative knowledge), (b) general strategies that might be used for different tasks (procedural knowledge), and (c) knowledge of the conditions under which these strategies might be used, as well as knowledge of the extent to which the strategies are effective (conditional knowledge) (Flavell, 1979; Pintrich, Wolters, & Baxter, 2000; Pressley, Van Etten, Yokoi, Freebern, & Van Meter, 1998).

McCormick (2006) argued further about the definition and delineates of characteristics represented within the three different types of meta-cognitive knowledge. The first, declarative knowledge includes knowledge about oneself as a learner and about what factors influence one’s performance. Good learners appear to have more knowledge about different aspects of memory, such as capacity limitations, rehearsal, and distributed learning. The second is procedural knowledge and refers to knowledge of how to do things. Much of this knowledge is represented as heuristics and strategies. Individuals with a high degree of procedural knowledge perform tasks more automatically, are more likely to possess a larger repertoire of strategies, to sequence strategies effectively, and qualitatively use different strategies to solve problems. The third is conditional knowledge, which refers to knowing when and why
to use declarative and procedural knowledge. Conditional knowledge is important because it helps students selectively allocate their resources and use strategies more effectively. Conditional knowledge also enables students to adjust to the changing situational demands of each learning task.

According to Swartz (2001), meta-cognition is a conscious examination of one’s thinking; an individual needs to be aware that the focus of their thinking is on their own thought processes. The purpose of meta-cognition is for one or more of the four intentions listed in the definition. These intentions represent increasingly more sophisticated stages of meta-cognition. The first intention of meta-cognition is, understanding which implies an awareness that one’s thinking is directed toward one’s own thought processes. The second intention, monitoring, is checking to see that the thinking is on the right track. This includes an awareness of the two modalities of thought – the type of thinking being used and the disposition of the thinker toward those thoughts and their consequences. Are the results of the thinking reasonable? Is the right type of thinking being used? Are appropriate habits of mind exercised? The third intention, evaluation, involves an assessment of how well the thinking is proceeding toward its objective. The final intention, regulation, involves adjusting the thought process to make sure the objective is attained and when the objective is attained, reviewing the thinking and modifying the thought process so that it will be even more effective the next time it is used.

Costa (2007) considered meta-cognition or “thinking about one’s own thinking” emerges as a vital element of effective thought.

Researchers working in the field of cognitive psychology have offered the following definitions of meta-cognition: “Awareness and management of one’s own thought” (Kuhn & Dean, 2004, p. 270). “The monitoring and control of thought” (Martinez, 2006, p. 696).

Khoramde et al. (2012) investigated validation of the meta-cognitions Questionnaire—Adolescent Version in an Iranian sample and compare of met cognitive beliefs between adolescents with anxiety disorders and normal sample. First the original version was translated into Persian then administered to 204(101boy&103 girl) Adolescent aged 13 through 17 years that by cluster randomly were selected from the schools of Isfahan city, together with Mood and Feelings Questionnaire and
Revised Children's Manifest Anxiety Scale. Also 17 adolescents with anxiety disorders and normal sample were completed the (MCQ-A). In order to assess reliability, method of internal consistency (chronbach’s alpha and split-half Coefficient) was used, and also in order to assess validity, Convergent Validity, criterion validity and confirmatory factor analysis was used. The results of correlation coefficient of Convergent Validity showed a relation between total factor of (MCQ-A) and its components with anxiety and depression except Cognitive self-consciousness. Data were indicative of appropriate level of Coranbach's Alpha and Split-half reliability coefficients of the (MCQ-A) and extracted factors. Results of multivariate analysis of variance showed that there were significant differences between the 2 groups with regard to the total factors except positive beliefs and negative beliefs about worry. The results of factor analysis by principle components analysis and using varimax rotation showed 5 factors that account for 0.45% of the variance. Conclusion: the (MCQ-A) has satisfactory Psychometric Properties in Iranian sample.

Yildiz et al. (2009) conducted a study to develop the meta-cognition Scale (MS) which was designed for primary school students. The sample of the study consisted of 426 primary school students in Izmir, Turkey. In order to examine the construct validity of the MS, exploratory factor analysis and confirmatory factor analysis were performed. For the validity of the MI, corrected item-total correlations were used. The corrected item-total correlations ranged from .35 to .65. In addition, t-tests between items’ means of upper 27% and lower27% points were compared. For each factor and each item, the differences between mean scores of upper 27% and lower 27% groups were significant. Finally, cronbach alpha correlation coefficients were used. The internal consistency of the MS was .96 for the entire scale. The MS has eight scales: declarative knowledge, procedural knowledge and conditional knowledge, planning, self-control, cognitive strategies, self assessment and self monitoring. According to these findings, the MS was appropriate for researchers or teachers whose aim was to measure his/her students’ meta-cognitive awareness and meta-cognitive abilities.

Cooper, Sandi-Urena & Stevens (2008) undertook a study to develop a reliable multi method assessment of meta-cognition use in chemistry problem solving. This paper describes an across-method-and-time instrument designed to assess the use of
meta-cognition in chemistry problem solving. This multi method instrument combines a self report, namely the Meta-cognitive Activities Inventory (MCA-I), with a concurrent automated online instrument, Interactive Multi-Media Exercises (IMMEX). IMMEX presents participants with ill defined problems and collects students’ actions as they navigate the problem space. Artificial neural networks and hidden Markov modeling applied to the data collected with IMMEX produce two assessment parameters: the strategy state, which is related to the meta-cognitive qualities of the solution path employed, and the ability which is a measure of the problem difficulty students can properly handle. The ability values are significantly correlated with the MCA-I scores, and groups of students who performed using more meta-cognitive state strategies had significantly higher mean MCA-I values than those using fewer meta-cognitive strategies. This evidence is indicative of convergence between the methods.

2.2.1 COMPONENTS OF METACOGNITION

There are three general objects of meta-cognition 1) the content of our thoughts, 2) the cognitive skills being used and 3) the conduct in support of thinking.

1) The Content of Thought:

When we think about our thinking we sometimes need to focus on the content of our thought; that is just what we are thinking about right now. It might be knowledge we already possess as we try to recall some specific information; it might be a concept we are trying to understand; it might be a problem we are trying to solve; it might be plans we are trying to formulate. We can think about our thoughts as we are thinking of them. The reason for meta-cognition about the content of our thought is to monitor our own understanding of concepts involved or to track the progress of thinking toward the objective of thought or to check for consistency with other knowledge.

2) The Cognitive Skill Being Used:

Sometimes we may need to think about the cognitive skill being used; that is the type of thinking we are engaged in to achieve our goal. The objective of meta-cognition about the cognitive skill being used is to ensure that the right thinking skills
are brought to bear on the problem at hand and to sharpen these skills for future use. We may also need to seek alternative means of reaching or justifying our conclusions.

3) Personal behavior supporting thinking:

In addition to the thoughts running through one’s mind we may also reflect on specific behaviors supporting the thinking. How diligent is the thinker? Is the thinker using all available resources? The objective of meta-cognition about one’s behavior is to develop these behaviors into the habits of a successful thinker. Table 2.1 summarizes the objectives and intentions of meta-cognition (Buoncristiani, 2008).

Table 2.1: Components of Meta-cognition (Buoncristiani, 2008).

<table>
<thead>
<tr>
<th>Components of Meta-cognitive</th>
<th>Objectives (reasons to meta-cognition)</th>
<th>Languages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>To monitor understanding concepts or track progress in problem solving or formulating plans</td>
<td>Discipline</td>
</tr>
<tr>
<td>Content</td>
<td>To achieve the objective of one’s thinking and to sharpen thinking skills</td>
<td>Types of thinking</td>
</tr>
<tr>
<td>Content</td>
<td>To develop the habits of a successful thinker</td>
<td>Habits of mind</td>
</tr>
</tbody>
</table>

2.2.2 ELEMENTS OF METACOGNITIVE SKILLS

Meta-cognition has two constituent parts: knowledge about cognition and monitoring of cognition (Cross & Paris, 1988; Flavell, 1979; Paris & Winograd, 1990; Schraw & Moshman, 1995; Schraw et al., 2006; Whitebread et al., 2009). Several frameworks have been developed for categorizing types of knowledge about cognition. Table 2.2 organizes components from each of these frameworks to facilitate comparisons among them. For example, Flavell (1979) defines cognitive knowledge as knowledge about one’s own cognitive strengths and limitations, including the factors (both internal and external) that may interact to affect cognition. He classifies such knowledge into three types: (1) “person” knowledge, which
includes anything one believes about the nature of human beings as cognitive processors; (2) “task” knowledge, which includes knowledge about the demands of different tasks; and (3) “strategy” knowledge, which is knowledge about the types of strategies likely to be most useful.

Table 2.2: Typology of Meta-cognitive Components (Flavell, 1979)

<table>
<thead>
<tr>
<th>Meta-cognitive Component</th>
<th>Type</th>
<th>Terminology</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge about</td>
<td>person and task knowledge</td>
<td>Flavell, 1979</td>
<td></td>
</tr>
<tr>
<td>Oneself as a learner</td>
<td>self-appraisal</td>
<td>Paris &amp; Winograd, 1990</td>
<td></td>
</tr>
<tr>
<td>And factors affecting Cognition</td>
<td>Epistemological understanding</td>
<td>Kuhn &amp; Dean, 2004</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Declarative Knowledge</td>
<td>Cross &amp; Paris, 1988</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Schraw et al., 2006</td>
<td>Schraw &amp; Moshman, 1995</td>
<td></td>
</tr>
<tr>
<td>Cognitive knowledge</td>
<td>Procedural knowledge</td>
<td>Cross &amp; Paris, 1988</td>
<td></td>
</tr>
<tr>
<td>Awareness and</td>
<td>Strategy knowledge</td>
<td>Kuhn &amp; Dean, 2004</td>
<td></td>
</tr>
<tr>
<td>Management of</td>
<td>Schraw et al., 2006</td>
<td>Schraw &amp; Moshman, 1995</td>
<td></td>
</tr>
<tr>
<td>cognition, including</td>
<td></td>
<td>Whitebread et al., 2009</td>
<td></td>
</tr>
<tr>
<td>knowledge about strategies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge about</td>
<td>Conditional knowledge</td>
<td>Schraw et al., 2006</td>
<td></td>
</tr>
<tr>
<td>Why and when to use A given strategy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identification and</td>
<td>Planning</td>
<td>Cross &amp; Paris, 1988</td>
<td></td>
</tr>
<tr>
<td>Selection of appropriate</td>
<td>or</td>
<td>Paris &amp; Winograd, 1990</td>
<td></td>
</tr>
<tr>
<td>Strategies and allocation</td>
<td>regulation</td>
<td>Schraw et al., 2006</td>
<td></td>
</tr>
<tr>
<td>Of resources</td>
<td></td>
<td>Schraw &amp; Moshman, 1995</td>
<td></td>
</tr>
<tr>
<td>Cognitive regulation</td>
<td>Monitoring or regulation</td>
<td>Cross &amp; Paris, 1988</td>
<td></td>
</tr>
<tr>
<td>Attending to and being</td>
<td></td>
<td>Paris &amp; Winograd, 1990</td>
<td></td>
</tr>
<tr>
<td>Aware of comprehenson</td>
<td></td>
<td>Schraw et al., 2006</td>
<td></td>
</tr>
<tr>
<td>And task performance</td>
<td></td>
<td>Schraw &amp; Moshman, 1995</td>
<td></td>
</tr>
<tr>
<td>Cognitive Experiences</td>
<td></td>
<td>Whitebread et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Assessing the processes</td>
<td></td>
<td>Cross &amp; Paris, 1988</td>
<td></td>
</tr>
<tr>
<td>And products of one’s</td>
<td></td>
<td>Paris &amp; Winograd, 1990</td>
<td></td>
</tr>
<tr>
<td>Learning and revising</td>
<td></td>
<td>Schraw et al., 2006</td>
<td></td>
</tr>
<tr>
<td>And revising learning</td>
<td></td>
<td>Schraw &amp; Moshman, 1995</td>
<td></td>
</tr>
<tr>
<td>Goals</td>
<td></td>
<td>Whitebread et al., 2009</td>
<td></td>
</tr>
</tbody>
</table>
Subsequent meta-cognition researchers have offered a slightly different framework for categorizing cognitive knowledge. For example, several researchers have used the concepts of declarative and procedural knowledge to distinguish cognitive knowledge types (Cross & Paris, 1988; Kuhn, 2000; Schraw et al., 2006; Schraw & Moshman, 1995). Kuhn and Dean (2004) characterized declarative cognitive knowledge broadly as epistemological understanding, or the student’s understands of thinking and knowing in general. Schraw et al. (2006) portrayed declarative cognitive knowledge as knowledge about oneself as a learner and what factors might influence one’s performance. Paris and Winograd (1990) discuss the process of self-appraisal as reflection about personal knowledge states to answer the question, “Do I know this?” Finally, Cross and Paris (1988) define declarative cognitive knowledge specifically within the context of reading as awareness of the factors that might affect reading ability.

On the other hand, procedural knowledge involves awareness and management of cognition, including knowledge about strategies (Cross & Paris, 1988; Kuhn & Dean, 2004; Schraw et al., 2006). Schraw et al. (2006), also distinguish conditional cognitive knowledge, which is knowledge of why and when to use a given strategy. The authors point out that cognitive knowledge is “late developing,” in the sense that children often exhibit deficits in cognitive knowledge. In addition, although the ability to explicitly articulate cognitive knowledge tends to improve with age, many adults struggle to explain what they know about their thinking. This latter result suggests that cognitive knowledge may not need to be explicit in order for people to access and use it.

