51-MJA-C001

Developing an Evaluation Rubric for Hands-on Science Education Activities

An Interactive Qualifying Project submitted to the faculty of WORCESTER POLYTECHNIC INSTITUTE in partial fulfilment of the requirements for the Degree of Bachelor of Science by:

Scott Cornman
Grace Holliday
Gregory Tersoff

April 30, 2013

Report submitted to:
Professor Michael Aghajanian, Advisor
Professor Holly Ault, Co-Advisor
Chris Krishna-Pillay, CSIRO Liaison
Carly Siebentritt, CSIRO Liaison

Abstract

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) conducts science education programs in schools across Australia. This project developed a formal rubric to evaluate the effectiveness with which the hands-on activities in these programs teach students science concepts in an engaging ways. Through educator interviews and surveys, program observations, and discussions with teachers and students, this rubric was developed and tested. The rubric was found to be valid in accurately conveying key information about activities, reproducible when used by different evaluators, and universally applicable to all age groups. The project team also provided recommendations for improvements to the rubric and used the rubric to inform design recommendations for two new activities.

Acknowledgements

The project team has an extensive list of people (and insects) to whom they owe their gratitude.

Firstly, we would like to thank our advisors, Professors Mike Aghajanian and Holly Ault, for their efforts in making sure our project and paper were the best they could be and for humouring our requests for more feedback on certain sections of the draft. We would also like to thank our ID2050 professor, Professor Nicola Bulled, for her endless helpful feedback on our project proposal in C-term.

We would like to thank all of the staff at CSIRO Education, Victoria who contributed to our project by taking time to fill out our surveys and speak with us. We would especially like to thank the following educators for letting us observe their programs and for not stranding us at various train and bus stations around Melbourne. Thank you to Simon, Sarah, Gemma, Kate, Jason, Elke, Claudie, and Mick for letting the Americans invade your program sessions and play with the activities. We'd also like to thank Ava for always being there to answer both the front door and all of our questions.

Additionally, we want to apologize to Cath for insinuating that she had never seen an icicle. No one told us you were originally Canadian. We're sorry!

We would like to acknowledge our stick-bug friends who kept us company through most of our long days in the Green Room and honour the ones who passed during our seven-week stay.

Finally, we owe many, many, many thanks to our project liaisons, Chris Krishna-Pillay and Carly Siebentritt for all of the help and support that they've given us this term. Thank you to Carly for your helpful advice and feedback on our work, and for the effort you put in to find us bookings and travel arrangements for all of the programs we needed to see. Without you, we never would have fulfilled our dream of visiting every train, tram, and bus station in the Melbourne metropolitan area. Thanks to Chris for sharing with us so much of your wisdom on program design, observation skills, presentation techniques, footy, Australia, and life. Your insights were invaluable to our project, and your comic relief was invaluable to our sanity. Our chats with you were always the best parts of our days at work. Many thanks in particular for saving the day when our initial project fell apart, helping us to put the pieces back together into a new project that was far more useful than the old one could have been. Lastly, thanks to you both for taking time on a Saturday to see a footy game with us, specifically to Chris for reserving our seats and to Carly for going along (and even getting your face painted!) despite having no interest in the sport whatsoever. We enjoyed the game and your company immensely!

Table of Authorship

All group members edited all sections.

Section Number	Section Title	Primary Author(s)
	Abstract	Scott
	Executive Summary	All
1.0	**Introduction**	All
2.0	**Background**	Greg
2.1	The Commonwealth Scientific and Industrial Research Organisation	Scott
2.1.1	CSIRO Science Education Programs	Grace, Scott
3.0	**Methodology**	All
3.1	Education Activity Research and Criteria Development	All
3.1.1	Analysis of Education Techniques	All
3.1.2	Development of an Activity Evaluation Rubric	All
3.2	Flexible solar cell activity design	All
3.2.1	Research into educational applications of flexible solar cell technology	All
3.2.2	Development of flexible solar cell activity recommendations	All
4.0	**Results and Discussion**	All
4.1	Rubric Design	All
4.1.1	Rubric Creation	All
4.1.2	Rubric Weighting	Scott, Greg
4.2	Assessment of Rubric	All
4.2.1	Validity of Rubric	Grace
4.2.2	Reproducibility of Rubric	Scott
4.2.3	Universality of Rubric	Greg, Grace
4.3	Activity Design Recommendations	All
4.3.1	Energy: Sources and Uses	Greg
4.3.2	Polymers and Nano-Chemistry	Greg, Grace
5.0	**Conclusions and Recommendations**	All
5.1	Project Summary	All
5.2	Recommendations for Rubric Implementation	All
5.2.1	Use of Rubric	Grace
5.2.2	Suggested Rubric Improvements	Grace, Greg

Table of Contents

List of Figures

List of Tables

Executive Summary

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) is Australia's national science agency, consisting of both research and education branches. The education branch offers a wide variety of informal science education programs for students of all ages in schools across Australia. In Victoria, CSIRO Education officers transport these programs into schools and present them to classes of 30 to 60 students. These informal science programs typically consist of three main components: an introduction with a few demonstrations of the main concepts covered by the program, a period of hands-on activities and experimentation, and then a wrap up discussion that may include another demonstration to conclude the program. These programs are designed to be interesting, engaging and educational.

Over time, CSIRO Education will adjust the makeup of these programs, adding or removing activities to fill new needs or strengthen a particular aspect of a program. Currently, CSIRO Education holds a yearly staff meeting, in which educators discuss the strengths and weaknesses of these programs and propose modifications. As this discussion includes no formal structure for analysing the strengths and weaknesses of programs, decisions come down to subjective conversation based on years of experience.

While these discussions are fairly effective, due to the staff's years of experience with specific CSIRO education programs and education as a whole, they can become difficult without some framework to guide the conversation. This project set out to create an effective way for CSIRO education officers to standardise this communication and allow them to quickly point out strengths and weaknesses of an activity. To accomplish this goal, we developed an activity evaluation rubric that enables educators to translate their experience with a program into quantified scores across several criteria that measure the effectiveness of the program's hands-on activities.

Similarly, CSIRO lacks a systematic means of designing new activities and programs to meet key objectives. This rubric will also give CSIRO educators a standardised starting point for the design of new activities and programs. To test the effectiveness of the rubric as a design tool, we developed design recommendations for activities involving new flexible solar cell technology designed by CSIRO's Materials Science branch.

To develop the rubric, we studied current CSIRO education programs through a combination of background research, discussions with CSIRO educators, and first-hand experience observing CSIRO science education programs as they are normally run. The main purpose of this research was to generate an informed list of important features of hands-on activities, at the same time determining the effects of each feature and how important it was to the success of an activity. Through this research, we were able to create the following list of main criteria that successful activities should exhibit, along with their measurable component sub-criteria:

- **Educational Value**
  - Curriculum Relevance
  - Potential for Teacher Extension
  - Encourages Student Thought
- **Logical Feasibility**
  - Health, Safety, and Environment
  - Cost
  - Ease of Setup/Takedown
  - Ease of Transport/Handling
  - Durability
  - Ease of Repair
- **Initial Draw** (No component sub-criteria)
- **Student Satisfaction**
  - Student Interaction
  - Novel Completion Process
  - Relevance to Students' Lives
  - Surprising/Unexpected Results
- **Ease of Completion**
  - Clear Instructions/ Expectations
  - Intuitiveness to Complete
  - Appropriate Length

These criteria measured the key elements of an effective program, covering how likely a student was to begin an activity, how easy the activity was to perform, how much was learned, how much the student would enjoy and remember the activity, and how many resources it required as part of the program. Each of the sub-criteria, which made up these requirements, was represented on the rubric as a 0-5 scale, with descriptions defining the traits that activities must exhibit to earn each possible score. These descriptions were made as clear as possible, to aid educators in giving similar scores based on their experiences, rather than subjective opinions on what those experiences represented. This allowed educators to begin the discussion of a program from the common ground of their rubric scores, before deciding on the significance of these scores to any proposed program modifications.

We ensured the rubric worked by evaluating a variety of programs and activities. Through these evaluations, we confirmed:

- The validity of the rubric in effectively and accurately measuring aspects of the main and sub-criteria
- The reproducibility of scores across educators to ensure that the rubric would give them a common ground in their assessment of activities
- The universality with which the rubric can be used to evaluate the wide range of programs and activities that CSIRO offers.

We determined that the large majority of criteria and sub-criteria were valid, reproducible and universally applicable. However, some sub-criteria are weaker than others in these aspects, and the project team has made recommendations on how the rubric can be improved for continued use by CSIRO education staff. These recommendations are:

- Adding a "coolness" sub-criteria to the *Satisfaction* main criterion
- Separation of *instructions/expectations are clear* into two sub-criteria
- Adding text to the *ease of repair* scoring descriptions to include replacement of broken parts
- Rewording the *student interaction* scoring descriptions to put less emphasis on experimental variables
- Rewording *relevance to students/familiarity of concepts* or separating it into two criteria
- Reducing weight of *Logical Feasibility* main criterion as the target audience of the program becomes older.
- Increasing weight of *Educational Value* main criterion as the target audience of the program becomes older.

The rubric we created will also be a useful design tool for future hands-on activities and new programs. To validate this claim, we used the rubric as a guideline for providing design suggestions for two future flexible solar cell activities. These design recommendations are specifically geared towards the inclusion of these flexible solar cell activities in the Energy: Sources and Uses program and Polymers & Nano-Chemistry program.

