(Article: 3)

Developing Performance-based Assessment Instrument in Science: A How-To Guide

GOURANGA SAHA[1*]

RODNEY L. DORAN[2*]

[1*]*Professor of Science & Technology Education, Department of Education, Lincoln University, Jefferson City, Missouri 65102-0029, USA; Email ID: sahag@lincolnu.edu*

[2*]*Professor Emeritus, Department of Learning and Instruction, State University of New York at Buffalo, New York, 14260-1000, USA*

Abstract

Performance-based assessment procedure directly corresponds to inquiry science learning for students to show not what they know but also what they can do. When constructivist science learning requires students to perform, inquiry science teaching becomes an appropriate platform for the students to perform. Inquiry teaching may be frustrating if the assessment activity (task) which tends to show learners' 'sciencing' skills jeopardizes them to perform. This study documents how to develop a doable inquiry assessment activity for all students in diverse instructional facilities. Science performance activities developed during this research process, demonstrated a high level of reliability and validity.

Key words: *Performance-based assessment, Inquiry teaching, constructivist classroom, reliability and validity, technical qualities, process skills, Diffusion.*

1. Introduction

Performance-based assessment has been tied to constructivist classroom as a valid tool in measuring students' knowledge and skills across all the content areas. Laboratory is a common and vital place for doing science ('sciencing'). Challenging performance assessment activities allow learners to demonstrate what they know and can do. Again, it is the performance activity that helps teachers measures the insights of their students' true level of capability in any domain of knowledge (Joyce, 2015). Because students are assessed during the inquiry activity on their process skills in constructing new knowledge, performance assessment is vital in evaluating the effectiveness of the inquiry science. Even in those instances where performance–based activities are used to measure inquiry, limitations of these activities hinder students to perform and assessment outcomes become unreliable and invalid (Reynolds *et. al*., 1996). For this reason, developing performance assessment tools that adequately measure science inquiry is essential to the quality science curriculum reform movement.

Introduction

Research in gifted education indicates that performance-based assessments have been found to be valid and reliable measures of student learning, including science acquisition (Adams & Callahan, 1995; Moon, Brighton, Callahan, & Robinson, 2005). Unlike most other forms of assessment, performance-based assessments do not have only one right answer. Instead, there are levels of proficiency which may be attained by students, which means instructors need an assessment tool to collect data on learning allowing them to rate each student's performance (Meyers, 2008) validly. A consistent scoring guide can accomplish this need. Extending Doran and Hejaily's (1992) work, this paper describes only the trial testing process to illustrate a laboratory-based performance-based activity on *'Diffusion'* (Saha, 2001) to document what and how we did, what worked, and what didn't during the research for developing reliable and valid performance-based tasks/activities for inquiry-based science assessment process. Although sticking to this order is not essential, the intent of the following stepwise guidelines is to help develop dependable and doable activities for assessing students' science performance (process) skills.

2. The process of development :

Inquiry assessment activity development process involves an iterative process of trial testing (Saha & Doran, 2009) that calls for frequent reviews, revisions, modifications and changes from feedback at several stages. For the convenience of describing how new activities are developed, the process of science assessment activity development and trial testing can best be viewed within several stages. However, each of these stages may be visited several times for a specific activity. These stages are presented in Fig. 1 and the guidelines and concerns at each of these steps can be found in figure 2.

3. Illustration of development process with *Diffusion* Activity:

The source of the *diffusion activity* was the set of NORC (National Opinion Research Center) activities of Hejaily and Doran (1993). The initial NORC activity used three different concentrations of potassium permanganate solutions ($KMnO_4$) to determine the relative diffusion rate in raw potato cubes. The major drawback of this activity was the difficulty for the students to measure the rate of diffusion of $KMnO_4$ solution into potato cubes. Time and unevenness of diffusion "lines" became critical factors. Need to develop a new or a modified version of this activity was felt mainly from these weaknesses of the activity.

Another stage of brainstorming evoked the idea of using gelatin and or agar medium for food color/$KMnO_4$ solutions to diffuse through. The activity was written using these materials and began the in-house nano-testing. This time to the diffusion rate was too slow to measure within 2-3 minutes. Mainly for the time factor these efforts went in vain. Another round of brainstorming began.