The other component of meta-cognition is monitoring of one’s cognition, which many researchers have argued includes activities of planning, monitoring or regulating, and evaluating (Cross & Paris, 1988; Paris & Winograd, 1990; Schraw & Moshman, 1995; Schraw et al., 2006; Whitebread et al., 2009). Planning involves identification and selection of appropriate strategies and allocation of resources, and can include goal setting, activating background knowledge, and budgeting time. Monitoring or regulating involves attending to and being aware of comprehension and task performance and can include self-testing. Finally, evaluation is defined as “appraising the products and regulatory processes of one’s learning,” and includes revisiting and revising one’s goals (Schraw et al., 2006, p. 114).
The research of Hartman (2001), McCormick (2006), as well as Schraw and Dennison (1984) identify four main types of strategic knowledge that are essential for students to learn to become effective meta-cognitive thinkers. These components include (a) planning, which helps the learner define what the problem is, and select an appropriate solution strategy, (b) monitor the effectiveness of the solution strategy, and (c) regulate themselves while learning in order to identify and overcoming obstacles to solving the tasks in front of them and (d) evaluating the end results. These four key components and the elements they include are shown in Table 2.3.

Table 2.3: Key Meta-cognitive Elements (McCormick 2006)

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Indicator Behaviors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Planning</td>
<td>1. Making predictions</td>
</tr>
<tr>
<td></td>
<td>2. Being aware of what is already known so appropriate strategies can be selected</td>
</tr>
<tr>
<td></td>
<td>3. Sequencing those strategies</td>
</tr>
<tr>
<td></td>
<td>4. Allocating time and attention that affect performance</td>
</tr>
<tr>
<td>Monitoring</td>
<td>1. Identifying the task</td>
</tr>
<tr>
<td></td>
<td>2. Checking one’s on-line awareness of comprehension and task performance</td>
</tr>
<tr>
<td></td>
<td>3. Deciding whether, in light of new information, a path already taken should be abandoned and what, if anything, can be salvaged from an abandoned attempt</td>
</tr>
<tr>
<td></td>
<td>4. Looking for previously overlooked information and identifying ways to combine information</td>
</tr>
<tr>
<td></td>
<td>5. Predicting the eventual outcome</td>
</tr>
<tr>
<td></td>
<td>6. Engaging in periodic self-testing</td>
</tr>
<tr>
<td>Regulating</td>
<td>1. Allocating resources and number of steps needed to complete a task</td>
</tr>
<tr>
<td></td>
<td>2. Being mindful of the intensity and speed with which a task must be completed</td>
</tr>
<tr>
<td></td>
<td>3. Using existing strategies to the learner’s best advantage</td>
</tr>
<tr>
<td></td>
<td>4. Increasing awareness of comprehension breakdowns</td>
</tr>
</tbody>
</table>
2.2.3. METACOGNITION AND LEARNING

When meta-cognition is applied to teaching, teaching and assessment become integrated. Similar to the learning-assessment integration model described above, teaching involves assessing (monitoring and evaluating) the extent to which learning is taking place, and adapting/adjusting (planning and regulating) to make teaching more effective. This integrative-meta-cognitive approach helps learners become more aware of what they learn, how they learn, and what helps them to learn. They become engaged in reflecting on the evidence of learning, setting their learning goals, sharing learning plans and criteria for success, and evaluating their progress through dialogue and self and peer assessment. As meta-cognitive learners-teachers-assessors, they are empowered to build a knowledge base for themselves, make connections between the different pieces of knowledge they have learnt and take more responsibility for their learning. Being active participants in the learning process, they gain a stronger ownership of their learning (Te Puna Puoru National Centre for Research in Music Education and Sound Arts, 2010).

This synergistic meta-cognitive learning-teaching-assessment model can be presented as a three prong integrated process:

![Learning-teaching-assessment three-prong model](image)

**Figure 2.1: Learning-teaching-assessment three-prong model (Schraw et al., 2005).**
The influence and relevance of meta-cognition in learning has been extensively demonstrated (Veenman et al., 1997; Georgiades, 2000; Pintrich, 2002; Schraw et al., 2005), and the findings suggest that it may even be more important for problem solving success than aptitude (Swanson, 1990). It has also been suggested that it may play a compensatory role for cognitive skills and motivation in the learning of chemistry (Schraw et al., 2005).

Two reports from the National Academy of Sciences (NAS, 2000, 2005) have identified three principles of learning that are important for helping students develop understanding. These principles include (a) engaging students’ preconceptions of how the world works, (b) developing a deeper understanding of content knowledge, and (c) teaching meta-cognitive skills that can help students learn to manage their own learning by defining learning goals and actively monitoring their progress in achieving them (Alexander & Winne, 2006). This study focused on this third principle.

Hoffmann (2000) studied that for whatever reason, “many college students have difficulty in figuring out a right strategy to learn in the many courses they take. Even after the instructor adopts good instructional practices that supposedly help students learn content specific information, learning may not occur fully unless a student learns to monitor his or her own learning” (p. 380). Second, Lovrich (2004) claims that “students often believe that if they just think harder about a problem a solution will follow. However, thinking about one’s thinking can be a more productive expenditure of mental energy” (p. 57). Third, Dearnley and Matthew (2007) report that “we must remember that these students are starting from the position of being silent receivers of knowledge instead of actively generating their own knowledge” (p. 383).

Hartman (2001) reported that students who don’t progress past the acquisition of basic study skills lack the meta-cognitive knowledge needed at the college level and “seem to have little knowledge of what they are doing when performing a task” (p. 35). These students generally have a hard time performing the following learning tasks:

1. Determining the difficulty of a task
2. Monitoring their comprehension effectively
3. Planning ahead
According to Leamnson (1999), an average only about half of the matriculating first-year college students in the United States are over graduate. This sizable attrition rate suggests that students who come directly from secondary education are not well prepared in a number of areas for college learning. One reason they may flounder may not be because they lack knowledge but rather strategies for using what they know to attain success by meta-cognitively guiding their learning. Therefore, to help these students gain the study skills necessary to move from surface learning to deep learning, teaching the meta-cognitive skills and strategies of planning, monitoring, regulating and evaluating, coupled with explicit reflection, is a viable option teachers can introduce (or reintroduce) to this population. Hartman (2001) claims that student’s perceptions of the amount of personal control they have over their own learning have important implications for student retention. He reports that a number of studies claim significant improvement in student learning when regulatory skills, like those emphasized in meta-cognitive instruction, as well as an understanding of how to use these skills, are included as part of normal classroom instruction.

Dunning, et al. (2003) considered meta-cognition as an important element in learning and a strong predictor of academic success. Students with good meta-cognition demonstrate good academic performance compared to students with poor meta-cognition. Students with poor meta-cognition may benefit from meta-cognitive training to improve their meta-cognition and academic performance.

Smilksteins' research (2003) revealed a difference between those students who use meta-cognitive strategies and their peers who do not. One important difference is that “meta-cognitive learners separate their intellectual capacity from their lack of knowledge and have confidence in their ability to learn when and if given the opportunity to experience or be exposed to a particular object of learning” (p. 111). One study, which focused on teaching meta-cognition to pre-service education students, (Wilburne, 1997) found that those students that were taught meta-cognitive
skills consistently scored higher on tests and showed significant improvement in other cognitive processes associated with learning over other students that were not taught meta-cognitive skills. According to Cooper (2004), researching meta-cognition in older, returning college-age students, and Gott, Lesgold, & Kane, (1996), who studied student transfer of technical competence, all report that meta-cognitively aware students are more active and independent as learners, engage in more reflecting, monitoring and assessing of their actions and cognitive processes than their less meta-cognitively aware peers who were found to be more passive and instructor-dependent.

These findings suggest that meta-cognitive knowledge plays a compensatory role that students can use to enhance their learning, leading both Cooper (2004) and Gott et al. (1996) to claim that a student’s knowledge of meta-cognitive skills and strategies may predict learning performance.

Yap (1993) suggested that cognitive strategies and self-checking behaviors are part of a series of state meta-cognitive learning behaviors that can enhance learning.

Dawson (2008) showed that relationship among meta-cognition, assessment, motivation and learning in the following diagram.

Tansaz & Tavou (2011) studied the effects of meta-cognition on educational achievement of the Azad university students in special texts in English language. After the final examination, 32 students with the higher grades and 32 students with the lower grades from the 128 students of the 4 classes were chosen and their meta-cognitive skills were compared. In this research, meta-cognitive skills were considered as independent variables, academic achievement as the dependent variable and sex as the control variable. The result showed a significant difference between strong and weak group according to the meta-cognitive levels. In addition, the sub-scales scores (awareness, cognitive strategy, planning and self-control) were significantly different in two group.

Khamene & Seif (2011) investigated the effect of teaching meta-cognition approaches of reading on the motivation and academic achievement of high school female students of Tehran. The current study was a pre test–post test study and it has been done with a control group and a experimental group. The population of this study was all of the high school female students of Tehran (Iran). The sample has been 56 first grade students that were chosen by random cluster method. In the pre test level the motivation and academic achievement of students in both control group and experimental group has been the same. After 10 session of teaching meta-cognition approaches of reading the difference between two groups became significantly high. The experimental group had very higher marks than the control group.

Zare & Mohammadi (2011) reported the effects of meta-cognition learning and its approaches in solving mathematic problems in first grade high school students in Yazd (Iran). The method used in this study has been quasi experimental. To do so 120 of first grade high school students participated in two control and experimental group. The experimental group had a 6 week and 4 hour per week program of meta-cognition methods. Before the experiment a pretest was taken from all 120 students, the tool used gathering data was a teacher-constructed math test that its content validity and reliability has been confirmed. The results have been analyzed with T-Test method and showed that there is a significant difference between two groups, as the scores mean of experimental group was significantly very better than the control group. In general this study indicated that meta-cognition methods of learning have a
positive effect on math problems solving of students and encouraging them to learn mathematics.

Shen & Liu (2011) examined the effect of the web-based learning environment in developing of meta-cognitive skills. A pretest-posttest quasi-experimental design was used in this study. Fifty-three college students were assigned into experimental and control groups. After four-week training period, the results of paired-samples t-test showed that experimental group’s posttest scores were significantly higher than the pretest scores in self-plan, self-monitor, and total score, while there was no significance in the control group. In addition, students in experimental group made significantly greater gains compared to control group in self-plan. Discussion and suggestions are also provided.

Ghiasvand (2010) conducted a study to compare learning strategies between under-achiever and upper-achiever students (including both genders in 3 school grades). This study was designed in retrospective framework. Subjects were high school students in Qazvin Province selected by random multi-level cluster sampling method. Among the samples, two 90-person groups were chosen as upper and under achievever students. Participants completed the Learning and Study Skill Inventory (LSSI) form. This inventory assesses cognitive and meta-cognitive learning strategies. To test the hypotheses, independent t-test, one way ANOVA and multivariate regression method were used. Findings showed that upper students used cognitive and meta-cognitive strategies more than the lower group (p<0.001). Girls used LSS more than boys (p<0.000). No significant difference was found among school grades in using of LSS. Meta-cognitive strategies predict academic achievement more effectively than cognitive strategies.

Javadi, et al. (2010) investigated the relationship between meta-cognition awareness of reading strategies and students’ academic status in Isfahan University of medical sciences. In this study 191 students were selected through stratified random sampling out of total students of schools of Isfahan University of Medical Sciences in 2008-2009 academic years. Meta-cognition Awareness of Reading Strategies Inventory (Marsi) was distributed among study sample. Academic status of students was defined based on their grand point average. Data was analyzed by SPSS version 10 software using Pearson correlation, independent t, and variance analysis tests. The
results showed the mean and standard deviation of meta-cognition awareness score of students was 66.62± 14.8. The mean score of meta-cognition awareness showed a significant correlation with students’ score of academic status meaning that students with average score higher that 17, had significantly higher meta-cognition awareness. Meta-cognition awareness showed no significant relationship with variables of age, gender, and residence place. But, there was a significant relationship with academic level, so as MS students had higher meta-cognition score compared to that of PhD students.

Chari, et al. (2010) investigated the reliability and validity of questionnaire of meta-cognition strategies. The population was from high school male students in Bandar Abbas (Iran); 297 samples of first year students were chosen by cluster sampling method. The tool of the study was a questionnaire consisting 30 questions in meta-cognition strategies. To investigate the validity of the questionnaire the exploring factor analysis method with varimax rotation was used. The indexes gained for the sample competence and variables for factor analysis were satisfactory. After that the strategies of general meta-cognition, supportive meta-cognition and solving meta-cognition factors were extracted and almost %59.27 of the general variance confirmed the meta-cognition strategies of reading meta-cognition. Also the reliability of the questionnaire gave us alpha coefficient as 0.70 and half as 0.75. In general the results of this study indicated that questionnaire of strategies of meta-cognition has been useful for the high school students of Bandar Abbas and can satisfy the needs to some point.

Safari & Marzoughi (2010) reported the effects of meta-cognition approaches of teaching on educational condition and meta-cognition intelligence of students in their elementary school students. In order to do so from a statistic population of 8271 students, 8 classes with 228 students have been chosen from the third grade of elementary school in district number 4 of educational ministry of Shiraz (Iran) by random cluster method. They were put in experimental group (111 persons) and control group (117 persons). The tool to collect data has been a questionnaire of analyzing the meta-cognition intelligence and academic achievement test which was confirmed in statistic population. The results have shown that the grades, academic achievement and meta-cognition intelligence of the experimental group were significantly higher than the control group (>p.001). The difference of the two groups
were not so much based on the gender difference for their educational condition but it was significant for their meta-cognition intelligence (>p.001).

Ozsoy & Ataman (2009) studied the effect of using meta-cognitive strategy training on mathematical problem solving achievement. The study took place over a nine-week period with 47 fifth grade students. The experimental group (n=24) instructed to improve their meta-cognitive skills. At the same time the students in the control group (n=23) received no additional activities and continued their normal lessons. Students were pre- and post-tested with the Mathematical Problem Solving Achievement Test and Turkish version of Meta-cognitive Skills and Knowledge Assessment (MSA-TR). The results indicated that students in the meta-cognitive treatment group significantly improved in both mathematical problem solving achievement and meta-cognitive skills.