1.0 Introduction

As the world becomes increasingly dependent on new advancements in science to solve problems, improved tools and technologies are needed to keep up with the demand for progress. In order for these improvements to be made, there needs to be a motivated group of potential scientists to help work on these problems, and a society that understands the effort and resources required to solve them. It is therefore important to stimulate student interest in science whenever possible, encouraging students to view science as a potential career path and giving them the science awareness necessary to make informed policy decisions. If interest in science lessens on a societal level, crucial problems in that society could be left unsolved.

Australia is one country where students' interests in science are declining. The percentage of students taking higher level science courses has been steadily decreasing since the 1980s (Ainley, Kos, & Nicholas, 2008). In contrast, the need for new scientists has increased as many areas of Australia seek to improve their society and environment through science-based endeavours such as carbon emission reduction programs. As a smaller portion of the population studies physics, biology, and chemistry, the number of people capable of solving these problems decreases. This is especially worrisome for any projects requiring a gradual shift towards new technologies that will, over time, require new solutions and scientists to champion these innovations.

There are many organisations already in Australia trying to reverse this trend of decreasing interest in science. The Commonwealth Scientific and Industrial Research Organisation (CSIRO) works to further the cause of science both through research and development, and by raising public awareness. The CSIRO Education Victoria branch designs and runs unique, non-formal, hands-on science programs that are used by teachers to supplement their curriculum. Through these activities, CSIRO endeavours to make science interesting for future generations by teaching lessons in innovative and interactive ways. While CSIRO has a number of programs on a variety of subjects, these programs are frequently added to and updated in order to improve how effectively they stimulate interest in science.

One such update is currently underway with CSIRO's existing programs, in order to include cutting edge flexible solar cell technology that CSIRO materials scientists have developed. While some programs have utilised solar technology in the past, silicon based solar panels were used, which are both more difficult to manipulate, and lack the unique characteristics of the polymers used in the flexible solar cells. CSIRO Education has not utilised flexible solar cell technology in its programs before and is interested in including novel activities that will allow for new demonstrations of the technology.

In addition, CSIRO has no standard by which to determine how programs should be updated. CSIRO currently relies on the instincts and intuition of educators to determine which portions of programs are effective and which programs need to have activities added, removed, or modified. A quantifiable method of judging activities that could be used by staff with a variety of backgrounds and experience will improve the evaluation process, requiring fewer abstract opinions and allowing educators to more easily explain positive or negative features of an activity.

Our project developed a rubric to judge activities based on importance-weighted design criteria, to be used by CSIRO educators to evaluate existing and proposed activities. Informed by these criteria, we also developed design recommendations for two new hands-on educational activities involving flexible solar cell technology that CSIRO could add to its existing *Energy: Sources and Uses* program and to its *Polymers and Nano-Chemistry* program.

2.0 Background

Declining science interest is evident internationally and is a concerning worldwide problem. Studies conducted by the organisation known as Relevance of Science Education (ROSE) surveyed secondary-school students' opinions on science in 40 countries. The surveys consisted of ranking a number of given statements on a scale from 1 (disagree) to 4 (agree) (Sjoberg & Schreiner, 2005). Responses to the following two statements show a concerning disparity between children's opinions on science: *Science and technology are important for society* (Figure 1) and *I would like to become a scientist* (Figure 2). Students worldwide recognise that science and technology benefit society, but have little interest in pursuing science themselves; a trend that is especially prevalent in the highly developed countries.

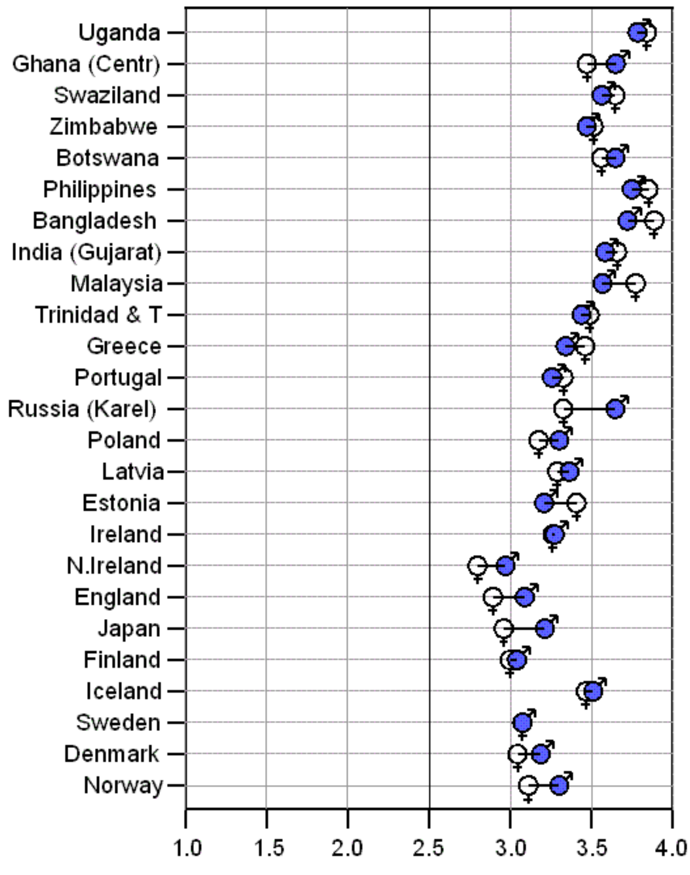


Figure 1: Science and Technology are important for society.
Average scores for boys (filled symbols) and girls (open symbols). "Trinidad & T" denotes Trinidad and Tobago.

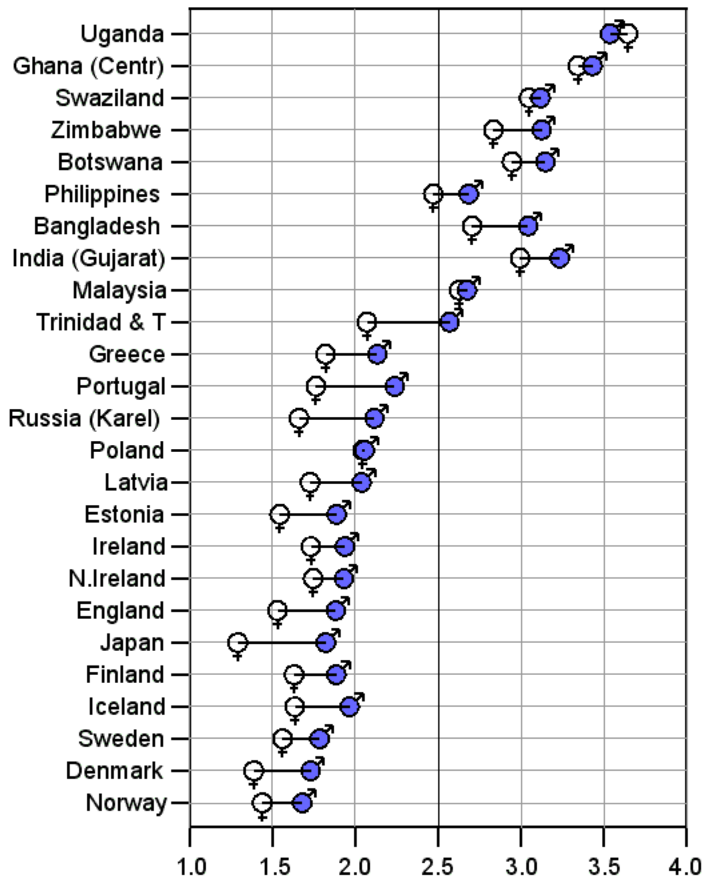


Figure 2: I would like to become a scientist.

Average scores for boys (filled symbols) and girls (open symbols). "Trinidad & T" denotes Trinidad and Tobago.

In Australia, student interest in science is also low. Since the early 1980s the percentage of students in Australia taking scientific courses has dropped, with the only major exceptions being psychology and several unique or infrequently provided science classes (Ainley et al., 2008). From 1983 to 2008, the proportion of students enrolled in physics classes dropped from 28.1% to 14.6%; in biology from 48.3% to 24.7%; and in chemistry from 32.1% to 18.0% (Ainley et al., 2008, pp. 15-16). None of the subjects have shown any periods of noticeable growth over the past 25 years, making a reversal of this downward trend extremely difficult (see Figure 3).