Illustration of development process with *Diffusion* Activity:

The process of diffusion is contingent on concentration gradient between two substances. Based on this concept or the "effect of concentration gradient" on the movement of particles of matter it was thought whether a solution of $KMnO_4$ would show a phenomenon of diffusion in ordinary tap water? In order not to let this idea escape, it was incorporated into the activity and began the usual in-house nano-testing. $KMnO_4$ solution diffuses in tap water under STP (Standard Temperature and Pressure) equally in all directions. The rate of this easily visible diffusion process can conveniently and quickly be measured using a metric ruler. It is essential to make sure that there should be a short time limit within which to measure the diffusion rate to avoid complete mixing up and distortion of the diffusion "line."

In the initial writing process an attempt was made to find the rate of diffusion of 1%, 3% and 10% $KMnO_4$ in tap water under STP. Similar volumes of water were taken in a petri dish and dropped one drop of the $KMnO_4$ solution in it. In the in-house nano-testing it was found that the rate of diffusion between 1% and 3% was not easy to distinguish. So testing with different concentrations began and ultimately found that 1%, 6% and 10% worked very fine. The distance that 6% $KMnO_4$ diffuses is greater than 1% and that of 10% is greater than the other two solutions.

Based on these findings the activity was reviewed and revised and then started the next phase of trial testing - the *in-house micro testing* with two other authors of this article. At this stage it was found that only fresh 10% $KMnO_4$ solution works best; however after several experiments it was realized that before an old $KMnO_4$ solution works without any visible problem, it should be shaken thoroughly for several minutes. The second observations at this stage helped determine the exact time - which is two minutes at which the diffusion rate should be measured. The third was how to facilitate measuring the diffusion rate by the students with more ease? A suggestion was made to create a grid of 1 mm square. A transparency was made from a millimeter graph paper and cut in to circles to fit the base of the petri dish. Questions to assess students' inquiry skills were also reviewed to include the effect of hot water on diffusion rate of any one concentration. Item 1 was changed to construct a data table in place of the initial question for prediction with one of the three solutions. A new item (#3) for *interpolation* from the student's graph to predict the diameter of a drop of 3% $KMnO_4$ was constructed. After trial testing, it was found that the activity took more than 17 minutes to complete. To reduce this time length the background information was made more compact and decided to decrease the number of pages using single space between the lines of text.

Illustration of development process with *Diffusion* Activity:

Fig. 1: Trial Testing Stages

As they move to the last stage,activities become more refined and translate into cognitively appropriate and thus fit more to the target program.

Stage V. Field Testing
Activities that are to be used with a large number of students (like district, state, etc.) need to be field tested among representative samples of target grade/course.

Stage IV. Pilot testing
The activities that survived micro and mini testing are given to several classroom teachers. These teachers set up testing for an entire (intact) classroom. Administration should preferably be scheduled ahead so that

Stage III. Mini-testing
The activities successfully micro tested are then ready to be shared with a few other classroom teachers for set up and administration to their students to a group of 8-10

(d) School-based micro-testing
These reviewed and revised activities are then sent to classroom teachers of 2-3 schools for further tuning. At this stage each classroom teacher micro-tests each of these activities at his/her own, examining their usability from his/her perspective

(c) In-house micro-testing with a group of school students
At this stage of the micro-testing a few school students of the same grade level/course for whom these tasks are meant are invited in the researcher's/developer's classroom.

(b) In-house micro-testing with colleagues
Because the developing assessment activities at this stage is prone to malfunctions, odd results, confusion, and other mishaps, 'friendly' colleagues with a keen interest in science assessment should be invited to trial test the activities. Based on their input, the activity is reviewed and revised accordingly.

(a) In-house pre-micro testing or Nano-testing
Researcher tests, reviews and revises the activities to see that they (activities) work as expected and fine tune them to a point where he/she feels that other colleagues can try them.

Stage II. Micro-testing
At this stage the most promising activities are selected from the activities idea pool, the necessary materials are gathered and then tested to see whether materials and equipment work the way they are supposed to work.

StageI. Pooling activities
Once the desired program/curriculum level is selected, a pool of activities is gathered that may be appropriate for a district or state curriculum. Inquiry skills and program/curriculum content areas, behavioral objectives, required materials, all safety precautions, questions, and directions are identified, developed and written.