Nbina (2010) examined the effect of instruction in meta-cognitive self assessment strategy on senior secondary school students’ Chemistry self-efficacy and achievement. The study also explored the interaction effects of instruction in meta-cognitive self assessment strategy and genders in their Chemistry self-efficacy and achievement. The study was guided by five research questions and four hypotheses. A non-equivalent control group pretest and posttest design involving one treatment and one control group was adopted. A total of 192 SS 2 students from Port Harcourt Education zone were used for the study. The Self Assessment Instructional Programmed (SAIP) was developed, validated and used for the study. Three instruments: Chemistry Achievement Test (CAT), Self Assessment Scale (SAS) and Chemistry Self-efficacy scale (CSS) were adopted, validated and used for data collection. The results suggested that instruction in the meta-cognitive self assessment strategy improve the students’ chemistry achievement and self-efficacy.

Erskine (2009) applied the meta-cognitive strategies to improve students learning and performance. Incoming college and university freshmen are not typically trained in using the meta-cognitive skills that could enhance their academic performance and their satisfaction with the college experience. This study attempted to assess first-year university students’ meta-cognitive awareness and usage at two levels: (a) After direct and specific meta-cognitive training, (b) after engaging in weekly meta-cognitive reflection assignments. Six classes of university freshmen
were studied in terms of their use of meta-cognitive skills and strategies as they progressed through their initial semester. Four of the six classes were trained in meta-cognitive skills and strategies using the Meta-cognitive Skill Instruction. Two of these four classes were prompted to specifically reflect on their use of meta-cognitive skills and strategies. The other classes were not prompted about their use of meta-cognition. Students’ meta-cognitive performance was assessed at the end of the semester using the Meta-cognitive Awareness Inventory. Results show there was no initial difference between groups yet a significant difference between posttest and retrospective pretest scores was found for all three groups at the end of the term.

Chalmers (2009) investigated the development of group meta-cognition by three small groups of middle-grade primary school students engaged in the collaborative construction of computer-based mathematical models. The three groups of students were part of a cohort of 30 Grade 4-5 students engaged in the construction of mathematical models within the context of a computer-supported collaborative learning (CSCL) environment. These three groups were chosen for the group meta-cognition study because they were seriously malfunctioning, little co-operation was evident between them and most of their time was spent on non-productive conflict. During the six week period of the study, the three groups were provided with sets of meta-cognitive scaffolds and strategies to facilitate group meta-cognition. The design of the meta-cognitive scaffolds and strategies was informed by a conceptual framework that was derived from the literature in the fields of meta-cognition, cooperative learning, cooperative group meta-cognition, and computer-supported collaborative learning. The study found that providing the students with meta-cognitive scaffolds and strategies resulted in positive changes in the students’ cooperative work and increased levels of knowledge-building activity. The students formed a ‘collective cognitive responsibility’ for their group work and developed an understanding of how to contribute effectively to the knowledge building progress of the group.

The central role of meta-cognition in learning and problem solving, in general and in chemistry in specific was studied by Santiago (2008). The first research question addressed in this work originates from this observation: Is it possible to reliably assess meta-cognition use in chemistry problem solving? This study presents the development, validation, and application of a multi-method instrument for the
assessment of meta-cognition use in chemistry problem solving. This multi-method is composed of two independent methods used at different times in relation to the task performance: (1) the prospective Meta-cognitive Activities Inventory, MCA-I; and (2) the concurrent Interactive Multi-Media Exercises software package, IMMEX. This work also includes the design, development, and validation of the MCA-I; evidence is discussed that supports its robustness, reliability and validity. Even though IMMEX is well-developed, its utilization as a meta-cognition assessment tool is novel and explained within this work. Among the benefits of utilizing IMMEX are: the automation of concurrent evidence collection and analysis which allows for the participation of large cohorts, the elimination of subjective assessments, and the collection of data in the absence of observers which presumably favors a more realistic deployment of skills by the participants. The independent instruments produced convergent results and the multi-method designed was proven to be reliable, robust and valid for the intended purpose. The second guiding question refers to the development of meta-cognition: Can regulatory meta-cognition use be enhanced by learning environments? Two interventions were utilized to explore this inquiry: a Collaborative Meta-cognitive Intervention and a Cooperative Problem-Based Laboratory Project. The former was designed and developed as part of this study; the latter is part of the curriculum of a two semester cooperative General Chemistry Laboratory course. Both interventions rely on two main axes to promote meta-cognition development: intense social interaction and induced reflection. In the first case, it is through small group collaborative work and guided and peer prompting; in the second one through cooperation and inquiry in the laboratory. The effect of both interventions was investigated using pre and posttest, control and treatment type experiments. The choice of assessment was the multi-method developed in the first part of this same study. Despite the differences between the interventions (length, nature of prompting, and relation to chemistry contents) both learning environments succeeded in enhancing the awareness and use of meta-cognition in chemistry problem solving. Findings support the assertion that the mechanisms that define the learning environments under study—social interaction and reflection—promote the enhancement of meta-cognitive skills. A significant corollary of this research is that it offers evidence of the laboratory as a learning environment where students can acquire high order thinking skills and develop content knowledge and understanding.
Gregory, David & Samson (2008) examined the development and evaluation of science students’ meta-cognition, learning processes and self-efficacy are important for improving science education. They report on the development of an empirical self-report instrument for providing a measure of students’ meta-cognition, self-efficacy and constructivist science learning processes. Research studies resulted in a final version of the Self-Efficacy and Meta-cognition Learning Inventory—Science(SEMLI-S) consisting of 30 items that can be used for either analyzing and focusing on any or all of its dimensions or for assigning scores to individuals that enable comparison between them in relation to their meta-cognitive science learning orientations.

Cubukcu (2008) studied the relationship of using meta-cognition strategies and students learning in schools. Research studies strongly support the fact that the using meta-cognition strategies impacts directly on the ability of students' learning in school. The purpose of this study was enhancing vocabulary development and reading comprehension through meta-cognitive strategies. The results of the study indicated that the impact of the meta-cognitive strategy training is important in developing vocabulary and bettering reading comprehension skills. The results of the present study have confirmed that reading comprehension could be developed through systematic instruction in meta-cognitive language learning strategies. Systematic explicit instruction about the concept of meta-cognition and learning strategies helped students of the experimental group to better comprehend this new approach and how to apply it to different learning tasks on reading.

Coutinho (2007) examined the relationship between mastery goals, performance goals, meta-cognition, and academic success. Research focused on how students learn with a view to improve learning tactics for students and encourage effective teaching practices by teachers. This study examined the relationship between three variables – achievement goal orientation that orient students towards a focus on mastering information or performing well, meta-cognition which is the learner’s monitoring of how well he or she is learning, and academic success which is reflected in class grades accumulated over the college tenure. These variables have been studied with elementary and secondary school students but not college students. College students are a different group from elementary and secondary students and may have different learning patterns that are based on their goals to finish college or
acquire skills for a job. The learning variables used in this study have not been studied in this combination. Regression analyses revealed a partial mediation effect in the relationship between mastery goals and academic performance. Performance goals were unrelated to academic performance. This study supported research findings suggesting that students with mastery goals reap the rewards of academic success.

The goal of Jacobs' study (2004) was to investigate the presence and growth of kindergarten children’s meta-cognition as they engaged in the writing process. The study was conducted in an environment that surrounded children with books, language, and print. Twice a month the teacher/researcher interviewed the children as they finished writing, asking questions designed to help them reflect on their thinking and strategies they used in their writing. Anecdotal records, observations, and individual writing folders were used to complete a checklist of writing strategies for each child. Interviews with the children confirmed that they were exhibiting and showing growth in their meta-cognition. They were able to provide appropriate answers to questions that required them to talk about their thinking and identify strategies that helped them in their writing.

Hollingworth & McLoughlin (2000) found a powerful relationship between the development of meta-cognitive skills in science and learning. Results of this research showed that students' meta-cognitive skills can be developed significantly by taking a proactive approach and by designing an environment specifically for problem solving and meta-cognition. They also emphasized that meta-cognition can be developed in contexts that engage students in self-monitoring their own problem solving approaches, in scenarios where they can ultimately use that knowledge. This requires creating real life anchors for development of problem solving skills and enabling students to explore, test and review their own strategies.

Chisholm (1999) examined how gender role identification, meta-cognition and critical thinking contributed to the achievement related behaviors of a group of rural early adolescents. Achievement was measured in terms of overall grade point averages, while achievement orientations were also examined by investigating me, educational, and career aspirations. The results of this study suggest that there has been a break down between the nontraditional and the traditional male and female gender roles. These females had slightly higher achievement levels in terms of overall
grade point averages. Further examination revealed that the contrast in males and females grades may be partially explained by the emergence in achievement orientations. There was a significant correlation between meta-cognitive and critical thinking skills when compared to students' grades. A regression analysis was performed which examined the contribution of gender, meta-cognition, and critical thinking to overall achievement levels. It was determined that when the influence of the other factors was controlled for, only critical thinking came out as a significant predictor of achievement.

Wiles (1997) investigated the influence and effectiveness of a meta-cognitive strategies program for adult-upgrading students on their perceived control and perceived success in the academic setting, as well as performance on a reading comprehension or math assessment. As evidenced in the research pertaining to children, academic success has been closely associated with meta-cognitive awareness, particularly related to perceived control and perceived success. These two areas are considered integral components of attribution, meta-cognition and adult learning. For educators to be able to assist adults returning to school, it is important to ascertain if meta-cognitive training indeed changes perceived control and success as well as performance on academic tasks. This research used a pre post experimental design. The experimental group was composed of four intact English Cognitive strategies classes as well as two Math Cognitive Strategies classes. Two intact English classes and one intact Math class made up the control group. No significant group differences were observed for perception of control. As well, no significant differences were noted between are experimental and control groups for the math area in any of the variables measured. Educational and research implications are discussed as well as limitations of the study.

2.2.4 METACOGNITION IN CHEMISTRY LEARNING

It has been claimed that meta-cognition characteristics make its role in chemistry learning fundamental to achieve deeper and fruitful understanding (Rickey & Stacy, 2000). Despite the consensus about its relevance (Francisco & Nicoll, 1998; Rickey & Stacy, 2000; Schraw et al., 2005), little work has been done to study meta-cognition in the specific context of chemistry learning. Tsai (Tsai, 2001) suggested the potential of using internet-based environments to facilitate practicing meta-
cognitive skills in chemistry courses, as example this author cited monitoring and reviewing of navigating and learning paths, and sorting out and classifying of information. In a study of the multiple teaching methods in a General Chemistry course, Francisco and Trautmann (Francisco & Nicoll, 1998) utilized three interactive approaches (cooperative learning, concept maps, and class discussion) in addition to traditional lecture. Using a survey, they found that students reported higher levels of engagement with the non-traditional teaching styles. Similarly, they found that students assigned different functions to the different formats employed. From these self-reports on level of engagement and functions served, the authors concluded that “multiple modes of learning foster meta-cognitive (sic) skills12 necessary for mastering General Chemistry”. However, the nature of this study is very speculative as it does not present data corresponding to the actual effect on meta-cognition use or course achievement. Often, it is assumed that meta-cognition occurs if that is the intention of the instructor or if an argument for the observed positive outcomes is necessary. In some instances, this amounts to conveniently equating the intended and implemented curricula with the learned curriculum rather than actually assessing meta-cognition.

Schraw et al. (2005) introduced the “Interactive Compensatory Model of Learning” (ICML) as a tool to improve chemistry teaching. This speculative model is based on empirical data and has five components: cognitive abilities, organized knowledge, learning strategies, meta-cognition abilities, and motivational beliefs. The purpose of their paper is to facilitate teachers’ understanding of how to develop more efficient learning environments and includes suggestions as how to improve the previously mentioned components. These suggestions are general in nature and could be implemented in different science education contexts. Another relevant aspect of the scarce literature in the field is that references to the use or the importance of meta-cognition in chemistry are not domain specific (Schraw et al., 2006). An exception in this regard is the instantiation of the MORE Thinking Frame in the chemistry laboratory (Rickey & Stacy, 2000). MORE (Model-Observe-Reflect-Explain) is believed to promote chemistry specific meta-cognitive skills, for example, establishing the connections between macroscopic and molecular level explanations (D. Rickey & Stacy, 2000), and in general is it assumed that it explicitly encourages reflection and meta-cognition (Tien, Teichert, & Rickey, 2007). In an experimental
study (D. Rickey, 1999), participants in a chemistry laboratory course designed around the MORE thinking frame showed “significantly enhanced meta-cognitive abilities, understanding of fundamental chemistry ideas, and abilities to solve examination problems”. The source of evidence supporting the effect on meta-cognition was essentially qualitative in nature: including analysis of video and audio recordings of student working in the laboratory and on transfer problems, and interviews that explored students’ perception of laboratory activities.

Meta-cognitive development in a chemistry related course has been described by Case and collaborators (Case, Gunstone, & Lewis, 2001). In their work, the authors modified a second year chemical engineering course (Materials and Energy Balances) to explicitly develop meta-cognition by “posing questions to the students”, “getting them to try problems on their own”, “discuss issues with their classmates”, “report back to the class”, and “ask questions”. In addition to these activities, students were required to keep a journal with weekly tasks aimed at encouraging their reflection about learning, progress and the course format. Sources of data were the journal entries of 90 students and five individual interviews. Case’s findings suggest that meta-cognition development occurred; however, the study does not provide information about the effect on achievement or understanding of material. Another aspect that should be considered is the fact that in the course program, during instruction, and through the assignments, the instructor’s expectations to develop meta-cognition were explicitly shared. This may be a source of bias for the students who are explicitly assessed on the fulfillment of the expectations of the instructor. In addition, researcher-bias questions may arise from this study design.