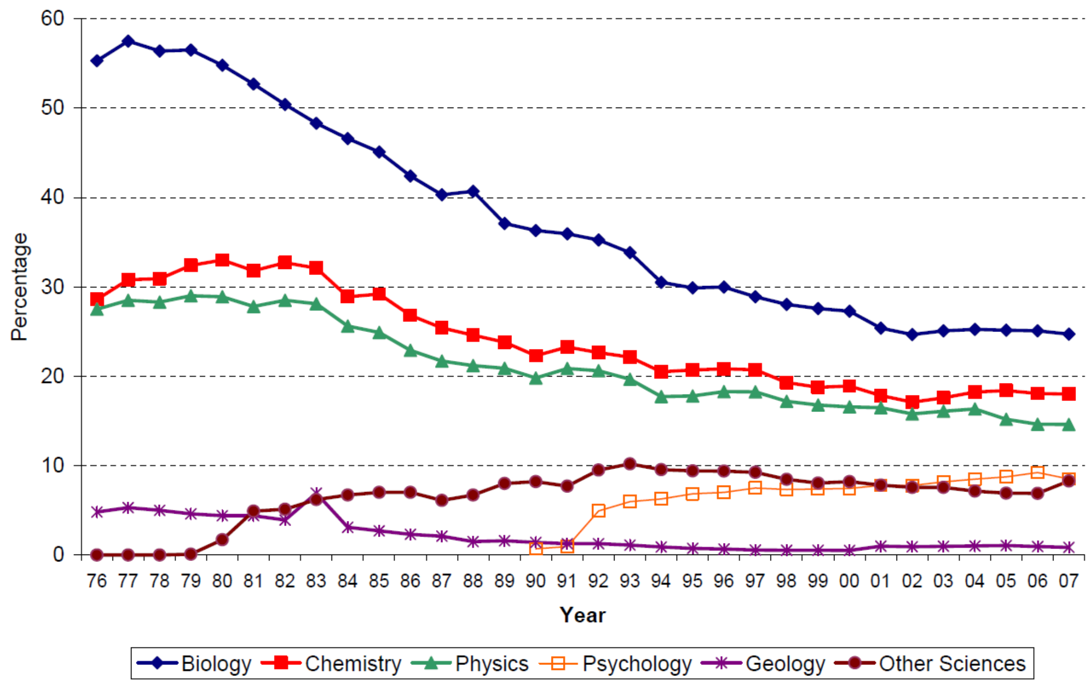


Figure 3: Percentage of Australian Year 12 Students Taking Science Courses by Subject

Graph from (Ainley et al., 2008)

Increasing enrolment in scientific courses through the creation of an environment that encourages scientific disciplines may inspire the pursuit of scientific careers. As fewer students take science classes, it is likely that fewer students will be interested in pursuing science as a career. While a number of steps could be taken to reverse this situation, they all tend to share the common link of making students more interested in science.

Being interested in science fosters the further pursuit of the sciences in university and as a career. A survey of Australian university graduates who self-identified as having a background in science indicated that more than 75% enrolled in physics courses, more than 85% in biology courses, and more than 95% in chemistry courses beyond year 10 of their studies (Harris, 2012). Enrolment in science classes (i.e., physics, biology, and chemistry) was independent of whether or not their degrees required a scientific background (Harris, 2012). As suggested by Harris (2012), a strong positive correlation exists between a scientific outlook and taking science courses in years 11 and 12.

Interest in science starts at a young age and can lead to a lifetime scientific outlook. Kerri-Lee Harris, a researcher at the Melbourne University Centre for the Study of Higher Education, found that one in four people who pursued a science degree in university cited an influence in their early lives from their parents; teachers and the school system; or other factors such as reading, extracurricular science activities and exposure to inspirational environments as a part of their science backgrounds (2012). Furthermore, according to Jon D. Miller, research scientist at University of Michigan's institute for social research, "Approximately 90 percent of the high-school students who did not plan to attend college failed to meet minimal criteria for interest in scientific issues or for cognitive knowledge of basic scientific constructs" (1983, p. 36). Making young students more interested in learning about science will both increase the number of students taking science classes at higher levels and increase the number of students pursuing careers in science. The Commonwealth Scientific and Industrial Research Organisation, through its non-formal education programs, hopes to create these early life experiences for children and inspire students to pursue science beyond the requirements of formal classroom learning.

2.1 The Commonwealth Scientific and Industrial Research Organisation

As Australia's national science agency, CSIRO is one of the largest government-affiliated scientific agencies in the world (CSIRO, 2009). Their mission is to develop scientific solutions for the betterment of Australia and society in general. To attain this, CSIRO consists of both research and education branches, which combine for more than 6500 employees spread over more than 50 sites all across Australia and overseas (CSIRO, 2011). CSIRO's educational branch offers a variety of science programs for schools and other settings, supplementary tools and materials for teachers, and science activities for the home. Through these non-formal education offerings, CSIRO Education strives to promote interest in science among students of all ages. To this end, CSIRO has developed numerous effective non-formal education programs covering a variety of subject areas including astronomy, robotics, environmental issues, forensics, and various chemistry and physics topics.

2.1.1 CSIRO Science Education Programs

In Victoria, CSIRO Education offers 28 different educational programs to groups of students from area schools, ranging in year level from Foundation and the primary school years through year 11 and 12 studies toward the Victorian Certificate of Education (VCE) requirements. Sessions of these programs typically run for 30 to 90 minutes, depending on the program, and are available at the CSIRO Science

Education Centre in Highett. Most programs are also offered as an "incursion program" on school grounds (CSIRO, 2012). These programs typically consist of a presenter-led introduction and demonstrations; an explanation of each of the hands-on activities; half an hour or more for students to do the hands-on activities; a brief wrap-up that summarises the most important concepts from the program; and another demonstration or two. The activities are designed to help students learn about important science concepts while keeping them interested and engaged.

Strengths of CSIRO's Educational Programs

Promotion of Scientific Literacy through Curriculum Relevance

All CSIRO Science education programs are carefully designed to touch upon key parts of the Australian national science curriculum. The Australian Curriculum, Assessment and Reporting Authority (ACARA) has recently designed and implemented a standardised national science curriculum that "provides opportunities for students to develop an understanding of important science concepts and processes, the practices used to develop scientific knowledge, of science's contribution to our culture and society, and its applications in our lives" (Australian Curriculum Assessment and Reporting Authority, 2012). This curriculum provides the structure for all Australian schools to develop their science programs by calling for these programs to cover ideas appropriate for the given year, within these three main topics: science understanding, science inquiry skills, and science as a human endeavour. With the increase in the use of technology in our daily lives, having some understanding of and appreciation for the science behind the technology is becoming increasingly important.

The guidelines set by ACARA state that science understanding becomes apparent when a person demonstrates the ability to select and utilise scientific information to explain and predict phenomena (ACARA, 2012). The Australian science curriculum breaks science understanding into four disciplines that cover the major areas within science: the biological, chemical, earth and space, and physical sciences. An understanding of the major concepts, theories, models and facts within each of these areas gives students basic scientific knowledge that contributes to scientific literacy.

Being literate in the sciences also consists of being able to gain new knowledge through curiosity about everyday events. Scientific inquiry is the way in which new scientific knowledge arises. The skills focused on in this sub-area of the Australian science curriculum include "identifying and posing questions; planning, conducting and reflecting on investigations; processing, analysing, and interpreting evidence; and communicating findings" (ACARA, 2012). This area also enhances the ability of students to assess the validity of scientific claims, investigate ideas, solve problems, draw valid conclusions and develop evidence-based arguments. Students who demonstrate competency in this area likely enjoy the scientific process and may further pursue the sciences.

Another part of scientific literacy is the capacity to identify scientific problems that affect local and national policy decisions and to express well informed opinions on those issues and other scientific issues affecting society. The "science as a human endeavour" topic in the Australian education science curriculum seeks to highlight scientific development as a distinctive means of problem solving and consists of two sub strands. 'Nature and development of science' seeks to develop recognition of the fact that science and scientific knowledge are unique and that current knowledge is the result of the effort of numerous people over time. 'Use and influence of science' is concerned with how the application of scientific knowledge affects peoples' lives and the implications of the relationship between science and society, especially how scientific literacy (or lack thereof) can impact decisions and actions on various levels.

These three topics taken together provide basic scientific literacy[1] skills necessary for a person to succeed in a scientific society. Each CSIRO program incorporates aspects of these three main topics and presents them in a coherent and engaging manner, with students utilising scientific inquiry skills to supplement their understanding of a given science topic. The students are then able to consider how the experiments they've performed relate to their lives and society as a whole. These curriculum links are arguably the most important part of any CSIRO program as curriculum relevance is what encourages teachers to book the programs. On the other

---

[1] The Board on Science Education (BOSE) offers a comprehensive definition of scientific literacy in the National Science Education Standards: "Scientific literacy means that a person can ask, find, or determine answers to questions derived from curiosity about everyday experiences. It means that a person has the ability to describe, explain, and predict natural phenomena. Scientific literacy entails being able to read with understanding articles about science in the popular press and to engage in social conversation about the validity of the conclusions. Scientific literacy implies that a person can identify scientific issues underlying national and local decisions and express positions that are scientifically and technologically informed. A literate citizen should be able to evaluate the quality of scientific information on the basis of its source and the methods used to generate it. Scientific literacy also implies the capacity to pose and evaluate arguments based on evidence and to apply conclusions from such arguments appropriately." (*National Science Education Standards*, 1996, p.22)

hand, if the programs don't keep students engaged and interested in what they are doing, they are unlikely to learn anything from the experience. Fortunately, with CSIRO, this is not the case, and the level of student engagement is a definite strength of all CSIRO science education programs.

Student engagement through hands-on activities

CSIRO's programs provide non-formal education, often through the use of hands-on activities, designed to supplement the Australian curriculum and inspire students' interest and engagement in science. All of CSIRO's science education programs consist of several activities for students to perform in groups during the hands-on portion of the program. These activities can vary, from looking through a book of pictures demonstrating natural disasters, to spinning gyro rings to learn about energy transfer, to mixing chemicals to produce a slime-like substance, to calculating the density of various volcanic rocks. Some activities in the programs are more popular amongst students than others, but all are engaging. According to science educator Donna Satterthwait (2010), hands-on science programs are effective because they include the three significant factors that have been identified as contributing to student learning: peer interaction through cooperative learning, object mediated learning, and embodiment (how humans interact with and make sense of their surroundings). By incorporating these three factors into their programs, CSIRO captures students' attention and motivates them to perform the experiments. Research also suggests that CSIRO programs are structured in an optimal way for generating interest. David Palmer, researcher in the School of Education at the University of Newcastle in Australia, has reported that students situational interest during a science lesson was highest during the experimental phase of the lesson, and second highest during the teacher led demonstration (Palmer, 2009, p.153). CSIRO programs incorporate both educator-led demonstrations and numerous hands-on activities for students, generating high amounts of situational, or short term, interest in doing science. Palmer suggests that additional research has shown that "multiple experiences of situational interest" can lead into longer-term interest by students (Palmer, 2009, p.148). By generating high amounts of situational interest, CSIRO sets students up to become more interested in science, even if only for short amounts of time.