The activity was reviewed and revised from the reflections of this stage of trial testing and conducted the in house micro testing with three 10^{th} grade students from a local school. Responses were analyzed and found that to investigate the problem with three concentrations and one also with hot water is time consuming. So, decision was made to use only 1% and 6% and only with tap water (room temperature) under STP. Items were also revised to include some completely new questions than the initial draft of the activity. Such as, “title of the graph” and the “grid of the graph” were separated from each other based on the observation that most of the students do not give title for their graph. To assess whether students do have planning skill of an experiment, it became essential to see whether the students could figure out at least one limitation of this experiment? Consequently, question/item #4 was framed (Identify one possible error that may have been made in this experiment). It was also found that students should have a cue in the direction for the unit to use in their measurement. With necessary review and revisions this stage of trial testing was repeated with two batches of student teachers. Based on recommendations of these students the procedure step 1 was revised and replaced the word *pour* with *take*; the exact height (2 cm) from which to drop the $KMnO_4$ solution in water was mentioned and also replaced the word *one* in place of *a* drop. Previously the outline of the graph was not provided that generated inconsistencies in the type of graph and units used. Interpolation for 3% was too easy as it is a half of 6% these teachers opined. So, it was replaced with 4%.

This reviewed and revised version of the activity was then given to three classroom teachers for micro testing. From the reflection of their trial testing, the activity was further reviewed and revised. In the procedure exact volume (25 ml) of water was included to avoid confusion and inconsistencies among the students. Previously directions were not specific as to where in the petri dish to drop the $KMnO_4$ solution, this confused many students when it came to measure the diffusion area. So revision was made to drop the solution at the center of the petri dish and all the directions were rephrased to make clear that one drop for 1% and one drop for 6% should be used in separate petri dish.

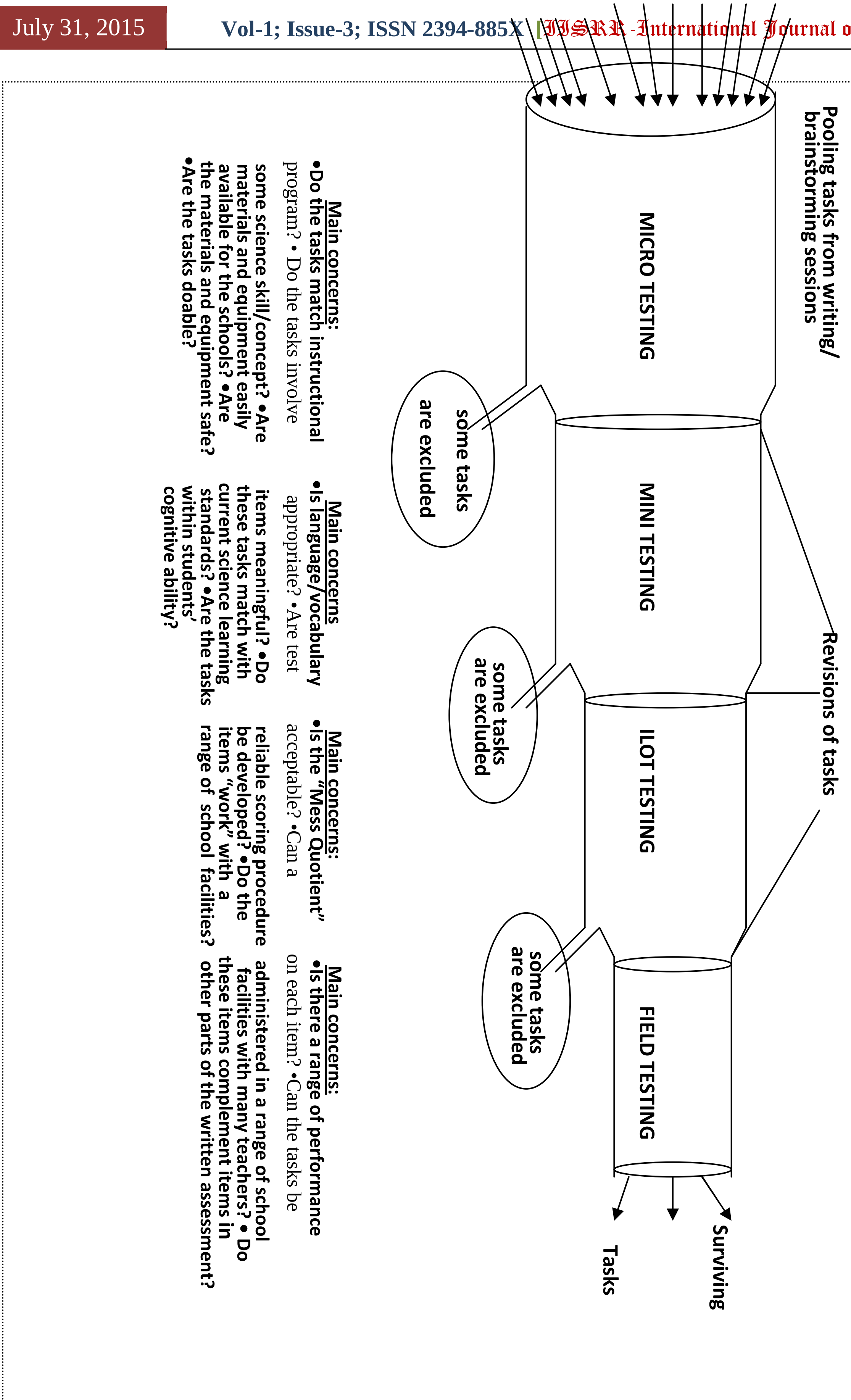
The revised activity was then forwarded to 4 classroom teachers for mini testing with greater number of their students. The reflections from the student responses facilitated further reviewing and revisions. For example, analyses indicated that to read the list of materials students take much time. So, decision was taken to draw a sketch of the materials for students to ascertain quickly what materials they needed and which one is missing, if any, from the station. In fact, a diagram or picture speaks thousand words. Procedure step A was rephrased as “place 25 ml of water in each of the two petri dishes” in place of “in each of the petri dishes.”