Utilizing a case study approach, Davidowitz & Rollnick (2003) investigated the effectiveness of using a metaphor, “The Competency Tripod”, and flow diagrams in enhancing chemistry students’ meta-cognition use. Students were presented with an analogy: The legs of a laboratory tripod are thought of as declarative knowledge, communicative competency, and procedural knowledge; the ring holding the legs together is the link that can be made between the three components or the coherence unifying concept; and the wire gauze represents other factors like the human interaction and time management. The Competency Tripod was assumed to enable meta-cognition by making students reflect about their learning process in a university second year chemistry laboratory. Implementation took place as part of their post
laboratory questions when students were prompted to use the model to reflect about the lab experience. As part of the pre laboratory, students were required to summarize the experiment in a flow diagram that had to be approved by the instructor before the student was allowed to proceed on to laboratory work. The authors collected multiple data from several sources but their analysis revolved around interviews with a stratified sample consisting of four students. Due to the nature of the theoretical framework chosen for the study, the analysis is entirely interpretative and the authors’ interpretation was that meta-cognition enhancement had occurred. However, this conclusion seems more a consequence of the ubiquitous belief that meta-cognition development simply happens if it is the instructor’s objective. In the authors’ own words: “It is not possible to establish directly if the Competency Tripod model was responsible for enabling meta-cognition in students but like dropping a pebble into a pond, its introduction certainly provided ripples which could be identified as meta-cognition” (Davidowitz & Rollnick, 2003).

2.3 ASSESSMENT

2.3.1 Concept and Goals

The purpose of the test provides an operational definition of the proposed assessment and guides decisions related to design activities (Downing, 2006). Volkwein (2003) gives a more specific definition [of assessment]: ‘Student outcomes assessment is the act of assembling and analyzing both qualitative and quantitative teaching and learning outcomes evidence in order to examine their congruence with an institution’s stated purposes and educational objectives’. “Effective assessment begins with clear, overall institutional goals and values”. “Assessment of basic skills needs to include exit proficiencies and tracking of students’ transition and performance in college-level courses. Such assessment should include comparisons between performances in college-level courses of students who transitioned from basic skills versus students who did not need remedial education (cognitive assessment).

Assessment as evaluation or appraisal is about making a judgment, identifying the strengths and weaknesses, the good and the bad, and the right and the wrong in some cases. It is more than simply giving marks or grades, although that may well be a part of it. And because it involves making a judgment it will almost inevitably
include an element of subjectivity by the assessor. However, we should strive to make assessment as objective, fair and transparent as possible. Assessment plays a crucial role in the education process: it determines much of the work students undertake (possibly all in the case of the most strategic student), affects their approach to learning and, it can be argued, is an indication of which aspects of the course are valued most highly (Rust, 2002).

In education, tests are often used to make inferences about the student’s (1) mastery of content in curriculum domains, (2) abilities in cognitive domains, and (3) potential in a future criterion setting (Millman & Greene, 1993).

An example of an educational assessment with a purpose targeted in the cognitive domain is Project STAR (Student Task Assessments and Rubrics). The assessment is a series of performance tasks developed to identify gifted students in underrepresented populations or students not identified by traditional assessments (VanTassel-Baska, Jonson, & Avery, 2002). Students’ verbal and nonverbal skills are assessed using an open-ended format. The verbal section includes problem solving, persuasive writing, analogies, verbal relationships, letter puzzles, and verbal reasoning. The nonverbal section includes arithmetic problem solving, number concepts, logic, proportional reasoning, patterns, number theory, spatial reasoning/visualization, spatial patterning, geometry, and transformations. Thus the tasks focus on skills typically associated with the cognitive domain.

Classroom-based assessments may be of two broad types: selected-response and constructed-response formats. Selected-response formats provide response items for students to choose from (such as multiple-choice, true-false, and matching items). Constructed response formats, on the other hand, ask students to develop a response, create a product, or conduct a demonstration (Feuer & Fulton, 1993; Frisby, 2001; Herman, Aschbacher, & Winters, 1992; McTighe & Ferrara, 1998). These types of assessments allow more than one correct answer to a problem and typically involve higher-order thinking skills.

In education, in a study of performance-based science tasks for the California Assessment Program, nine tasks scored by one rater produced reliable scores (Brennan & Johnson, 1995). Thus in an assessment composed only of performance tasks, 9–10 tasks may be required for examinees’ scores to be reliable.
2.3.2. TYPES OF ASSESSMENT

According to Child (1973/2004), four types of assessments are used depending on different purposes. They are pre-task assessment, formative assessment, diagnostic assessment and summative assessment. First, pre-task assessment aims at discovering the level of knowledge and skills of students before learning. It is often required that teachers, especially inexperienced teachers, devise a pre-task assessment because it helps them have an idea about the level at which to pitch the course. Second, formative assessment, often used interchangeably with the term formative evaluation, is an on-going assessment method to assess the progress made in knowledge and skills during the learning process. In this type of assessment, the teacher’s intention is to optimize feedback by making students aware of their weak and strong points, thereby guiding them to make improvements. Third, diagnostic assessment occurs particularly to pinpoint the cause of difficulties displayed by students who are struggling in a specific area during the learning process and to help them overcome such obstacles. This type of assessment mostly occurs during the formative assessment period. Child (1973; 2004) also stresses that formative and diagnostic assessments are more process oriented approaches where students benefit the most since it provides feedback to remedy students’ shortcomings and to reinforce their strengths. Fourth, summative assessment, also referred to as summative evaluation, normally occurs in the middle or at the end of a course or lesson to grade the students. Summative assessment, a more product-oriented approach, does not usually analyze difficulties nor provide subsequent feedback to the students but serves as useful information for students, teachers and employers to measure the student’s learning results (Child, 1973; 2004: 361-363).

In this research we used different types of assessment but the emphasize was on formative performance assessment.

2.3.3 ASSESSMENT STRATEGIES IN CHEMISTRY

A range of assessment strategies are suggested for chemistry in pre-university. The same strategy can be used both for formative and summative assessment, depending on the purpose of the assessment. Suggested assessment strategies that can be used in the chemistry classroom are discussed in detail in the following section. Teachers are
encouraged to develop their own assessment for Senior Years science based on their students’ learning requirements and the prescribed student learning outcomes.

• **Observation**

  Observation of students is an integral part of the assessment process. It is most effective when focused on skills, concepts, and attitudes. Making brief notes on index cards, self-stick notes, or grids, as well as keeping checklists, help teachers maintain records of continuous progress and achievement.

• **Interviews**

  Interviews allow teachers to assess an individual’s understanding and achievement of the prescribed student learning outcome(s). Interviews provide students with opportunities to model and explain their understandings. Interviews may be both formal and informal. Posing science-related questions during planned interviews enables teachers to focus on individual student skills and attitudes. Students reveal their thinking processes and use of skills when they are questioned about how they solved problems or answered science questions. Using a prepared set of questions ensures that all interviews follow a similar structure. It is important to keep a record of student responses and/or understandings.

• **Group/Peer Assessment**

  Group assessment gives students opportunities to assess how well they work within a group. Peer assessment gives them opportunities to reflect on one another’s work, according to clearly established criteria. During the peer assessment process, students must reflect on their own understanding in order to evaluate the performance of another student.

• **Self-Assessment**

  Self-assessment is vital to all learning and, therefore, integral to the assessment process. Each student should be encouraged to assess her or his own work. Students apply known criteria and expectations to their work and reflect on results to determine their progress toward the mastery of a prescribed learning outcome. Participation in setting self-assessment criteria and expectations helps students to see themselves as scientists and problem solvers. It is important that teachers model the self-assessment process before expecting students to assess themselves.
• **Performance Assessment**

Performance tasks provide students with opportunities to demonstrate their knowledge, thinking processes, and skill development. The tasks require the application of knowledge and skills related to a group of student learning outcomes. Performance-based tests do not test the information that students possess, but the way their understanding of a subject has been deepened, and their ability to apply their learning in a simulated performance. A scoring rubric that includes a scale for the performance of the task helps organize and interpret evidence. Rubrics allow for a continuum of performance levels associated with the task being assessed.

• **Science Journal Entries**

Science journal writing provides opportunities for students to reflect on their learning and to demonstrate their understanding using pictures, labeled drawings, and words. These journal entries can be powerful tools of formative assessment, allowing teachers to gauge a student’s depth of understanding.

• **Rubrics/Checklists**

Rubrics and checklists are tools that identify the criteria upon which student processes, performances, or products will be assessed. They also describe the qualities of work at various levels of proficiency for each criterion. Rubrics and checklists may be developed in collaboration with students.

• **Visual Displays**

When students or student groups prepare visual displays, they are involved in processing information and producing a knowledge framework. The completed poster, concept map, diagram, model, et cetera, is the product with which teachers can determine what their students are thinking.

• **Laboratory Reports**

Laboratory reports allow teachers to gauge the ability of students to observe, record, and interpret experimental results. These tools can aid teachers in determining how well students understand the content.
• **Pencil-and-Paper Tasks**

Quizzes can be used as discrete assessment tools, and tests can be larger assessment experiences. These written tasks may include items such as multiple choice questions, completion of a drawing or labeled diagram, problem solving, or long-answer questions. Ensure that both restricted and extended expository responses are included in these assessment devices.

• **Research Report/Presentation**

Research projects allow students to reach the learning outcomes in individual ways. Assessment should be built into the project at every stage, from planning to researching to presenting the finished product.

• **Interpretation of Media Reports of Science**

Short pieces extracted from newspapers could be used to assess the following: whether pupils understand the scientific content of the piece; whether they can identify and evaluate the possible risks and quality of the evidence presented; whether they can offer well-thought-out reactions to the claims; and, finally, whether they can give their opinion about future action that could be taken by individuals, government, or other bodies (Millar and Osborne, 1998, p. 26).

• **Demonstration of an Understanding of the Major Explanatory**

Stories of Science Questions should seek to examine observable results such as the following: whether students have understood what the particle model of matter is; whether they can give a short account of it; whether they can use it to explain everyday phenomena; and whether they can explain why it is an important idea in science (Millar and Osborne, 1998, p. 26).

• **Asking and Answering Questions Based on Data**

Such questions should assess students’ abilities to represent data in a variety of ways; to formulate and interpret the messages that can be extracted from data; and to detect errors and dishonesty in the way data are presented or selected. The ability to manipulate and interpret data is a core skill that is of value, not only in science, but also in a wide range of other professions and contexts (Millar and Osborne, 1998, p. 26).
• Recognizing the Role of Evidence

At the heart of scientific rationality is a commitment to evidence. Contemporary science confronts the modern citizen with claims that are contested and uncertain. Questions based on historical or contemporary examples can be used to investigate students’ understanding of the role of evidence in resolving competing arguments between differing theoretical accounts (Millar and Osborne, 1998, p. 26

2.3.4 ASSESSMENT AND LEARNING

Assessment tasks must be a part of the regular teaching and learning program (Victorian Curriculum and Assessment Authority, 2010).

There is a widely held belief that assessment drives student learning (Joughin, 2010; Boud, 1990) and that students’ experience of assessment in higher education is that it is central to their learning experience (James, McInnis & Devlin, 2002). If we are to accept the point of view that students place value, artificial though it may be, on their assessments, it offers us, as their instructors, the opportunity to improve the student experience by designing assessments that foster the type of learning we want to encourage (Segers, Martens & Bosscech, 2008).

According to Bloxham & Boyd (2007), the four purposes of assessment are: Certification, student learning, quality assurance and lifelong learning capacity. While the model of assessment proposed does not meet the criteria for certification nor quality assurance, it focuses on the other two components: student learning and lifelong learning capacity.

Learning environments that are knowledge-centered and learner-centered, and that take into account the role of assessment in learning, lay the foundation for a reflective classroom (Bransford et al., 2000).

Many of researches have shown strong links between the implementation of performance assessment and high quality learning (Darling-Hammond & Snyder, 2000; Ridley & Stern 1998; Brown, Collins & Duguid, 1989). The use and implementation of authentic assessment has two significant features; it has the ability to reengage students in the development of content-based knowledge through strengthened links with the outside world; and, it has the capacity to enhance student
learning through the provision of skills such as Meta-cognition, critical thinking and creativity (Darling-Hammond & Snyder, 2000).

Boud et al. (2010) indicated development of professional skills such as problem solving, critical thinking, creativity, meta-cognition, autonomy in learning, and authenticity in learning through innovative forms of assessment (Dochy, Segers & Sluijsmans, 1999); these are the skills that will be necessary for our students to thrive in the 21st century. For most students, assessment is the most important aspect of their coursework (Lamprianou & Athanasou, 2009; James et al., 2002); therefore, we as educators need to use this element of student perception to maximize the learning potential it harbors. If we can engage students through what they value in our courses and ensure that the assessments we assign are authentic and inspire the development of autonomous learning and the skills of critical and lateral thinking, we can maximize the learning.

Five key assessment factors have been identified by Black and Wiliam (1998) in their synthesis of research in assessment that enhance learning: (1) effective feedback, (2) students’ involvement in learning, (3) modifying instruction based on assessment data, (4) recognition of assessment’s influence on students’ motivation and self-esteem, and (5) students’ self-assessment practices. According to Butler and McMunn (2006), quality assessment tasks should have the following essential characteristics: (1) have a clearly defined purpose, (2) aligned with the learning targets embedded in the curriculum, (3) function on a more general level to improve students’ ability to comprehend, and deal with more complex material, (4) allow for more than one right answer, (5) lead to deep understanding, not surface learning, (6) improve students’ cognitive abilities and encourage higher order skills, and (7) improve students’ meta-cognition. Authentic assessment plays an important role in the learners’ developmental process by engaging them in tasks that reflect real-world situations and present “worthy problems or questions of importance, in which students must use knowledge to fashion performances effectively and creatively” (Wiggins, 1993, p. 229).

Teaching and assessment go hand in hand. In the classroom, teaching cannot be truly effective if it is not linked to some form of authentic assessment. Likewise, assessment is useless if it is not based on what has been, or is to be, taught. Although
this may sound obvious, teachers sometimes forget the close relationship between the two.