Weaknesses of CSIRO's Educational Programs

Lack of Systematic Evaluation of Program Strengths and Weaknesses

CSIRO's educational programs are constantly changing and evolving, with new programs designed, out-dated programs retired, and activities added, removed, or changed to keep programs interesting for students and relevant to teachers' changing curriculums. This means that the staff members at CSIRO Education frequently make decisions on which activities to add, remove, or modify. Unfortunately, they currently do not have a systematic way of evaluating their educational programs and the activities that make them up. This leads to decisions about program content and about which hands-on activities will best teach material up to the instinct and intuition of educators at CSIRO. The experience that the CSIRO educators have makes their initial feelings about activity success fairly well informed, but variations in experiences and personal preferences mean that consistency across educators and decisions may be lacking. The creation of a standardised activity evaluation rubric that can be applied to all hands-on activities in CSIRO science education programs would greatly benefit educators as they seek to update and design programs to be as successful as possible.

Lack of focus on contemporary research

CSIRO programs are regularly modified to include new activities and remove ones that haven't been working as anticipated. However, the additions typically rely on tried and tested technologies, rather than what is currently being researched. CSIRO programs have a reputation for including specialised equipment or technologies that teachers and students typically do not have access to in a school setting. Such technologies allow the programs to demonstrate real-world applications of the scientific concepts that students are learning in the formal education setting and excite students about the possibilities of scientific research. CSIRO has a keen interest in promoting scientific discoveries and having the public recognise the connection between new technology and its applications in their daily lives. CSIRO has recognised the lack of current research topics in their programs and is in the process of introducing activities with a focus on contemporary research. Two programs, *Energy: Sources and Uses* for primary school students and *Polymers and Nano-chemistry* for year 11 VCE chemistry students, are prime candidates for the introduction of an activity relating to flexible solar cells that researchers from CSIRO's Materials Science and Engineering division are currently developing.

Flexible Solar Cell Technology

The topic of solar energy lends itself well to CSIRO's educational program goals and has been incorporated into programs before. Solar energy has been used on a small scale for decades, but continuing research into increased efficiency and cost-effectiveness suggests that this technology has great potential to become a leading sustainable energy source for the future. The earliest discoveries of solar energy conversion date back to the late 19th century, but only after the oil embargo of 1973 did a consistent commercial market for photovoltaic cells develop (Fraas & Partain, 2010). Even so, as of 2010, solar energy still accounted for less than one tenth of one per cent of the world's energy usage (Xue, 2010). Materials scientist, Jiangeng Xue, comments that covering only 0.1% of the earth's land with 10% efficient solar cells could singlehandedly satisfy the world's energy demands, but explains that high upfront costs of solar electricity production continue to hinder the spread of the technology (2010).

Despite these economic barriers, solar energy use is spreading quickly, producing approximately 35-40% more electricity globally each year (Xue, 2010). In remote population centres around the globe, particularly in the developing world, solar cells are used to power home lights and water pumps that would otherwise have no means of obtaining electricity (Fraas & Partain, 2010; Watkins, 2012). Seeking solutions to the costliness of traditional solar panels, scientists at CSIRO and all over the world are turning to the rapidly growing subject of organic photovoltaic (OPV) cells, which tend to have much lower material costs than conventional silicon-based cells. According to the solar cell efficiency tables of Green and colleagues', *Progress in Photovoltaics,* the efficiencies of OPV's (which reached 10% in 2012), continue to lag behind those of silicon-based solar cells (boasting efficiencies of up to 25%) (2012). Despite this lack in efficiency, the significantly lower costs of OPVs give them the potential to generate a great deal more energy for the same price.

Along with other members of the Victorian Solar Cell Consortium research group (VICOSC), CSIRO researchers have been working on the development of thin-film, printable organic photovoltaic cells. Scott Watkins, stream leader for organic photovoltaics at CSIRO, explains that thin and lightweight OPVs, with active layers only 100nm thick, “can be made using the same process used to print food labels or Australia’s polymer banknotes” (2012). This potential for cheap mass-production, he argues, in conjunction with the flexibility of the material, could open up solar energy to the developing world, and make it affordable for many more consumers in the developed world. The goal of the VICOSC partnership in the upcoming few years is the inexpensive printing of these solar cells directly onto materials such as plastic and steel, allowing for the conversion of an entire roof or other surface into a solar panel (Watkins, 2012).

The versatility in use of solar energy, combined with its potential to cleanly harness large amounts of energy that would otherwise go to waste, make it an excellent topic for CSIRO’s non-formal education programs. Hands-on activities involving flexible solar cells have the capacity to further develop a variety of scientific literacy skills in students. The easy scalability of photovoltaic cell technology allows students to generate electricity themselves in a hands-on activity, giving them a sense of accomplishment and a deeper personal understanding of science. The chance to explore and experiment with the unique properties of flexible solar cells encourages students to further their science inquiry skills by testing their ideas and drawing their own conclusions. Lastly, the chance to get a sneak peek at on-going, world-changing research through experimentation with flexible solar cell technology can open students’ minds to the human and societal applications of science. This imparts in students a greater sense of purpose behind their studies and might just inspire children to become Australia’s next generation of scientists.

3 Methodology

The goal of this project was to assist the Commonwealth Scientific and Industrial Research Organisation (CSIRO) in providing engaging informal science education programs for students across Victoria by meeting the following objectives:

- Develop and test a rubric for evaluating the effectiveness of existing and proposed hands-on activities for programs delivered by CSIRO to students in foundation year through year 12 VCE science classes.
- Utilise the rubric to inform design recommendations for two hands-on activities involving flexible solar cell technology to be incorporated into CSIRO programs in the future.

3.1 Educational Activity Research and Criteria Development

To assist CSIRO educators in assessing the effectiveness of their hands-on activities, we conducted extensive research into the elements that contribute to an activity's success. We used this information to develop, test, and refine a clearly defined rubric with which all CSIRO hands-on activities can be evaluated. Based on a weighted set of important evaluation criteria, this rubric will allow educators to better analyse the strengths and weaknesses of each activity in their programs and to better focus their design of new activities on key objectives that these activities should strive to meet.

3.1.1 Analysis of Education Techniques

Creating a set of criteria with which to effectively and practically measure hands-on educational activities required knowledge of how such activities functioned and performed in both a general sense and specifically when designed and lead by CSIRO staff. We educated ourselves in a variety of ways: we studied previous research on non-formal education techniques, interviewed CSIRO staff about important characteristics of the organisation's activities, and observed relevant programs to study design features and learn which characteristics could and could not be easily observed. In addition, much of this research allowed us to create a list of educational and design tools to apply to our activity design recommendations.

Literature Review and Science Museum Excursion

The first step in developing our rubric criteria and activity design recommendations involved research into similar activities and educational methods that help such programs teach students concepts in engaging and effective ways. We also began to look for features and goals consistently represented by non-formal education activities that could be used as a comparison metric. To accomplish this, we first conducted a literature review, examining both writing on general education techniques and descriptions of existing hands-on science education programs from CSIRO and elsewhere. Using educational article databases such as the Educational Resources Information Centre and the National Science Teachers Association, we examined scholarly articles for general educational strategies for increasing student interest and learning, which were used to determine different types and levels of effectiveness to watch for. We also examined the Australian science curriculum in detail, noting the level of understanding that CSIRO programs should supplement for each age level, and identifying key concepts that could be reinforced by activities involving flexible solar cells. Additionally, we studied the teacher notes for several of CSIRO's educational programs, getting a first look at the nature of the activities that these programs include. These literature reviews prepared us for the makeup of educational activities, allowing us to look for the details relevant to design elements as they applied to our criteria.

We also visited the EcoTarium, a children's science and nature museum in Worcester, Massachusetts, observing how children interacted with hands-on science exhibits. As part of our observations, we noted which types of exhibits seemed to be favourites among students and were able to hold their interest for significant lengths of time. We also examined the exhibits ourselves, focusing on the techniques that were used to convey science concepts to visitors to better inform ourselves of the nature and effectiveness of these techniques. Based on these observations, we developed a list of common features among engaging and educational exhibits that could be applied to our activity design recommendations and the categorisation and scoring of the rubric criteria.

From the information obtained through all of these methods, we then developed a list of elements that can contribute to the effectiveness of educational activities. The entirety of this list can be seen in Appendix A. In the creation of this list, we focused our research into the following three primary categories:

communication of educational objectives as they relate to the Australian curriculum, wowing and exciting students about science, and logistics of implementation. These three categories were eventually expanded into the five main criteria on the evaluation rubric. For the first category, we examined the Australian curriculum goals for various age levels to assess what types of important information students might be able to learn from non-formal education programs. From the CSIRO program teacher notes that we examined, we also saw examples of the kinds of curriculum links that these programs contain. The second category consisted of elements of existing science education activities that we have found to keep students interested and engaged, helping us to get a sense of which elements to look for when judging student interaction with activities. The third and final category took into account the physical constraints and details which must be considered in the development of CSIRO activities, allowing us to begin generating design standards activities should be expected to meet. This list of educational objectives and design criteria formed the basis for the rubric that we later developed to evaluate the effectiveness of hands-on activities with regards to these three important categories.