Illustration of development process with *Diffusion* Activity:

Since many students pushed the petri dishes or the working bench during their experiment the diffusion circle was not distinct and easy to measure. So the sentence “Don’t tilt or move the petri dish” in part B of the procedure was added. Item 2 was revised to add the sentence “Be sure to include a title” and instead of the sentence “Label the two axes with appropriate variable and unit” a new sentence “Label the axes with appropriate scales” was added. The word ‘interpolate’ was removed as most of the students did not understand the meaning and asked to use their graph to predict the diffusion diameter with a drop of 4% $KMnO_4$. The wording of item 5 was not clear to many students so we rephrased as “Write a hypothesis for a new experiment that investigates the effect of temperature of water on the diffusion of $KMnO_4$ “For a new experiment” part was not there previously.

Thus the diffusion task that evolved for the field-testing can be described as a largely new one if all its revisions and changes are tracked and compared. This new version is nothing but the metamorphosed task of the initial *Diffusion* task that the researcher started with. Data (Saha & Doran, elsewhere) from the field-testing on performance tasks/activities thus developed from this arduous and iterative procedure, were found to be technically sound activities to use as performance-based assessment tools for all students in any facility factor.

Fig. 2. A model of laboratory-based assessment activity development process



References

1. Doran, R. L. & Hejaily, N. (1992). Hands-on evaluation: A how to guide. Science Scope, 15(6), 9-11.
2. Hejaily, N. & Doran, R. L. (1993). *Science laboratory test: biology (NORC)*. State University of New York at Buffalo.
3. Joyce, V. (2015). Arguments for and against the common core state standards*. Gifted Child Today*. 38(1), 60-62).
4. Kubiszyn, T. & Borich, G. (2001). *Educational testing and measurement* (7th ed.). John Wiley & Sons, Inc.
5. Meyers, S. (2008). Research Starters Education (Online Edition).
6. National Research Council (NRC, 1996). *National science education standards.* Washington DC: National Academy Press.
7. NORC (National Opinion Research Center) (1991). *Science laboratory test: Biology*. UAB/NORC and New York State Alternative Assessment in Science.
8. Reynolds, D. S., Doran, R. L., Allers, R. H. & Agruso, S. A. (1996). *Alternative assessment in science: A teacher's guide.* New York State Education Department, University at Buffalo.
9. Saha, G. (2015). A technology supported multi-sensory approach to science teaching model. *ISSR- International Journal of Research*, 1, 51-56. V; Issue- Published in January 2015.
10. Saha, G. (2001). *Implementing the science assessment standards: Developing and validating a set of laboratory assessment tasks in high school biology*. Unpublished Ph.D. Dissertation, State University of New York at Buffalo.
11. Saha, G & Doran, L. (2009). Developing and validating performance-based assessment tasks in science: A how to guide. In proceedings of International Science Education Conference (ISEC), 24 to 26 November, 2009 at National Institute of Education, Singapore.