Kabba (2008) argued that performance-based assessment requires students to demonstrate their learning and understanding by performing an act or a series of acts. This type of assessment is appropriate to use in a project-based, problem-based, or inquiry-based science classroom because it is consistent with the way students learn—by investigating a question or problem using tools and materials (i.e., performing an act). Since students in a project-based classroom learn by producing a product or performing an act, it is only fitting for them to be assessed using methods similar to those used to teach them—thus, aligning assessment with instruction.

According to Lombardis' study (2008), assessment is an integral component of any successful teaching effort. Research has shown that students engage with subject matter based in part on their expectations about how their achievement will be evaluated. Educators who strive to bring authentic learning experiences to their students must devise appropriate and meaningful measures to assess student learning and mastery of concepts at hand. Although some barriers must be overcome, numerous examples point to the opportunities available for effective assessment of authentic learning initiatives. These approaches to assessment are vital to ensuring that models of teaching and learning with technology see their full potential.

Lubezky, Dori & Zoller (2004) indicated that switch from traditional algorithmic lower-order cognitive skills (LOCS) teaching to higher-order cognitive skills (HOCS) learning in science, chemistry, engineering and environmental education, requires an in accord shift in students' assessment. HOCS-promoting assessment is expected to enhance students’ evaluative thinking in terms of system, critical thinking – problem solving and decision making.

Assessment should focus on achievement of authentic learning outcomes. The focus adopted for assessment has a substantial impact on the instructional process and the realization of intended learning outcomes through the channeling it encourages of teacher and student effort. That is, assessment signals to teachers and students what is important in learning. Vice versa, assessment tasks need to relate to and take account of actual teaching and learning processes. Relationships between learning goals, teaching activities, learning processes and assessment procedures can be depicted in
the form of a tetrahedron as in Figure 2.3. In such a system of interrelationships, all four components are in dynamic tension or balance. That is, adjustment of one component requires sympathetic adjustment of the other three. Sympathetic adjustment implies alignment of the underlying rationale or theoretical assumptions of each component. That is, systemic validity inheres in the consistency of the underlying rationales of the four components. In particular, in terms of authenticity, the underlying rationale or theoretical assumptions relating to the authenticity of the learning goals and the assessment procedures not only need to match each other but also those relating to the teaching processes and the nature of learning and achievement.

![Diagram](image)

**Figure 2.3:** The teaching, learning, assessment domain (Newmann & Archbald 1992).

The authentic achievement outcomes are drawn from a cognitive theory of learning and performance. The content of their discussion refers to cognitive processes and generic abilities such as higher-order thinking, embedded in a prior knowledge base. While Newmann and Archbald (1992) have extracted the essence of current cognitive learning theories, nuances in variations of these theories are important. Such different theoretical interpretations, particularly regarding the nature of knowledge and learning, lead to variations in the constructions of authenticity and the implementation of authentic assessment. One major interpretation of authentic achievement and authentic assessment relate to: performance and performance assessment.
According to Kearney & Perkins (2011), assessments' task must be authentic; that is, it must have a direct correlation or relevance to the students’ world outside of the classroom. In designing authentic assessment tasks, we inevitably encourage learning that has applicability outside of the classroom, which makes the learning sustainable.

Sari & Wiyarsi (2011) aimed to determine: (a) whether there is any difference between chemistry learning achievements of students taking the chemistry class with the implementation of Performance Assessment and ones joining the class without the implementation of Performance Assessment if the prior knowledge was statistically controlled, (b) the effectiveness of the implementation of Performance Assessment on the motivation improvement to learn chemistry of high-school (SMA) students in Yogyakarta, and (c) the quality and profile of Performance Assessment of SMA students in DIY in the terms of experimental works, presentations, discussions, and quality of the report. The experiment was conducted in 5 high schools in Yogyakarta, each of them was taken from a high school of each regency including Gunungkidul, Sleman, Bantul, Kulonprogo, and the municipality of Yogyakarta. The high schools selected in the research were determined by using purposive sampling including SMA N 6 Yogyakarta City (65 students), SMA N 1 Prambanan (64 students), SMA N 2 Bantul (72 students), SMA N 2 Wonosari (50 students), and SMA N 2 Wates (64 students). Research instruments used included RPP, motivation questionnaires, questions on chemistry learning achievements, observation sheets, and assessment rubrics of Performance Assessment. Data collection technique was through documentation for data of prior knowledge, questionnaires for motivation data, tests for achievement data, and observations for Performance Assessment data. Data analysis was performed by t test, Anakova test, and descriptive analysis of the percentage. The results showed: (a) there were significant differences on chemistry learning achievement with and without the implementation of Performance Assessment on students of class XI IPA at SMA N 6 Kota Yogyakarta, SMA N 1 Prambanan, SMA N 2 Bantul, SMA N 2 Wonosari, dan SMA N 2 Wates while at SMA Negeri 2 Bantul, there was no significant difference. However the average value of the chemistry learning achievements in the classes with the implementation of Performance Assessment was higher than ones without the application of Performance Assessment. (b) Based on the statistical analysis of the t-test of same
subjects, it showed that there was a significant increase of scores of students’ motivation to learn chemistry in classes with the implementation Performance Assessment at SMA N 6 Yogyakarta, SMA N 1 Prambanan, SMA N 2 Bantul, SMA N 2 Wonosari, and SMA N 2 Wates, while in classes without the application of Performance Assessment, there was no significant increase chemistry learning motivation. (c) the performance profile of the students of class XI IPA in semester 2 at SMA Negeri 6 Yogyakarta showed the mean of students’ performance quality in the categories: (94% very well, 6% good), SMA Negeri 1 Prambanan (51% very well, 49% good), SMA Negeri 2 Bantul (95.4% excellent, 4.6% both), SMA Negeri 2 Wonosari (25% excellent, 55% good, 20% moderate, and SMA Negeri 2 Wates (90.6% excellent, 9.4% good).

The following table summarizes some of the changes in the area of assessment in chemistry in pre-university.

**Table 2.4: Changing Emphases in Assessment of Student Learning in chemistry**

*National Science Education Standards (1996)*

<table>
<thead>
<tr>
<th>LESS EMPHASIS ON</th>
<th>MORE EMPHASIS ON</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessing what is easily measured</td>
<td>Assessing what is most highly valued</td>
</tr>
<tr>
<td>Assessing discrete knowledge</td>
<td>Assessing rich, well-structured knowledge</td>
</tr>
<tr>
<td>Assessing scientific knowledge</td>
<td>Assessing scientific understanding and reasoning</td>
</tr>
<tr>
<td>Assessing to learn what students do not know</td>
<td>Assessing to learn what students do understand</td>
</tr>
<tr>
<td>Assessing only achievement</td>
<td>Assessing achievement and opportunity to learn</td>
</tr>
<tr>
<td>End-of-term assessments by teachers</td>
<td>Students engaged in ongoing assessment of their work and that of others</td>
</tr>
<tr>
<td>Development of external assessment by measurements experts alone</td>
<td>Teachers involved in the development of external assessments</td>
</tr>
</tbody>
</table>
Kearney & Perkins (2010) examined the relationship of authentic assessment and students learning in the classroom. Research conducted in the School of Education at the University of Notre Dame, Australia where 280 undergraduate primary education students were surveyed prior to undertaking ASPAL and after undertaking ASPAL (Authentic Self & Peer Assessment for Learning). The results indicated students that were more engaged, had increased efficacy and felt that they were a part of the educative process, rather than being subjected to it. Research has shown strong links between the implementation of authentic assessment and high quality learning.

Chuang (2009) investigated Taiwanese college EFL students' perspectives on using oral performance assessment in class. The results indicate that test anxiety occurs when students engage in oral tests. Regarding test type, the majority of the students feel comfortable in paired role-playing setting rather than individual interview; their performance however can be somewhat influenced by their partner. Students agree that the oral testing is necessary in English courses, and many students believe that their performance more or less can be affected by test format/task type. The results also reveal that students are willing to take oral performance test instead of traditional written test: they get to realize the importance of being able to use the language for meaningful communication orally in real-life situations. In this study, the researcher sheds light on using oral performance assessment as a necessary and practical way to enhance students' speaking skills and ability.

She et al. (2007) conducted a study to investigate the students’ learning outcomes in Organic Chemistry Laboratory; a performance-based assessment was developed and implemented to a group of 222 chemistry-majored sophomores from a research university in Taiwan. A multi-step organic synthesis experiment was chosen, and the basic organic manipulative skills, such as reflux, distillation, extraction, filtration, and re-crystallization were incorporated in the laboratory activities. According to the mean scores of on-site assessment, the students’ manipulative skills of extraction and filtration should be reinforced. Students paid less attention on laboratory safety, and liked to rinse the apparatus with large amount of acetone, did not set up the equipments in proper position and did not work in the fume hood from our observations. Those indicated that we should more emphasize on the laboratory safety guidelines in the class. From the analysis of paper-and-pencil science-concept
test, students were unable to give reasonable explanations for certain Lab observations, lacking the ability of rationalizations.

Elwood (2006) undertook a study to consider the fundamental relationship between learning and assessment and also introduces views of mind as a further key concept in understanding assessment practice. These concepts are building blocks for fully understanding emerging trends and practice in assessment. For example, the Assessment for Learning movement that promotes different forms of assessment practice tends to focus on the individual student, their learning and how this can be developed through particular assessment practices. This movement has a particular view of learning, the learner and of assessment but the associated view of mind still sees the student as separately analyzable after the learning has taken place.

Andrade & Du (2005) conducted a study to use rubrics for supporting their own learning and academic performance. In focus groups, fourteen undergraduate students discussed the ways in which they used rubrics to plan an approach to an assignment, check their work, and guide or reflect on feedback from others. The students said that using rubrics helped them focus their efforts, produce work of higher quality, earn a better grade, and feel less anxious about an assignment. Their comments also revealed that most of the students tend not to read a rubric in its entirety, and that some may perceive of a rubric as a tool for satisfying a particular teacher’s demands rather than as a representation of the criteria and standards of a discipline. Students’ comments regarding rubric use were consistently positive. They liked the fact that rubrics let them know “what’s expected,” and contrasted it with the “guessing game” they felt they had to play when teachers did not provide a rubric or some sort of guidelines for an assignment. In fact, the most commonly cited purpose of rubrics was to communicate the teacher’s expectations and thereby provide “direction.” Students also noted that rubrics help identify strengths and weaknesses in their work when used to give feedback, and that knowing “what counts” made grades seem fair.

Luca & McMahon (2004) proposed an approach to assessment through student contracts that are designed to both address issues of fairness of assessment as well as promote the planning, monitoring and evaluation integral to enhancing metacognition. By negotiating their assessment in ways that involve feedback that is
internal, parallel and external to students, they have the opportunity to develop clearer understandings of themselves as learners and their own learning processes. The model is supported with online technology to help create an easy and confidential manner in which peer feedback can be collated. Initial findings suggest that students perceive themselves to be meta-cognitive; however, they may not be fully engaged in the processes that underpin this psychological state. It is proposed that online environments that support negotiated assessment expose these processes in ways that may enhance meta-cognitive outcomes, and lead to further research that identifies the nature of such processes and their value in the assessment approach.

Lubezky, Dori & Zoller (2004) examined the switch from traditional algorithmic lower-order cognitive skills (LOCS) teaching to higher-order cognitive skills (HOCS) learning in science, chemistry, engineering and environmental education, requires an in accord shift in students' assessment. HOCS-promoting assessment is expected to enhance students’ evaluative thinking in terms of system, critical thinking – problem solving – decision making, followed by a responsible action, accordingly. The focus of this study was on the development and implementation of HOCS-promoting environmental chemistry related exam items/questions, included in traditional exams in undergraduate chemistry, chemistry teaching methods and chemistry-related environmental education courses. These exams were administered to three groups of undergraduate, pre- and in-service science majors and science chemistry teachers at an engineering-technology and liberal arts university, respectively. The first and the third groups were not formally exposed to chemistry-related environmental issues in their courses. Our findings suggest that our undergraduate students appear to be weak on both, “making connections” and system thinking, with respect to these issues. Yet, pre-post improvement of their HOCS capability was found, particularly for the students who scored low on the pre-test. The implications for future action purposed persistent HOCS-promoting teaching and assessment, within which relevant environmental issues are integrated, have the potential of inducing ‘HOCS learning’ in science, technology, environmental, society (STES)-oriented chemistry teaching.

Klein et al. (1997) examined whether the differences in mean scores among gender and racial/ethnic groups on science performance assessments are comparable to the differences that are typically found among these groups of traditional multiple-
choice tests. To do this, several hands-on science performance assessments and other measures were administered to over 2,000 students in grades five, six, and nine as apart of a field test of California’s statewide testing program. Girls tended to have higher overall mean scores than boys on the performance measures, but boys tended to score higher than girls on certain types of questions within a performance task. In contrast, differences in mean scores among racial/ethnic groups on one type of test (or question) were comparable to the differences among these groups on the other measures studied. Overall, the results suggest that the type of science test used is unlikely to have much effect on gender or racial/ethnic differences in scores.

2.4 AUTHENTIC ASSESSMENT

2.4.1 Concept and Goals

A new approach to evaluation is authentic assessment. This modality connects teaching to realistic and complex situations and contexts. Also called performance assessment, appropriate assessment, alternative assessment, or direct assessment (Olfos & Zulantay, 2007).

According to Aiken (1996), authentic assessment of educational achievement directly measures actual performance in the subject area. It was developed as a result of criticism of multiple-choice tests, which usually only provide a superficial idea of what a student has learned and do not indicate what a student can do with what was acquired.

Wiggins (1993) reported that authentic assessment is a form of assessment in which students are asked to perform real-world tasks that demonstrate meaningful application of essential knowledge and skills. Authentic measures are “engaging and worthy problems or questions of importance, in which students must use knowledge to fashion performances effectively and creatively.