Interviews with CSIRO Education Staff

To further inform our construction of the educational activity evaluation rubric, we gathered more information about the specifics of what CSIRO educators desire to convey through their activities, and about the common traits exhibited by those that are most successful. To learn the most information in a short period of time about what makes a CSIRO activity a success, the ideal interview candidates were CSIRO Education employees themselves. At the CSIRO Science Education Centre in Highett, experienced staff members were easily accessible, and their knowledge and experience greatly benefited the project team's efforts to develop successful rubric criteria. These interviews allowed us to successfully accomplish our goal of understanding the important criteria that an informative and exciting hands-on activity should follow.

The interviews took place during business hours at the Highett CSIRO Science Education Centre, in offices and laboratories. They were conducted in an unstructured format, in which we talked with the educators individually about what types of educational activities tend to be successful at engaging students and reinforcing educational concepts. We leveraged the interviewees' extensive experience with CSIRO's programs, asking about the common elements of the successful and unsuccessful activities that they have encountered in the past. We also discussed some of our preliminary activity design ideas and asked them about anything in particular that they would like to see in a flexible solar cell activity.

Once these interviews were completed, the responses were organised and examined for similarities. The data organisation process involved listing each potential factor that the interviewees felt could contribute to the success of an activity and each suggestion for using the flexible solar cell technology. Where multiple educators raised the same or similar points, a tally was kept to indicate how commonly-mentioned each factor was. The tables used to organise these data can be found in Appendices B and C. This weighted list was then compared to our initial lists of design elements and rubric criteria, which were subsequently modified and/or added to in order to incorporate the information gleaned from the interview sessions.

Initial Program Observation

Though the interviews with CSIRO educators gave us a great deal of useful information about CSIRO's hands-on science education programs, it was difficult to put this information into context without experiencing such programs first-hand. Thus, over a three-week period, we observed programs at the education centre in Highett and also travelled with educators to schools to observe several different educational programs for a variety of age groups. The purpose of these observation sessions was threefold. Firstly, watching these programs gave us a much better idea of how they run and brought to light some logistical considerations that we needed to be aware of while designing our evaluation rubric and developing activity recommendations. Secondly, observing students interacting with many different activities allowed us to note the kinds of activities that kept students most interested and engaged. Finally, these excursions offered us a chance to practice our observation skills so that inexperience would not interfere with our ability to effectively test the usability of our rubric later in the data collection process.

3.1.2 Development of Activity Evaluation Rubric

Based on the research we described previously, we developed a list of criteria that we found to go into a successful activity design. These criteria consisted of five main categories, each of which had a number of measurable sub-criteria. The main categories were:

- Educational Value
- Logistical Feasibility
- Initial Draw
- Satisfaction
- Ease of completion

We then turned these criteria into a rubric that could be used to evaluate how well CSIRO activities met these main criteria.

Analysis of Existing Activities based on the Rubric

In order to confirm that our rubric accurately represented the components of hands-on activities, we evaluated existing CSIRO activities through a review of the Australian science curriculum, observations of students interacting with activities, focus group discussions with students, and individual conversations with teachers. Together, these various methods of data collection allowed us to give informed scores for all criteria on the evaluation rubric, by using different combinations of review, direct observation, opinion gathering, and interviewing to ascertain each piece of necessary information.

Analysis of Activity Curriculum Relevance

The first step in evaluating activities was to determine how relevant each activity was to teaching a concept in the curriculum and how clearly it communicated that concept. To do this, we conducted a review of the Australian national science curriculum, the notes CSIRO provides to teachers who book programs, and any student worksheets related to the programs, looking for concepts and overarching themes displayed by all sources. Activities that clearly communicated important concepts received higher scores than activities that related to the program and communicated concepts clearly, but were not explicitly covered by the curriculum or activities that did not communicate concepts clearly, but were relevant to the curriculum.

Observation of Student Interaction with Activities

Observing students as they performed activities was crucial to data collection, as it was the only way to measure how well each activity met some of the rubric criteria in practice. Observations also allowed us to judge the reactions of students without requiring them to self-analyse or respond to instructors, as these direct interactions could potentially change their normal behaviour. Each of us discreetly followed a different group of students during the activity phase of the CSIRO programs we observed, recording how the selected group reacted to each activity. Initial interest in each activity, the ease with which the activity could be completed, and the satisfaction obtained from the activity, were all measured through observation of more specific sub-criteria. By studying multiple groups, we were able to account for differences in students and differences in scoring amongst observing team members. The results of these observations were crucial to much of the rubric scoring process. The observation sheet that was used for this data collection can be found in Appendix D.

Focus Group Discussions with Students

At the end of some of the program sessions that we observed, we were able to conduct focus group discussions with the students to hear from them directly about their experience. Focus groups provided valuable information in terms of determining what the students as a whole thought of the activities in a program and allowed us to identify whether students were obtaining an understanding of the science concepts that each activity was designed to teach. Since the discussions were held as a group, it allowed the younger students to easily participate and gave the older students more opportunities to share their opinions. Christensen and colleagues (2005) described the use of "kiddie focus groups" as an effective, creative way of gathering data to address the issue of multiple goals in non-formal education settings by acquiring feedback directly from those affected by the program to determine the most valued goals. These focus groups also gave an opportunity for student collaboration that may have yielded clearer ideas about each activity than individual interviews with the students would have. In order to not disrupt the flow of the programs, we had the presenter ask our series of questions to the class during the program wrap-up. The questions that the presenter asked the class were intended to assess the satisfaction of the students with each activity, as well as to gain insights into what they had learned. We first asked students for their favourite activities, prompting them to explain what they liked about them and what they learned from doing them. To ensure that we obtained useful data for a wide range of activities, we also asked the students specific questions about some of the other activities, aiming to once more judge their satisfaction with and understanding of these activities. See Appendix E for the list of questions we asked each class.

Conversations with Teachers

As teachers know their students and those students' personalities better than we could through short observations, we decided to talk to them during the programs' hands-on portion to gain some of their insights into how students were acting and interacting with activities as well as their motivations behind booking the program. If a teacher had booked the program as an introduction to a particular unit, we asked questions about some of the specific activities in the program and if they thought they could use them as something to refer back to during a future lesson. We also asked about how the program related to the curriculum they were currently covering. This information provided good insights into the educational value of the activities and of the program as well as a context for how teachers were utilising the programs. The general conversation guidelines can be seen in Appendix F.

Weighting the Evaluation Criteria Based on Importance

Once we had determined the evaluation criteria that make up our rubric, we set out to ascertain the relative importance of each of them. This allowed the rubric to weight and combine the scores given to an activity for each criterion into an overall score that gives more emphasis to those criteria that are deemed more important to the success of the activity.

To inform this weighting process, we surveyed about a dozen CSIRO educators to obtain their input on the relative importance of each of the criteria that we had developed. The survey they were given first listed each of our five main criteria categories and prompted educators to rate the importance of each of them from 1 to 10, where 1 indicated that the criterion was not at all important and 10 meant that the criterion was critically important to the success of an educational activity. Next, each category was broken down into its individual sub-criteria, and educators were asked to repeat the process for each category, rating the sub-criteria on how important they are to achieving the main goal of the category. The full criteria importance survey can be found in Appendix G.

Based on a combination of the results of this survey and our personal experience from using the rubric to evaluate several educational programs, we ranked each of the criteria in order of importance. Using pairwise comparisons, we first compared the relative importance of the five main categories in contributing to the overall success of an activity. Once the categories had been ranked in order of importance, we then assigned weights to each of them. Finally, we repeated the process for each category, comparing and weighting the sub-criteria within a category with respect to each other. These numbers, when combined with the weights determined for the overall categories, produced the final weighting of each sub-criterion with regards to its overall importance. The pairwise comparison charts that we used for this process can be seen in Appendix I.

Evaluation of Criteria and Rubric Effectiveness

After using the rubric to evaluate and score each activity in the five educational programs that we observed, we analysed the results to assess the effectiveness of our rubric. First, we assessed the validity of the rubric scores in accurately measuring the important qualities of the activities by comparing the overall and category scores earned by each activity to our qualitative observations on the activities' relative effectiveness. Secondly, we examined the reproducibility of the rubric scores across multiple evaluators by giving our rubric to a handful of CSIRO educators to evaluate the same set of activities and comparing the results. Finally, we assessed the universality with which the rubric could be applied to educational activities for all age groups and topics by comparing the results of our analyses of five different programs and seeking to explain any trends that we identified.

3.2 Flexible Solar Cell Activity Design

Informed by our research into and analysis of successful CSIRO hands-on activities, we were able use a structured design process to thoughtfully develop design recommendations for new activities based on flexible solar cell technology, to best meet the identified criteria of activity success. The development of these recommendations began with research into possible applications of the solar cells, as well as analysis of the common elements of successful educational activities that we have identified. Building from this knowledge, activity ideas were brainstormed, organised, and described in more detail. Next, based on further analysis of existing CSIRO programs, these ideas were combined with design concerns emphasised by our activity rubric. Finally, these concerns were refined into concrete design requirements and suggestions for CSIRO for the two activities, addressing strengths and weaknesses of any activity involving flexible solar cells when intended for specific age groups.