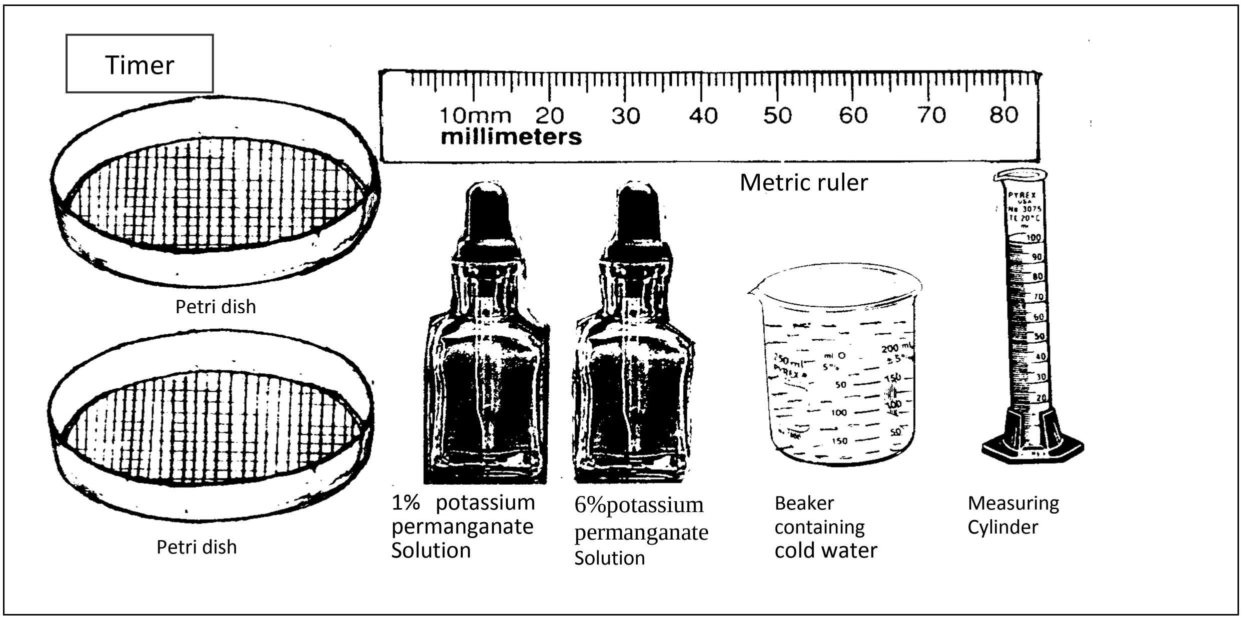
Appendix

Diffusion Task

Task: You will investigate the rate of diffusion of two different samples of a solution.

Background: Diffusion is the net movement of molecules in a fluid (gas or liquid) from a region of high concentration to a region of low concentration driven by the concentration difference. The distance of the movement of molecules in a given time can measure the rate of this movement.

Materials:



Procedure:

A. Pour 25 ml of water in each of the two petri dishes.

B. Using the dropper attached to the bottle, drop very carefully from a height of about 2 cm (so that it does not splash) one drop of 1% potassium permanganate solution ($KMnO_4$) in the center of the first dish and one drop of 6% $KMnO_4$ in the center of the second petri dish. Don't tilt or move the petri dish.

Diffusion Task

Answer Sheet

Questions:

1. Observe what happens to the two drops after two minutes. Record the distance the color spreads in the following table (use the largest distance of the color):

Concentration		
Diameter		

2. Graph your data. Be sure to include a **title**. Label the axes with the appropriate scales:

Title of the graph:

3. Use the data in your graph to predict the diameter of a drop of 4% $KmnO_4$.

4. Identify one possible error that may have been made in this experiment?

5. Write a hypothesis for a new experiment that investigates the effect of temperature of water on the diffusion of $KMnO_{4.}$

STOP

Table1. Scoring Guide

Item	Answer	Scoring	Item value (points)
1	Different rate	1 point for completing each column	2
2a	Diffusion Rate of two solutions	1 point for correct answer	1
2b	Plotting Graph	1 point each for correct plotting, scale, unit, X & Y axes	4
3	Interpolation falls between 1% and 6%; Prediction falls + or – 5 mm or + or – .5 cm of the correct 4% value	1 point for each correct answer	2
4	Relevant limitation, desk tilts, drop splashes	1 point for correct answer	1
5	**Warmer temperature equals more concentration**	**1 point for each – independent and dependent variable and link to their effect.**	**3**