Authentic assessment can provide genuine accountability. All forms of authentic assessment can be summarized numerically, or put on a scale, to make it possible to combine individual results and to meet state and federal requirements for comparable quantitative data (National Center for Fair and Open Testing, 1992).

In general, assessment refers to “any judgment (appraisal, or evaluation) of a student’s work or performance made by a teacher or a competent person, whether for
purposes of improvement or certification” (Sadler, 1987, p. 191). Authentic assessment is one of the many forms of assessment that lecturers can do in their subject, together with performance-based assessment, portfolio assessment, and a host of other assessment types.

In relation to differences between traditional assessment and authentic assessment, below fundamental question will be considered:

**How is authentic assessment different from traditional assessment?**

Mueller (2010) presented the distinction, along a continuum of attributes, between authentic assessment and traditional assessment in this way:

**Table 2.5: Distinction between authentic assessment and traditional assessment**

(Muller 2010)

<table>
<thead>
<tr>
<th>TRADITIONAL</th>
<th>AUTHENTIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Selecting a response</td>
<td>Performing a task</td>
</tr>
<tr>
<td>Contrived</td>
<td>Real-life</td>
</tr>
<tr>
<td>Recall-Recognition</td>
<td>Construction/Application</td>
</tr>
<tr>
<td>Teacher-Structure</td>
<td>Student-Structure</td>
</tr>
<tr>
<td>Indirect evidence</td>
<td>Direct evidence</td>
</tr>
</tbody>
</table>

That is, traditional assessment tends to fall towards the left end of the continuum while authentic assessment to the right. By recognizing these attributes, lecturers can think about assessment tasks that they can use that are more about ‘doing’ than simply ‘knowing’.

Traditional assessment commonly refers to ‘paper-and-pencil’ tests. They take a variety of forms but commonly they are true-false or simple multiple-choice type tests that tend to focus on testing lower order thinking skills. These types of tests often only require memorized concepts or facts. By using such forms of assessment knowledge can be de-contextualized as bits of information stored in memory. Learners may find it difficult to reproduce that knowledge and apply it in different decision-making contexts.
Calma (2010) argued that another way to distinguish authentic assessment from traditional assessment is by examining the tasks related to each. In traditional assessment, often students are asked to recall what they have memorized from texts. In authentic assessment, because the tasks are derived from real settings, the tasks require students to act as if they are the key players involved.

Shepard (2000) stated traditionally in higher education, Graff says, “our assumption has been that most students will not learn what we teach them, that given human nature this is to be expected, and that ultimately this is not our problem” (2009 p.160). Traditional forms of assessment focus on scientific principles that seek to be objective and are seen to be separate from learning and built on uniformity and fairness.

According to Bagnato's study (2007), authentic assessment provides an alternative approach to gathering performance information and designing developmentally-appropriate curriculum for students. Conventional tests and assessment have been shown to lead to the mis-measurement of students through the use of norm-referenced testing practices

Hart (1994) studied in a problem-centered curriculum; new assessment techniques that align with classroom activities and seek to assess students' understanding of science concepts are required. Since traditional tests often focus only on the answer or the use of a suitable algorithm to reach the answer, authentic assessment techniques need to be employed to provide a broader range of measures. In this article, the term authentic assessment is used to describe assessment of this type: assessment that involves students in tasks that are worthwhile, significant, and meaningful and that resemble learning activities. Such assessment activities also encourage risk taking and provide the opportunity to demonstrate the application of knowledge in unfamiliar settings. Other terms that are often used for authentic assessment include alternative, performance-based, or outcome-based assessment.

To capture the multi-faceted aspects of the problem-solving approach, a broad range of assessment techniques needs to be employed, such as structured interviews, concrete models, group problem solving, creative projects, and portfolio evaluations, as well as paper-and-pencil tests (Clarke, 1992; Goldin, 1992; Lampert, 2001). Authentic assessment includes the use of open-ended problems, scoring rubrics,
student self-assessment, science portfolios, and student journals. When a wide variety of these assessment techniques is used, teachers gain insight into the student's thinking and understanding of science and students learn to describe their own problem-solving strategies (Maher, Davis, & Alston, 1992; Schoen, Cebulla, Finn & Fi, 2003).

2.5 Performance assessment

Performance assessment permeates much of our world. Children encounter performance assessment as they move through school grades, in particular by the use of writing assessments, which now are often a part of final graduation requirements and college entry. Assessing performance is not a recent phenomenon; examples of performance assessment are available from early Chinese culture. In writing about performance assessment, Fitzpatrick and Morrison (1971) stated that “The performance is a sequence of responses aimed at modifying the environment in specified ways” (p. 239). This “sequence of responses” includes the behaviors of an examinee—what an examinee says, does, or creates (Messick, 1994; Mislevy, Steinberg, & Almond, 2002; Mislevy, Wilson, Ercikan, & Chudowsky, 2003).

ACS (2011) stated performance assessment includes research work, internships, simulations, and projects. In such instances, students are placed in an environment similar to that which they will be functioning after graduation and are asked to perform the task or skill for which they are being educated. Performance type assessments are based upon cognition (knowledge); however, it also includes an action component in which that knowledge is put into practical use.

The global economy of the 21st century requires a different set of knowledge and skills than workers needed in the past. These labor-market demands are driving education reforms, and helping students develop these skills necessitates a shift in teaching and assessment. Today, students must develop the ability to access, inquire into, think critically about, and carefully analyze information (Business-Higher Education Forum, 2007; Partnership for 21st Century Skills, 2005). Performance assessments address these societal demands by providing complementary information to traditional test data; it measures what students are able to do with the content they learn in school, not only whether they have learned the curriculum content. In other words, performance assessments can provide insight into how well students are able
to apply knowledge to solve new problems. Performance tasks are hands-on activities that require students to demonstrate their ability to perform certain actions. This category of assessment covers an extremely wide range of behaviors, including designing products or experiments, gathering information, tabulating and analyzing data, interpreting results, and preparing reports or presentations.

Herrington and Herrington (1998) noted that the terms “performance assessment” and “authentic assessment” also tend to be used interchangeably. While performance assessment and PBA are simply shortened versions of performance-based assessment, there is a notable difference between authentic assessment and performance-based assessment. Gulikers, Bastiaens, & Kirschner (2004) cited previous literature to explain the difference between performance assessment and authentic assessment.

Some see authentic assessment as a synonym to performance assessment, while others argue that authentic assessment puts a special emphasis on the realistic value of the task and the context. Some researchers point out that the crucial difference between performance assessment and authentic assessment is the degree of fidelity of the task and the conditions under which the performance would normally occur. Authentic assessment focuses on high fidelity, whereas this is not as important an issue in performance assessment. These distinctions between performance and authentic assessment indicate that every authentic assessment is performance assessment, but not vice versa (Herrington & Herrington, 1998).

Performance assessment can be defined as the direct, systemic observation of an actual learner performance and the rating of that performance according to previously established performance criteria. (Bachman, 2002; Bonk & Ockey, 2003; McNamara, 1996; Lumley, 2002).

O’Malley & Valdez-Pierce (1996, p. 5) stated that performance-based assessment includes the following characteristics:

1. Students make a constructed response.
2. They engage in higher-order thinking, with open-ended tasks.
3. Tasks are meaningful, engaging, and authentic.
4. Both process and product are assessed.
5. The depth of a student’s mastery is emphasized over breadth.

In a performance assessment, examinees demonstrate their knowledge and skills by engaging in a process or constructing a product. More broadly, a performance assessment is a system composed of (1) a purpose for the assessment, (2) tasks (or prompts) that elicit the performance, (3) a response demand that focuses the examinee’s performance, and (4) systematic methods for rating performances (Ruiz-Primo & Shavelson, 1996; Shavelson, Solano-Flores, & Ruiz-Primo, 1998; Stiggins, 1987a).

Performance assessment can be defined as a method of evaluating students’ knowledge, concepts, or skills by requiring them to perform a task designed to emulate real-life contexts or conditions in which students must apply the specific knowledge, concepts, or skills (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; U.S. Department of Education, 1993).

Not only does performance assessment allow students to demonstrate their abilities in a more genuine context than is required by other types of assessment, performance assessment has other advantages over the traditional assessments that are more commonly used in schools today. Students are able to recognize real-life connections with performance assessments. Additionally, students are generally more motivated by high-quality performance assessments, which have the capacity to measure higher-order thinking skills and other abilities needed to achieve success in the contemporary workplace.

Baker & O’Neil (1994, p. 15) described performance-based assessment as incorporating higher-order thinking and authenticity of purpose and elements of ‘real world’ performance:

... Complex learning, higher order thinking, stimulation of a wide variety of active responses of students, tasks requiring multiple steps, and significant commitments of student time and effort.
Performance assessments are as a type of assessment that “requires students to actively accomplish complex and significant tasks, while bringing to bear prior knowledge, recent learning and relevant skills to solve realistic or authentic problems.” Performance assessments provide an indication of what students “can do with what they know,” compared with traditional assessments, which often only tap “what students know” (Herman, Ascbacher & Winters, 1992).

McBrein & Brandt (1997) found that definitions of performance assessment vary, although all refer to the notion of authentic tasks such as activities, exercises, or problems that require students to show what they can do. Defined by the U.S. Congressional Office of Technology Assessment (OTA) (1992, p. 5), performance assessments are “testing methods that require students to create an answer or product that demonstrates their knowledge and skills and take many forms including: conducting experiments, writing extended essays, and doing mathematical computations.” Further, according to the OTA, performance assessment is best understood as a continuum of assessment formats ranging from the simplest student-constructed responses to comprehensive demonstrations or collections of work over time. According to Robert Lukhele and his colleagues (1993, p. 2), constructed responses may be understood as “any question format that requires the test taker to produce a response in any way other than selecting from a list of alternative answers. Constructed-response tasks may be as simple as a short-answer question, or adding an arrow to a diagram. They may require the test taker to organize and write an essay, solve a multi-step mathematics problem, draw a diagram or graph, or write an explanation of a procedure.” Whatever the format, common features of performance assessments involve students’ construction rather than selection of a response, direct observation of student behavior on tasks resembling those commonly required for functioning in the scientific field, and illumination of students’ learning and thinking processes. Finally, the OTA explains that performance assessments must measure what is taught in the curriculum and that there are two terms that represent core or essential attributes of a performance assessment:

1. **Performance**: A student’s active generation of a response is observable either directly or indirectly via a permanent product.
2. **Authentic**: The nature of the task and context in which the assessment occurs is relevant and represents "real world" problems or issues.

The global economy of the 21st century requires a different set of knowledge and skills than workers needed in the past. These labor-market demands are driving education reforms, and helping students develop these skills necessitates a shift in teaching and assessment. Today, students must develop the ability to access, inquire into, think critically about, and carefully analyze information (Business-Higher Education Forum, 2007; Partnership for 21st Century Skills, 2005). Performance assessments address these societal demands by providing complementary information to traditional test data; it measures what students are able to *do* with the content they learn in school, not only whether they have learned the curriculum content. In other words, performance assessments can provide insight into how well students are able to apply knowledge to solve new problems.

Calfee & Perfumo (1993), and Bedir, Polat & Sakacı (2009) studied that using the performance assessment in one lesson, improves students' interests, motivation and confidence towards learning, and eventually serves students to become lifelong learners.

During much of the 20th century, testing programs used multiple-choice items in their assessments. Today testing programs in education often use both multiple-choice and performance assessment. Incorporating both performance tasks and multiple-choice items into an assessment may be due in part to assessment practitioners’ conceptualizing multiple-choice items and performance assessment as part of an item format continuum (Fortune & Cromack, 1995; Gronlund, 2006; Johanson & Motlomelo, 1998; Messick, 1996; Snow, 1993). Table 2.6 shows a continuum of item formats and examples of qualities to consider when determining the appropriate form of assessment for a given testing situation. For example, if authenticity and cognitive complexity are key qualities for an assessment that you are designing, then one of the performance assessment item formats will better address these qualities.
Table 2.6: Item-format continuum and qualities of the formats. Fortune and Cromack (1995); Gronlund (2006); Johanson and Motlomelo (1998); Snow (1993).

<table>
<thead>
<tr>
<th>Performance assessment formats</th>
<th>selected-Response formats</th>
</tr>
</thead>
<tbody>
<tr>
<td>More authentic</td>
<td>Less authentic</td>
</tr>
<tr>
<td>More cognitive complex</td>
<td>Less cognitive complex</td>
</tr>
<tr>
<td>More in-depth Content coverage</td>
<td>More breadth In content coverage</td>
</tr>
<tr>
<td>Examinee constructed Response structure</td>
<td>Test developer Response structure</td>
</tr>
<tr>
<td>More expensive</td>
<td>Less expensive</td>
</tr>
</tbody>
</table>

By combining performance assessment with multiple-choice items, testing programs benefit from the strengths of each form of assessment (Lane& Stone, 2006).

Performance-based assessment is suitable for assessing nearly all types of science learning. It is particularly useful for assessing science-process skills, such as the ability to formulate a hypothesis; think critically; solve problems; design and conduct investigations; use instruments; collect, analyze, and interpret data; and document and present findings from research projects. Performance-based assessment allows the student to construct his or her own answers as opposed to choosing from a group of answers. Both student and teacher are made aware of the skills and knowledge to be learned as well as the criteria for judging performance. Performance-based assessments “can be a learning experience in themselves. They can actually motivate students to learn more about the subject matter” (Doane, Rice & Zachos 2006). Science is also a performance-based enterprise, and we should expect our students to demonstrate scientific knowledge and understanding through performance.

Cognitive research Gardner (1993) indicated that most learning goes on within an active, rather than a passive, context and “that children construct knowledge from their actions on the environment”. Performance-based assessment is therefore suitable for assessing nearly all types of learning because it allows students to demonstrate
their competency in ways compatible with their learning experience. In project- or inquiry-based teaching, this method of assessment is particularly applicable for assessing science-process skills, such as the ability to ask questions, formulate hypotheses, design experiments, analyze and interpret data, use instruments, and present findings (Wadsworth 1989, p. 156).