3.2.1 Research into Educational Applications of Flexible Solar Cell Technology

The first step in the activity design process was obtaining the preliminary information necessary to develop informed design ideas. In achieving this goal, we relied upon many of the same data and sources that we used in the development of general activity criteria, but with a new analytical focus on how to apply this information toward creating requirements for a currently non-existent activity. From the initial literature review, we explored specific techniques by which teachers and other educators keep students interested in and engaged with their lessons, looking for ways that CSIRO could leverage these techniques in flexible solar cell activities. By examining the teacher notes for the two CSIRO programs into which these activities would be added, we were able to get an initial idea of the types of activities that these programs contain, to ensure that our recommendations would account for strengths and weaknesses in the respective programs. From the interviews with CSIRO educators, we learned about some of the specific elements that have made activities successful with children in the past experience of the educators. These data were backed up with the initial observations of various CSIRO educational programs, in which we were able to see children interact with activities with varying degrees of enthusiasm and understanding. The analysis of the data allowed us to draw upon proven examples in the development of recommendations for potentially successful activity designs. Our design recommendations greatly benefited from this research, but they also required a working knowledge of the solar cells themselves, in order to account for their specific strengths and weaknesses.

Meeting with CISRO Materials Science Researchers

The simplest method to learn about the capabilities of the flexible solar cells proposed for use was to contact the source directly. After completing background research on organic photovoltaic cells in general, we travelled to the CSIRO complex in Clayton to speak with Scott Watkins, leader of CSIRO's organic photovoltaic research team. At this meeting, we were shown how to use the flexible solar cells and examples of prototype applications. We asked several questions about the solar cell technology to help us better understand what can be done with them.

We discussed with him the various polymers and other chemicals that make up the solar cells, learning about the unusual properties of each and about the reasons for their use. We also learned about the physical properties of the technology, focusing on the factors that affect the output of the panels, such as the shape or the temperature. Finally, we learned about the amount of power that the panels can produce under different lighting conditions and potential ideas for things he'd like to see in an activity making use of this technology.

The answers provided helped us determine both the physical limitations of what we could do with these cells, as well as curriculum-relevant knowledge of the underlying chemistry that could be integrated into the VCE Chemistry activity. After the meeting, Scott Watkins directed us to some additional papers and articles on the topic, so that we could further research the technology that the researchers were working on. He also gave us a small number of weak or non-functional solar cells, with which we could test their behaviour and prototype some of our design ideas.

Experimentation with Prototype Flexible Solar Cells

Though talking to the researchers responsible for the flexible solar cells provided us with a lot of useful information on the technology, in order to better understand how to apply this information, we took some time to experiment with the sample solar cells. Though the solar cells that we were supplied with for prototyping purposes were not efficient enough to power anything substantial, we measured the current and voltage provided by each sample under various lighting conditions and configurations. We also compared these results to the power outputs of traditional, silicon-based solar cells. We experimented with the flexibility of the already-non-functional solar cells, determining how far they could be bent without creating too much strain. This process allowed us to see the capabilities of the solar cells first-hand so that we could develop more realistic recommendations for hands-on activities.

3.2.2 Development of Flexible Solar Cell Activity Recommendations

Once the necessary background information on solar cells had been gathered, the next step was to evaluate a multitude of activities against a set number of important criteria, enabling us to obtain concrete information on which activities were the most exciting, interesting, educational, or relevant to students. To accomplish this, we observed a number of sessions of both the target educational programs and of other CSIRO programs of similar formats. We also administered focus group discussions for the students at these programs to get their feedback on the various activities that made up the programs. Once all of this data had been compiled, we were able to analyse it to look for common elements among activities that scored highly and among those that scored poorly in each of the criteria. When possible, we modified our design suggestions to include more of these positive elements, while noting the elements that seemed to have a negative effect. Once this analysis had been performed, the list was examined again to remove any suggestions that no longer seemed at all realistic.

The final step in our design process was to analyse our design recommendations critically with respect to the design criteria that we developed for our activity evaluation rubric. For each criterion, a final list of related design recommendations was made, based on previously established lists. This helped to organise thoughts and point out gaps in our recommendations where they failed to address specific aspects of our rubric criteria. These final recommendations were then combined into a unified list of design suggestions, in which each of the design criteria from our rubric was addressed by one or more specific recommendations.

4.0 Results and Discussion

4.1 Rubric Design

Our final rubric assigns scores to activities based on several criteria, which are broken down into sub-criteria. Each sub-criterion is scored on a scale from 0 to 5, based on descriptions for what constitutes each possible score for that criterion. The full evaluation rubric, containing all of these descriptions, can be found in Appendix H. Each sub-criterion was also given a weight, and a final score for each activity was calculated by multiplying each sub-criterion's score by its weight, and then summing the results of those multiplications for all of the sub-criteria together. The scoring sheet used to record and calculate these scores and the weighting sheet used to change the weights of the criteria can be found in Appendices J and K, respectively.

4.1.1 Rubric Creation

Our literature review and educator interviews led us to identify five main categories of criteria that a good educational program should have. These five categories are:

1. Educational value
2. Logistical feasibility
3. Initial draw
4. Student satisfaction
5. Ease of completion

Through our educator survey, program and activity observations, student discussions and teacher conversations we were able to validate the importance of each of these criteria and weight them relative to each other.

Educational Value

One of the main reasons teachers book CSIRO programs is because they have a high educational value (Carney, Hyman, Mello, & Snieckus, 2011). As such, measuring the educational value of each activity in a program gave a good indication of where the program's educational strengths and weaknesses lie. Through our literature review and educator interviews, we learned that important aspects of educational value are *curriculum relevance, encouraging student thought, and* *potential for teacher extension*, which deals with the ability of the teacher to build upon the covered concepts. Taken together, these three aspects can be thought of as how well a given activity teaches science concepts, an important result in regards to program outcomes.

The curriculum relevance of each activity was assessed by comparing the Australian science curriculum against CSIRO's teacher notes and student worksheets, looking for concepts each activity covered in the curriculum. For example, the Ball Smasher activity in the Energy: Sources and Uses program demonstrates the transformation of gravitational potential energy into kinetic energy, which is then transformed into sound and heat. This activity also demonstrates four different types of energy. These are curriculum-relevant points that CSIRO has used to create an interesting activity while teaching students about science. Activities with a high amount of relevance to the national science curriculum scored highly, while activities that were interesting and related to the program, but not the curriculum, scored lower.

An activity that encourages students to think through what they are doing and why they are doing it will be a better activity than one where students mindlessly press buttons. This was mentioned in a number of our educator interviews. How well each activity encouraged students to think through the process of doing the activity was measured through a combination of student observations and responses to the group discussion questions. Through these methods we were able to make reasonable educated guesses as to which activities students actually thought through. Activities that students did not need to think about at all received low scores on our rubric, while activities that students had to do some thinking about scored higher.

The ability of a teacher to build upon concepts covered by an activity is a very important aspect of educational activities that should not be overlooked during the design process. Some teachers who book CSIRO programs use them as a fun introduction to a new unit. These teachers then have the ability to build off the concepts presented by the activities in the program while referring back to a time the students enjoyed and can likely easily recall. While conversing with teachers, we were able to ask about which activities they thought they might refer to in a future classroom setting. These activities generally earned higher rubric scores than those that were not mentioned during discussions.

Logistical Feasibility

From the beginning of our research, we knew that any evaluation of an educational activity would not be complete without taking into account logistical considerations. The logistical feasibility of an activity measures how easily it can be implemented and run as part of a program. This category of criteria is different from most of the other criteria that are used to evaluate the effectiveness of an activity. While other criteria, such as educational value and satisfaction, can have a great impact on an activity's success, the logistical feasibility of an activity determines whether the activity can be used at all. The feasibility of an activity can be broken down into a variety of logistical factors. To assess the logistical feasibility of each activity, we analysed its *cost, durability, ease of repair, ease of setup and takedown, and ease of transport,* as well as any *health, safety or environmental* concerns associated with the activity.

Though difficult to measure from observation and often overlooked, the cost of an activity is a critical factor in its feasibility. Few of the educators that we interviewed mentioned cost as a factor relevant to the success of an activity, as it seems to go without saying, but one educator specifically mentioned the importance of ensuring that an activity is inexpensive enough, not only to create, but also to maintain and replace consumables. The importance of cost is abundantly clear when examining the overall budget for an educational program. If an activity were to cost more than the allotted program budget, then it is clearly not only unfavourable, but impossible to run without additional outside funding. Furthermore, even if such an activity does not exceed the budget on its own, but uses up a large portion of the total program money, it can be very difficult to fit into the program, as all of the other activities must be paid for from the same budget. In rating the cost of activities, we assigned low scores to activities that use up significant portions of the budget or require additional funding to be possible, average scores for those that cost a reasonable share of the total budget, and high scores for those whose costs are negligible.

As we gathered information to inform the construction of our rubric, no logistical concern was emphasised more than durability. During our visit to the EcoTarium, we noticed that student interaction with exhibits could easily become destructive. When presented with a giant chessboard on which to play the classic strategy game, many students began to throw and kick the giant chess pieces across the room. Fortunately, the pieces involved were durable, and did not break or become damaged by the abuse. From this experience alone, it was immediately apparent that any activity used with children needs to be durable to be successful. In our interviews with CSIRO staff members, the durability of an activity was mentioned as an important criterion by more interviewees than any other logistical factor. On our evaluation rubric, activities that are damaged every few sessions receive low scores for durability, while those that almost never need to be repaired or replaced receive high scores.