Pierce (2002) conducted a study on performance-based assessment (PBA), which uses a constructed-response format, has as its primary purpose the improvement of learning. Performance based assessment links assessment to instruction through the use of meaningful and engaging tasks. The results showed using performance based assessment increase learning function.

Authentic assessment is a type of PBA, promotes application of knowledge and skills in situations that closely resemble those of the real world (Frisby, 2001; McTighe & Ferrara, 1998; Wiggins, 1998). Authentic assessments are potentially more motivating than other types because they engage students in realistic uses of science concepts. Authentic assessment and other types of PBA can be used in the service of education to promote transfer or generalizability of learning from facts and procedures to applications in meaningful contexts. A large range and number of tasks are needed over time, however, to ensure the generalizability of PBAs.

Can performance-based assessments be used to monitor and support the learning of students? A number of factors make PBAs more appropriate for student than traditional testing formats (Frisby, 2001; Hamayan & Damico, 1991; O’Malley & Pierce, 1996). Well-constructed performance tasks are more likely than traditional types of assessment to do the following:

- provide comprehensible input to students
- Use meaningful, naturalistic context embedded tasks through hands-on or collaborative activities
- show what students know and can do through a variety of assessment tasks
- allow for flexibility in meeting individual needs
- use criterion-referenced assessment for judging student work
- provide feedback to students on strengths and weaknesses
- generate descriptive information that can guide instruction
• provide information for teaching and learning that results in improved student performance

Further, PBAs have the potential to provide in-depth information about a student’s ability to integrate knowledge for specific curriculum objectives or standards.

2.5.1 HISTORY OF PERFORMANCE ASSESSMENT

The history of performance assessment in education in North America includes oral recitations and can be traced to the beginnings of college education in North America (Rudolph, 1990). Oral exams were used in Massachusetts as early as 1709 (Morris, 1961). Horace Mann replaced the oral exam in the Boston public schools with an essay exam in 1845. Prior to the Civil War, testing requirements for candidates for accounting positions at the Treasury Department included writing a business letter (Hale, 1982). In the early 1900s, to assign military personnel to specialized tasks and rate officers for appointment and promotion, John B. Watson and other psychologists developed a series of tests based on oral questions (Hale, 1982). In these oral examinations, military personnel answered questions based on various trades, photographs, and sample tasks.

Early in the 20th century essay tests were challenged, in part due to studies that showed the unreliability of the scoring of essay exams (Starch & Elliot, 1912, 1913). Frederick Kelly’s introduction of multiple-choice items in 1915 facilitated the development of standardized, norm-referenced tests, further challenging the use of essays (Madaus & O’Dwyer, 1999; Office of Technology Assessment, 1992).

The 1980s and 1990s saw a resurgence in the use of performance assessment in education. For example, in the 1980s, state testing programs replaced multiple-choice measures of language usage, grammar, and mechanics with writing samples (Stiggins, 1987b).

Continuing the trend in the early 1990s, state departments of education and school districts incorporated mathematics and writing portfolios into their testing programs (Gong & Reidy, 1996; Koretz, Stecher, Klein, & McCaffrey, 1994; LeMahieu et al., 1995).
In licensure, the 1980s marked a decline in performance testing due to the expense and the high correlation between scores on performance and written tests (Knapp & Knapp, 1995). However, return to the use of performance assessment in the form of clinical examinations occurred due to “a mistrust of paper-and-pencil or multiple-choice tests, a need to see a candidate work with people, and a need to see the candidate perform in a work setting integrating the physical and cognitive skill areas” (Fortune & Cromack, 1995, p. 155). Today performance assessment plays an important role in examinees’ lives from their entry into public school, through their matriculation into a university, and into their professional lives.

2.5.2 CHARACTERISTICS AND BENEFITS OF PERFORMANCE TESTS

A number of the authors in the literature describe what they believe to be important characteristics of performance tests. Interestingly and perhaps naturally, the aspects they choose to describe differ widely. For instance, in general education literature, Quellmalz (1991) lists the following:

1. Significance [cognitive components; meta-cognitive components (e.g., planning, self-monitoring, evaluation, and reflection); and dispositional components (e.g., attitudes, perseverance flexibility)]
2. Fidelity [Contextual components; complete performance assessment (full range of process)]
3. Generalizability [of a task to a range of other tasks]
4. Development appropriateness in terms of getting representative sample of achievement milestones
5. Accessibility of criteria for educators, students, parents, community
6. Utility

O’Neil & Abedi (1996) stated that there are many assumed advantages of alternative assessments, for example, that such assessments should result in more effort expended and perhaps less anxiety. Further, such assessments should engage students in higher level thinking or meta-cognitive skills.

In education circles, Wiggins (1989) argued that authentic educational measurement (including performance assessment):
1. Involve collaborative elements
2. Is contextualized and complex
3. Measures real-world tasks
4. Establishes standards that are clear to students and authentic


Khattri, Reeve, and Kane (1998, pp. 26-27) stated that “initiators and supporters of performance assessments claim that the assessments influence and inform instruction, because they:

1. Help teachers and other educators conduct a comprehensive evaluation of students’ achievement, including students’ strengths and weaknesses.
2. Support instruction and curriculum aimed at teaching for understanding by providing good pedagogical templates.
3. Help teachers and other educators better assess students’ understanding of procedural knowledge, which is not so easily judged through traditional assessment methods.”

2.5.3 TRADITIONAL ASSESSMENT VS. PERFORMANCE ASSESSMENT

The increase in popularity of performance assessments during the late 1980s and 1990s came about in part because of dissatisfaction with traditional, multiple-choice tests (Kahl, 2008). By the end of the 20th century, performance assessment had moved “from a trendy innovation to an accepted element of good teaching and learning” (Brandt, 1998). With the increase in standardized testing after the No Child Left Behind Act of 2001 was signed into law, educators took a renewed interest in different types of alternative assessments, including performance assessment. As seen in Table 2.7, performance assessment has a number of advantages over traditional assessment for evaluating individual students. Most notably, performance assessment has the capacity to assess higher-order thinking and is more student-centered than traditional assessment.
Table 2.7: Attributes of Traditional Assessments and Performance Assessments

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Traditional assessment</th>
<th>Performance assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessment Activity</td>
<td>Selecting a response</td>
<td>Performing a task</td>
</tr>
<tr>
<td>Nature of Activity</td>
<td>Contrived activity</td>
<td>Activity emulates real life</td>
</tr>
<tr>
<td>Cognitive Level</td>
<td>Knowledge/Comprehension</td>
<td>Application/Analysis/Synthesis</td>
</tr>
<tr>
<td>Development of solution</td>
<td>Teacher-structured</td>
<td>Student-structured</td>
</tr>
<tr>
<td>Objectivity of scoring</td>
<td>Easily achieved</td>
<td>Difficult to achieve</td>
</tr>
<tr>
<td>Evidence of Mastery</td>
<td>Indirect evidence</td>
<td>Direct evidence</td>
</tr>
</tbody>
</table>

Advocates of performance assessment also emphasize that it is more in line with instruction than traditional assessment (Palm, 2008). While Popham (2001), and other experts (Haladyna, Nolen, & Haas, 1991; Mehrens, 1991) agree that teaching to the test—Popham refers to the practice as item-teaching—is highly unethical in preparation for traditional assessments, teaching to the test is actually encouraged when it comes to performance assessments (Mueller, 2008). With performance assessment, students have access to scoring rubrics in advance so they will know exactly how their performance (e.g., oral or written response, presentation and journal) will be evaluated. Teachers should also allow their students to preview examples of high-quality and poor performance products to use as models, provided the product cannot be mimicked.

In November 2008, the National Academy of Education (NAE) hosted the Education Policy in Transition Public Forum in Washington, D.C. The purpose of the forum was to facilitate a discussion between educational researchers, policy leaders, and advisers to Congress and the new administration on the most critical issues in education policy. One of the main concerns mentioned by the panel on Standards, Accountability, and Equity in American Education was that:

Accountability tests over-represent what is relatively easy to measure (such as basic skills) and under-represent highly valued reasoning skills such as problem solving. Because there are consequences for schools and school districts (and sometimes students and teachers) for how well students
perform on the tests, the accountability system establishes strong incentives for schools to focus almost exclusively on what is tested.

The panel followed with this recommendation: “The federal government should support a program of research and development of the next generation of assessment tools and strategies for accountability systems” (NAE, 2008).

Palm (2008) maintained that performance assessment “is viewed as having better possibilities to measure complex skills and communication, which are considered important competencies and disciplinary knowledge needed in today’s society.” In short, performance assessments are better suited for measuring the attainment of 21st century skills than are traditional assessments.

However, Gewertz (2008) noted, “assessing students’ grasp of 21st century skills is tricky.” Critics of performance assessment routinely call attention to the fact that scoring performance assessments can be highly subjective (Liskin-Gasparro, 1997).

In traditional classrooms, teachers communicate knowledge by lecture and demonstration, while students take notes, observe, perform drills, memorize, and take tests. Although these strategies characterize the traditional approach to teaching and emphasize rote learning, research has identified that “understanding [italics added] not memory should be the goal of instruction” (Alexander & Winne, 2006, p. 5).

Traditional “low-inference testing, based on the assumption that knowledge could be de-contextualized, was replaced by contextual assessment methodologies in science education, such as performance assessments, not on account of direct criticism, but rather on. . .the change from behavioral to cognitive psychology, developments in the philosophy of science, and the rise of constructivism” (Klassen 2006).

2.5.4 TYPES OF PERFORMANCE ASSESSMENT

Teachers using PBAs in the classroom have three types to choose from: products, performances, or process-oriented assessments. (McTighe & Ferrara, 1998).
Products are works produced by students that provide concrete examples of their application of knowledge, for example, writing samples, projects, art or photo exhibits, and portfolios.

According to Feuer & Fulton (1993) product form of performance assessment is the ubiquitous essay. Essays are a form of extended-response that requires the examinee to write a description, analysis, or summary in one or more paragraphs. Other extended-response products include term papers, laboratory reports, drawings, and dance performances. Products also include short, constructed-response items that require an examinee to briefly answer questions about reading passages, solve a mathematics problem, or complete a graph or diagram.

Performances allow students to demonstrate application of their knowledge and skills under the direct observation of the teacher. One approach in authentic assessment is observation. Performance assessment places great importance on observation as a key tool to creating developmentally-appropriate curricula for infants and toddlers. Employing a systematic approach, caregivers regularly collect information about each child’s behaviors and skills. However, unlike many forms of assessment this observation takes place in the context of everyday activities and a child’s natural setting. Authentic assessment observations focus on the process children use to learn, explore and interact, documenting each child’s individual physical, social, emotional, cognitive and language development. Observations help caregivers understand a child’s behavior and make an assessment of a child’s abilities (Muskie School of Public Service, 2010).

Students may engage in tasks that are useful outside of school, such as asking for directions by telephone, demonstrating a process, or arguing a position. All of these can demand high levels of science skill. Examples of performance tasks include oral reports, skits and role-plays, demonstrations, and debates.

Stiggins (1987) argued that performance concept include both oral assessments and demonstrations. In these instances, the assessment focuses on process. An example of an oral assessment occurs when a teacher assesses a student’s oral reading fluency and speaking skills.
Deaf (2006) considered the use of demonstration as a performance assessment occurs in the National Interpreter Certification examination. In this examination, candidates for certification demonstrate their ability to sign for the Deaf.

According to Pierce (2002), Process-oriented assessments provide insight into student thinking, reasoning, and motivation. They can provide diagnostic information on how well students use learning strategies and may lead to independent learning when students are asked to reflect on their learning and set goals to improve it. Some examples of process-oriented assessments are think-aloud, self-assessment checklists or surveys, learning logs, and individual or pair conferences. Products, performances, and process-oriented assessments can all be used to generate rich information on students' ability to transfer learning and meet state and local standards.

According to Feuer & Fulton (1993), sometimes, performance assessments involve a combination of performances and products. For example, experiments require students to engage in the scientific process by developing hypotheses, planning and executing experiments, and summarizing findings. The focus of the assessment can be the execution of the experiment (i.e., a demonstration), the laboratory report that describes the results (i.e., a product), or both.

Fitzpatrick & Morrison (1971) agreed the term “performance assessment” as shorthand for the type of assessment referred to as performance and product evaluation.

2.5.5 DEVELOPING PERFORMANCE ASSESSMENT

The development of performance assessments involves a general process that has been described by a number of authors (Allen, 1996; Brualdi, 2000; Herman, Aschbacher, & Winters, 1992; Moskall, 2003). The three basic steps in this process—defining the purpose, choosing the activity, and developing the scoring criteria—will be explained in the next sections.

1- Defining the Purpose

The first step in developing performance assessments involves determining which concepts, knowledge, and/or skills should be assessed. The developer needs to know what type of decisions will be made with the information garnered from the assessment. Herman et al. (1992) suggested that teachers ask themselves five
questions as they narrow down the myriad of possible learning objectives to be considered:

- **What important cognitive skills or attributes do I want my students to develop?**
- (e.g., communicate effectively in writing; employ algebra to solve real-life problems)
- **What social and affective skills or attributes do I want my students to develop?**
- (e.g., work independently, appreciate individual differences)
- **What meta-cognitive skills do I want my students to develop?** (e.g., reflect on the writing process, self-monitor progress while working on an independent project)
- **What types of problems do I want them to be able to solve?** (e.g., perform research, predict consequences)
- **What concepts and principles do I want my students to be able to apply?** (e.g., understand cause-and-effect relationships, use principles of ecology and conservation)

2- **Choosing the Activity**

The next step in the development of a performance assessment is to select the performance activity. Brualdi (2000) reminded teachers that they should first consider several factors, including available resources, time constraints, and the amount of data required to make an adequate evaluation of the student’s performance. In her synthesis of the literature on developing classroom performance assessments, Moskall (2003) made several recommendations:

- **The selected performance should reflect a valued activity** (i.e., a real-life situation).
- **The completion of performance assessments should provide a valuable learning experience.**
- **The statement of goals and objectives should be clearly aligned with the measurable outcomes of the performance activity.**
- **The task should not examine extraneous or unintended variables.**
- **Performance assessments should be fair and free from bias.**

3- **Developing the Scoring Criteria**

Wiggins & McTighe (2005) argued that last step in constructing a performance assessment is developing the scoring criteria. While traditional
assessments are comprised mostly of items for which the answer is either right or wrong, the difference is not as clear-cut with performance assessments (Brualdi, 2000). Rubrics are used to evaluate the level of a student’s achievement on various aspects of a performance task or product. A rubric can be defined as “a criterion-based scoring guide consisting of a fixed measurement (4 points, 6 points, or whatever is appropriate) and descriptions of the characteristics for each score point. Rubrics describe degrees of quality, proficiency, or understanding along a continuum”.