A closely related and also important factor in the success of an activity is the ease with which it can be repaired or replaced. This criterion measures how much time, effort, and equipment go into replacing or repairing activities that have become damaged. Even if an activity is quite durable and rarely breaks, it can still be problematic if it cannot be repaired and will take weeks to replace. Conversely, even fragile activities that break frequently can be successful if they can be quickly repaired mid-program. In our educator interviews, while most of the interviewees focused on ensuring that activities are tough to damage in the first place, one specifically mentioned the benefit of those that can be fixed easily and quickly. Our initial program observations demonstrated this importance as well. In one primary-school program that we observed, an activity was accidentally dropped and broken by a student. The activity could not be repaired on-site and was removed from the program for the remainder of the day, meaning a majority of students missed out on it, the consequence of not having a quickly repairable or replaceable activity. For the ease of repair criterion on our rubric, activities that can be repaired during the program earn high scores, activities that must be fixed in between sessions or back at the CSIRO facility receive lower scores, and those that cannot even be repaired by CSIRO staff and take excessive amounts of time to replace or have repaired earn the lowest scores.

Another logistical factor that contributes to the feasibility of an activity is the ease with which an activity can be set up and taken down. While observing a variety of CSIRO educational programs during our initial observation period, we learned that educators have a lot to accomplish to set up a program in the allotted hour before it begins. A single presenter runs each program and must assemble as many as 20 activities from parts spread across three large boxes. For some programs, presenters must move quickly and efficiently to set up each activity in as little time as possible.

After the program, the same process must be done in reverse, as the presenter must quickly clean up and put away every activity in a limited amount of time. Do to the hurried nature of this process, it is clear that activities with long setup or takedown times could take up too much of the precious time allotted for the setup and takedown of the program as a whole, and thus make the program unfeasible. The results of our educator interviews supported this idea, as multiple educators commented that a good activity should take a reasonable amount of time to put together. The ease of setup/takedown criterion on our rubric rates activities based on the time that they take to prepare and put away. Activities that take more than 10 minutes to setup or takedown receive the lowest possible scores for this criterion, while those that can be set up and taken down in less than two minutes receive much higher scores.

While an activity must be relatively easy to set up and take down at a program site, it should also be easy to transport to and from schools. The smaller and lighter that an activity's components are, the easier they are to fit into boxes and the easier these boxes are to lift. When activities cannot fit in the boxes where most of the program material is kept, the additional things that need to be carried may force the presenter to take multiple trips to bring everything inside, thus cutting into the time needed to set up the program. On our evaluation rubric, an activity earns the highest score in the ease of transport criterion if it weighs less than 500 grams and/or takes up less than ten per cent of the space in one of the large storage containers CSIRO uses to store activities in cars, which are designed to fit on trollies and fit through doorways. Activities that weigh more than ten kilograms or are too large to fit into one of the storage containers earn the lowest possible score.

The final criterion that we found to contribute to the logistical feasibility of an activity consists of the health, safety, and environment concerns of the activity. Safety is always the top priority in any setting, and almost all hands-on activities that CSIRO runs have low risk of causing injury or harm. Some activities, though, require warnings or careful watch by the presenter to prevent students from improperly using the equipment in a way that could be dangerous. Other activities involve chemicals or other materials that must be handled carefully and disposed of properly. Students must be clearly warned how to dispose of these materials so as to avoid potential environmental risks. With respect to the health, safety, and environment criterion in our evaluation rubric, high scores are awarded to activities with virtually no risk, while average scores are given to activities in which students must be given some kind of safety warning and low scores given for activities that can actually be dangerous to students if the presenter does not take extreme care and precaution.

Initial Draw

Initial draw represents elements that cause a student to become interested in performing an activity. These can include anything from actively drawing a student's attention with light or sound to having an interesting appearance or an interesting start to the activity. An activity with a high initial draw is able to encourage students to approach the activity and to interest students enough upon arrival that they will actively want to perform the activity. Our literature review indicated the importance of getting a student initially interested for motivational purposes, and interviews with CSIRO staff made it clear that a high initial draw was greatly sought after when designing activities. In these interviews, educators brought up appearances and various forms of conceptually exciting designs, because they were usually effective at getting students to begin the activity with an open mind. Many factors end up going into the initial draw of an activity, but measuring any of them on their own as sub-criteria obscures the end result and adds unnecessary complication to the evaluation process. Thus, initial draw was scored on the rubric as a single criterion; with instructions in the rubric encouraging anyone using it to record their own notes to justify their decision based on those numerous factors. Scoring is scaled with the likelihood of drawing interest; discouraging students on average scored zero points, a neutral activity scored minimal points, and points increased with the number of students likely to be interested.

Satisfaction

Satisfaction represents design elements that lead to a student enjoying or appreciating an activity. During background research, the enjoyment, memorability, or meaningfulness of an activity was brought up repeatedly by multiple sources. Interviews with CSIRO educators confirmed that a good activity should be enjoyable. The need to keep activities interesting and fun for students was one of the topics that was directly mentioned by educators most frequently, and every educator that was interviewed spoke about this aspect indirectly. They revealed that the enjoyable aspects help to make both the program and science as a whole more interesting, and increases the likelihood that a student will be willing to complete the entire activity. Rather than simply have a rubric category that was "did the student find this fun", the criteria was broken into four sub-criteria to capture likely sources of enjoyment and meaningfulness: *Novel completion process, relevance of the activity to the student's life, unexpected/surprising results, and degree of student interaction*.

Novel completion process covered whether the activity is new to students, and as a result is exciting, unpredictable, and interesting. Interview notes repeatedly brought up how doing a new or unknown task tends to be inherently interesting, and both students and teachers enjoy the chance to access educational equipment or toys that are normally not available in schools, often due to either budget concerns or lack of required training. The scoring for the novel completion criterion was based on how much of the activity was similar to tasks or actions that students perform in their daily lives. With this system, an activity consisting of a straightforward everyday action earns no points, variations on everyday actions earn more points as the variations grow larger, and activities completely different from anything an average student would be likely to do earn full points.

Relevance and familiarity measure how likely students are to be able to link an aspect of the activity to something that they have encountered in their daily lives. This helps to increase student interest by explaining or expanding upon elements of the world around them and helping the students to understand or remember the information through an external reference point. Background research indicated that the applicability of a task or lesson helped in the process of education and enjoyment, and the majority of CSIRO educators stated similar opinions. While this criterion could be misinterpreted as the opposite of novel completion, the two are not mutually exclusive, as a completely new activity can still apply a principle of science relevant to something in a student's life. The scoring for relevance was based on how familiar and relevant the program was to the students' lives, with no, little, or tangential familiarity or relevance scoring few points and relevance to a core aspect of the students' lives scoring very well.

The unexpected results criterion covers activities with an unanticipated ending. Background research indicated that these kinds of activities were more memorable for roughly the same reason as novel completion processes. While CSIRO educators did not bring this up as often in interviews as some other aspects, it was mentioned as an important factor and was seen often in our preliminary observation sessions. An unexpected result contributes to the educational value of an activity by requiring students to acknowledge that they did not fully understand whatever lesson the activity was teaching before the activity began. This encourages them to pay more attention to be sure they understand the lesson by the end of the activity. The scoring for the unexpected results criterion is based upon how likely the student is to expect the outcome of the activity. If the result is mundane and expected, or the principle behind the otherwise unusual result is in fact common enough knowledge that most students already know the "trick" ending, the activity earns few points for this criterion. An activity with a result that leaves students clearly stunned and intrigued would score maximum points.

The student interaction criterion refers to how much control a student has over the activity. Background research and CSIRO educators both mentioned the importance of being able to change and measure variables to aid understanding through practice, and hands-on aspects are strongly associated with the style of teaching used by CSIRO programs. The scoring for student interaction refers to how much control the student has over variables in the activity; an activity with no variables the student can change earns no points, and binary or trivial variables earn very few, but an activity with multiple meaningful variables, over which students have significant control, earns full points.

Ease of Completion

Ease of completion covers how easy it is for a student to understand how to perform an activity, and how likely they are to then do it successfully. Interviews with CSIRO educators, along with observations of programs, confirmed the need for activities to be easily completed, as activities that are difficult to complete can disrupt programs. A misleading setup can prevent activities from effectively engaging students and teaching the relevant concepts. Ease of completion was broken up into 3 subcategories: *clarity of instructions and expectations, intuitive to complete, and appropriate length*.

Clarity of instructions and expectations covered how effective instructions (both written instruction sheets and oral presenter descriptions) were. While this was already understood from life experiences, CSIRO educators emphasised the need to consider how easy it is for students to understand how to perform an activity. Low scores are given in this category to activities where most students do not understand what to do, even when following instructions, and high scores are given to activities where the instructions make it obvious exactly how to complete the activity.

Intuitive to complete measured how likely a student was to correctly perform an activity based on his or her own intuition. When discussing possible criteria, educators voiced concerns that a student would falsely assume that an activity was straightforward and complete the wrong task, because they didn't find it necessary to follow the given instructions. If more students are likely to incorrectly perform the activity or if educators have to strongly stress the correct way to do an activity, that activity scores fewer points, but if an activity can almost always be performed the correct way with almost no instruction, that activity gets a high score for this sub-criterion.