Before creating or adopting a rubric, it must be decided whether a performance task, performance product, or both a task and product will be evaluated. Moskal (2003) explained that two types of rubrics are used to evaluate performance assessments: “Analytic scoring rubrics divide a performance into separate facets and each facet is evaluated using a separate scale.

Developing performance tasks or performance assessments seems reasonably straightforward, for the process consists of only three steps. The reality, however, is that quality performance tasks are difficult to develop. With this caveat in mind, the three steps, with a brief discussion of each, follow.

Step 1. List the skills and knowledge you wish to have students learn as a result of completing a task.

As tasks are designed, one should begin by identifying the types of knowledge and skills students are expected to learn and practice. These should be of high value, worth teaching to, and worth learning. In order to be authentic, they should be similar to those which are faced by adults in their daily lives and work.

Researchers suggest that educators need to ask themselves five questions as they identify what is to be learned or practiced by completing a performance task. Their questions, with examples, follow:

1. What important cognitive skills or attributes do I want my students to develop? (e.g., to communicate effectively in writing; to analyze issues using primary source and reference materials; to use algebra to solve everyday problems).
2. What social and affective skills or attributes do I want my students to develop? (e.g., to work independently, to work cooperatively with others, to have confidence in their abilities, to be conscientious).

3. What meta-cognitive skills do I want my students to develop? (to reflect on the writing process they use; to evaluate the effectiveness of their research strategies, to review their progress over time).

4. What types of problems do I want them to be able to solve? (to undertake research, to understand the types of practical problems that geometry will help them solve, to solve problems which have no single, correct answer).

5. What concepts and principles do I want my students to be able to apply? (e.g., to understand cause-and-effect relationships, to apply principles of ecology and conservation in everyday lives).

**Step 2.** Design a performance task which requires the students to demonstrate these skills and knowledge. The performance tasks should motivate students. They also should be challenging, yet achievable. That is, they must be designed so that students are able to complete them successfully. In addition, one should seek to design tasks with sufficient depth and breadth so that valid generalizations about overall student competence can be made.

There is a list of questions which are helpful in guiding the process of developing performance tasks. Those questions, with their recommendations, follow:

1. How much time will it take students to develop or acquire the skill or accomplishment? The authors recommend that assessment tasks should take at least one week for students to complete. Others recommend that worthwhile tasks require far more time.

2. There are no rules regarding the appropriate length or complexity of a task; however, there are problems associated with developing overly complex and creative performance tasks. To begin with, relatively modest performance tasks are easier to develop. Furthermore, if they are well crafted and reasonably short (a few days rather than a few weeks), they are more likely to hold the interest of students. Finally, if a task fails to accomplish its purposes, it is best if the task is limited in duration.
3. How does the desired skill or accomplishment relate to other complex cognitive, social, and affective skills? Priority should be given to those which apply to a variety of situations.

4. How does the desired skill or accomplishment relate to long-term school and curricular goals? Skills or accomplishments which are integral to long-range goals should receive the most attention.

5. How does the desired skill relate to the school improvement plan? Priority should be given to those which are valued in the plan.

6. What is the intrinsic importance of the desired skills or accomplishment? Emphasis should be given to those which are important, while others should be eliminated.

7. Are the desired skills and accomplishments teachable and attainable for your students? Priority should be given to tasks which represent realistic goals for teaching and learning.

**Step 3.** Develop explicit performance criteria which measure the extent to which students have mastered the skills and knowledge.

It is recommended that there be a scoring system for each performance task. The performance criteria consist of a set of score points which define in explicit terms the range of student performance. Well-defined performance criteria will indicate to students what sorts of processes and products are required to show mastery and also will provide the teacher with an “objective” scoring guide for evaluating student work. The performance criteria should be based on those attributes of a product or performance which are most critical to attaining mastery. It also is recommended that students be provided with examples of high quality work, so they can see what is expected of them (Wisconsin Education Association Council, 1996).

**2.5.6 SCORING PERFORMANCE ASSESSMENT**

Two features of performance-based assessment help support the development of mental habits that lead to independent learning. The first is referred to as visible criteria. A fundamental tenet of performance-based assessment is the sharing of standards and making the criteria for evaluation visible to students. Teachers share their expectations for student work and performance in as explicit terms as possible,
using a scoring rubric, checklist, or other assessment tool and representative samples of student work. This approach is especially important with students, who have been shown to benefit from the teacher’s sharing of the assessment criteria in advance of the assessment itself (Kolls, 1992). When teachers state expectations for learning in terms of specific outcomes—in language the students can understand—and show them examples of excellent work, the likelihood of students attaining the criteria is greatly increased (McTighe & Ferrara, 1998; O’Malley & Pierce, 1996; Stiggins, 2002).

A variety of methods have been developed to score complex student performances, including both holistic and analytic approaches. In some cases, students are assessed directly on their performance; in other cases, assessment is based on a final product or oral presentation.

1) **Checklists**, rating scales and scoring rubrics facilitate judging and recording observations from performance assessments. Checklists are appropriate when the process or product can be broken into components that are judged to be present or absent, or adequate or inadequate (Oosterhof, 2003).

2) **Rating scales** are an assessment technique often associated with observation of student work or behaviors. Rather than recording the "presence" or "absence" of a behavior or skill, the observer subjectively rates each item according to some dimension of interest. For example, students might be rated on how proficient they are on different elements of an oral presentation to the class. Each element may be rated on a 1 to 5 scale, with 5 representing the highest level of proficiency.

3) **A scoring rubric** is used when performance is to be scored holistically. Performance assessments usually are scored more quickly with scoring rubrics, although the resulting scores tend to have lower reliability than when checklists or rating scales are used. Scoring procedures should be constructed to facilitate the efficient use of these scoring instruments. Checklists, rating scales, and performance assessments should be formatted so that brief comments can be quickly inserted during observations. Usually, qualitative scoring of performance is preferable, although numerical scores can be derived when needed, such as for assigning letter grades. Rubrics are sets of criteria that evaluate performance. Points are assigned
according to how well each criterion is fulfilled, and are then used to provide the quantitative values (Wiggins 1998).

Pierc (2002) stated the second key element of performance-based assessment is self-assessment, which is essential for teaching students how to manage their study habits, use learning strategies, and reflect on progress toward learning goals. The goal of self-assessment is to produce students who can learn independently of the teacher and become lifelong learners. To accomplish this, teachers need to provide students with specific feedback, opportunities to give and receive feedback from peers, and time to set learning goals. Self-assessment also plays a role in motivating learners to continue learning and building self-confidence in their ability to learn.

Dawson (2008) indicated the processes of teaching and learning metacognition that causes development of independent learners.

2.5.7 PERFORMANCE ASSESSMENT AND METACOGNITION

The educational benefit of using performance assessments has been demonstrated by a number of researchers (Lane et al., 2002; Niemi et al., 2007; Stecher et al., 2000). When students are given the opportunity to work on meaningful, real-world tasks in instruction, students have demonstrated improved performance on performance assessments. Performance-based assessment allows the student to construct his or her own answers as opposed to choosing from a group of answers (Lingtan & Towndrow, 2009).

One of the ways of promoting meta-cognition is through assessment. Haefner (2004) described an approach to assessment that engages planning, monitoring and evaluation, through three different mechanisms of assessment feedback. These engage students in setting goals, evaluating their performance and monitoring their understandings through techniques that are internal such as performance assessment.

Assessment practices would need to keep pace with developments in the sciences of learning and measurement including metacognition (Pellegrino, 2006; Tough & Reed, 2006; Cunliffe, 2005; Brown, 2003; Pintrich, 2002).

Educational reform in the 1980s was based on the premise that too many students knew how to repeat facts and concepts, but were unable to apply those facts and concepts to solve realistic problems that require complex thinking and reasoning.
skills. Assessments need to better reflect students’ competencies in applying their knowledge and cognitive skills to solve substantive, meaningful tasks. Promising advances in the study of both cognition and learning in content domains and of psychometrics also prompted individuals to think differently about how students process and reason with information and how assessments can be designed to capture meaningful aspects of student learning. Performance assessments that assess complex cognitive skills were also considered to be valuable tools for educational reform by policy makers and advocates for curriculum reform (Linn, 1993; Resnick & Resnick, 1982). They were thought of as vehicles that could help shape sound instructional practice by modeling to teachers what is important to teach and to students what is important to learn. Carefully crafted performance assessments that measure complex thinking and reasoning skills can serve as exemplars of assessments that stimulate and enrich learning rather than just serve as indicators of learning (Bennett & Gitomer, in press; Black & William, 1998). Performance assessments are needed to assess the types of thinking and reasoning skills that are valued by educators, and cannot be assessed by other item formats such as multiple-choice items.

According to O’Neil & Abedis’ study (1996) there are many assumed advantages of alternative assessments, for example, that such assessments should result in more effort expended and perhaps less anxiety. Further, such assessments should engage students in higher level thinking or meta-cognitive skills. We believe there is a need to measure such assumed advantages directly and explicitly.

Martinez & Katz (1996) stated there is evidence that the format of the assessment affects the type of thinking and reasoning skills that are used by students, with performance assessments being better suited to assessing high-level, complex thinking skills.

Performance assessments require students to perform a task such as conducting a science investigation as described above, or to construct an original product or response such as writing an explanation of one’s solution to a mathematics problem or writing a persuasive essay. Proficiency can be explained by the cognitive processes and skills involved in solving the performance task as well as the strategies chosen for a solution, having the potential to provide rich information for diagnosing strengths as well as gaps in understanding for individual students as well as groups of
students. When working on well-designed performance tasks, students may be engaged in applying their knowledge to real world problems, evaluating different approaches to solving problems, and providing reasoning for their solutions. A prevailing assumption underlying performance assessments is that they serve as motivators in improving student achievement and learning, and that they encourage instructional strategies that foster reasoning, problem-solving, and communication (Frederiksen *Performance Assessment: The State of the Art* & Collins, 1989; National Council on Education Standards and Testing, 1992; Resnick & Resnick, 1982).

Choi (2006) described the meta-cognitive evaluation method that can affect both the teaching and learning process of learners, especially novice learners, in consecutive interpretation. The basic idea of this study is to argue that evaluation of novice learners should be based on standards differentiated from those geared towards professional interpreters. The purpose and limitations of evaluation from a pedagogical standpoint are examined, followed by an overview of evaluation in the interpretation classroom. This study is noteworthy in that it attempts to propose a framework for performance assessment and to introduce the learning curve concept as part of assessing the learning process, which are presented as the main elements of the meta-cognitive evaluation method.

An assumed advantage of performance assessments is that they result in higher level thinking or meta-cognitive skills. This advantage should be measured directly and explicitly. Unfortunately, few standardized measures of meta-cognitive skills (planning, monitoring, cognitive strategies, and awareness) exist.

O'Neil & Abedi (1996) investigated the reliability and validity of a state meta-cognitive inventory: potential for alternative assessment. In this study for 12th graders, alpha reliability estimated and factor analysis indicated that our meta-cognitive subscales were reliable (alpha above .70) and one dimensional (one factor per subscale). Because the subscales had only 5 items each, they met brevity standards. Construct validity of state meta-cognitive inventory was acceptable. Results indicated that state meta-cognitive inventory yielded useful information about both the assessment and students.
O’Neil & Brown (1998) investigated the effect of item format on meta-cognitive and affective processes of children in the context of a large-scale mathematics assessment program. Mathematical items were presented in both multiple-choice and open-ended formats to 8th-grade students (N=1,032) as part of the California Learning Assessment System. Meta-cognition and affect were measured following each format for boys and girls of various ethnic groups. Results indicate that open-ended question formats have differential effects. Open-ended questions induced more cognitive strategy usage, less self-checking, and greater worry than did multiple-choice questions. These effects did not vary substantially as a function of gender and ethnicity.

Some researchers argue that the performance-oriented question format also challenges students to think critically and allows students opportunities to draw on prior knowledge and relevant skills to solve problems (Herman, Klein, Heath, & Wakai, 1994; Miller & Legg, 1993). Another claim is that these stimulate students to engage in complex thinking and thus reflect higher standards of excellence than traditional multiple-choice testing formats. However, the introduction of such assessment approaches brings questions of possible disparate impact on various ethnic groups (Baker & O'Neil, 1994; Winfield & Woodard, 1994).

2.6 SUMMARY

This chapter provided an overview of the pertinent literature in the fields of meta-cognition and learning, assessment and learning and meta-cognition and performance assessment. It also presented work that overlaps the fields of chemical education research and meta-cognition. Meta-cognitive skillfulness is widely deemed as necessary to achieve deeper understanding and to evolve into an autonomous learner; nevertheless advancement in its assessment does not match this acceptance. This lack of adequate assessment, caused primarily by the inherent difficulties of measuring a non-overt construct, has also limited the assessment of the effectiveness of meta-cognitive instruction. There is a tacit assumption that making meta-cognitive learning environments available does develop meta-cognition use in students. However, there is no direct evidence for this assumption.