Appropriate length is a sub-criterion that measures how long an activity takes to complete. While a more complex and satisfying activity may naturally take longer, if all other things are equal, a longer activity could disrupt the workflow of students, or cause students to avoid the activity. CSIRO staff also largely thought positively of shorter activities, both out of concerns of workflow and to accommodate shorter attention spans, especially amongst younger students. Because different programs are targeted at different age groups, which have differing attention spans, what constitutes an "appropriate" length for an activity is often relative to the program as a whole. As a result, scoring for this criterion is based on how disruptive the activity would be, relative to the total time for activities and the expected length of the activities combined. Activities that single-handedly take up the majority of students' time score fewer points, and activities that can easily be done while waiting for a different activity to be opened up score highly.

4.1.2 Rubric Weighting

Final criteria importance weightings were determined using a mix of educator survey results and personal experiences from observing activities. The first weights that were determined for the rubric were those of the five main criteria categories. Table A shows the pairwise comparison data used to rank these categories in order of importance. The table shows the process by which we determined which categories were more important than others. When filling in the pairwise comparison table, we compared the relative importance of each category in the leftmost column with that of each other category in the topmost row. When we determined, based on the results of our surveys and observations, that the category on the left was more important, we filled the intersecting cell with green. Similarly, when the category on the left was found to be less important, we filled the cell with red, and when we felt that the two categories were equal or approximately equal in importance, we coloured the intersecting box yellow. Once all rows had been filled in, the criteria were ordered from most to least important, based on the number of green and yellow cells in their respective rows. All of the pairwise comparison charts used in the weighting process can be found in Appendix I.

Main Criteria	Educational Value	Initial Draw	Student Satisfaction	Ease of Completion	Logistical Feasibility
Educational Value					
Initial Draw					
Student Satisfaction					
Ease of Completion					
Logistical Feasibility					

Key:

More Important	Equally Important	Less Important

Table A: Pairwise Comparison Chart for Category Weighting

Within this ordered list of categories, we next ascertained how much more important each category was than the next, and based on these comparisons, each category was given a weight relative to the others. These weights were then normalised to determine the percentage of the final score contributed by each category. The final score percentages, along with the maximum points that can be earned for each category, are shown in Table B, below. A graphical representation of the breakdown of weights by overall category can be seen in Figure 4.

Criteria	Percentage of Final Score	Maximum Score
Educational Value	24.8%	248
Logistical Feasibility	21.4%	214
Initial Draw	11.9%	119
Satisfaction	25.2%	252
Ease of Completion	16.7%	167
Total	**100%**	**1000**

Table B: Per cent of Final Score and Maximum Score for Main Categories

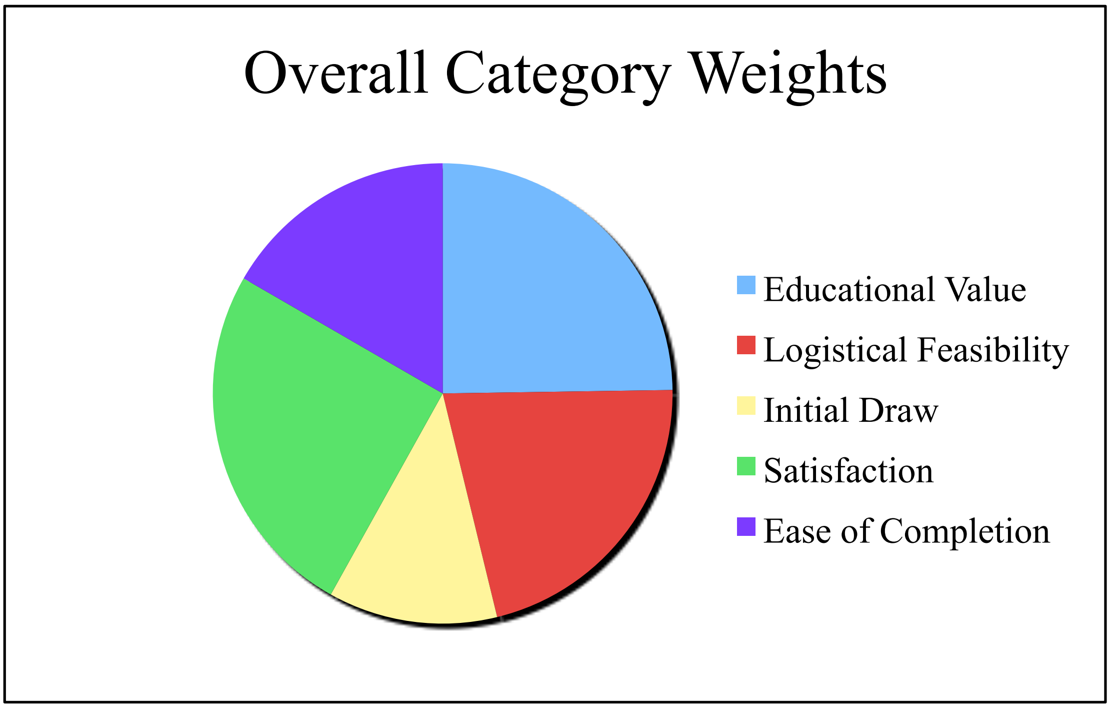


Figure 4: Relative Importance Weighting of Overall Categories

The final rankings that we determined for the main criteria were as follows:

- *Educational value* and *satisfaction* were both considered the factors with the highest importance, as both are absolutely crucial to a student taking away valuable lessons from the program. When determining the precise weights for these two criteria, we ultimately decided to give *satisfaction* a marginally higher weighting than educational value, because, though *educational value* is necessary for students to learn something from the program experience, *satisfaction* is critical for students to remember an activity at all.
- *Logistical feasibility* was given the third highest weight to show the impact it has on whether a program can run effectively and smoothly. While logistical

hurdles can sometimes be overcome in order to allow for more educational and satisfying activities, no activity can educate or satisfy students if it cannot be run at all.

- *Ease of completion* of an activity was given the fourth highest weight, as it aids education and satisfaction by making sure activities are completed as intended, but it is not as critical as *logistical feasibility*, as deficiencies in this category can more easily be overcome by the effort of the presenter.
- *Initial draw* was given the lowest weight since it was the only main criterion that was not truly required for an activity to meet the main goal of teaching science in an engaging way. Enthusiasm to begin an activity helps students stay focused on finishing the activity, but the *satisfaction* upon completion of an activity ultimately matters much more than the initial draw.

Educational Value

Once the weightings were determined for the overall categories, we then examined each category to assign weights to each of the sub-criteria within the category. Expected outcomes for the *educational value* category stressed potential for applications in future lessons As a result, encouraging student thought and potential for teacher extension had the largest weights within this category, since the former caused students to learn by performing the activity, and the latter aided teachers in helping students learn. While keeping an activity relevant to the curriculum was important, teachers, especially at the primary school level, generally cared more about being able to fit a lesson around an activity in order to make sure something was learned. The weights chosen for the sub-criteria of the educational value category are shown in Table C.

Educational Value Sub-Criterion	Percentage of sub-criteria score	Percentage of total score	Weight Multiplier	Maximum Points
Curriculum Relevance	28.3%	7.0%	14	70
Potential for Teacher Extension	35.8%	8.9%	17.8	89
Encourages Student Thought	35.8%	8.9%	17.8	89
Total	**100%**	**24.8%**	**49.6**	**248**

Table C: Weights of Educational Value Sub-criteria

Logistical Feasibility

For *logistical feasibility*, weights focused on ability to function rather than ease of function. As a result, the *health, safety, and environment* sub-criterion was considered to be the most important, a decision backed by almost all of the educators that were surveyed. The *cost* and *durability* sub-criteria both scored only slightly below *health, safety, and environment,* as both affect how well the program can be run, with lower-cost activities freeing up resources for other activities, and high durability scores decreasing the likelihood of a program being unable to be run due to damaged equipment. *Ease of repair* was the next most important because it was often crucial to keep a program running, but its importance depends heavily on the activity's durability. Since an average durability activity will likely go months, if not longer, before breaking, ease of repair or replacement would come into play infrequently. *Ease of setup/takedown and transport/handling* were both given low scores, as they are not crucial to the effectiveness of a program. A small number of consistently low scores in setup/takedown and transport/handling have an almost negligible effect on the program; only when a program consistently scores low does the inconvenience during setup actually create difficulties in running a program. The weights chosen for the logistics sub-criteria are shown in Table D, below.

Logistical Feasibility Sub-Criterion	Percentage of sub-criteria score	Percentage of total score	Weight Multiplier	Maximum Score
Health, Safety, and Environment	23.0%	4.9%	9.9	49
Ease of Setup and Takedown	10.3%	2.2%	4.4	22
Ease of Transport and Handling	10.3%	2.2%	4.4	22
Durability	20.7%	4.4%	8.9	44
Cost	20.7%	4.4%	8.9	44
Ease of Repair	14.9%	3.2%	6.4	32
Total	**99.9%**	**21.3%**	**42.9**	**213**

Table D: Weights of Logistical Feasibility Sub-criteria

Satisfaction

Within the *satisfaction* category, higher weights were given to sub-criteria that had something to do with the science the activity demonstrates, linking enjoyment to science as a subject. Thus, *student interaction* and *surprising results* were given the most weight, because they are most effective at getting students to enjoy an activity while also making the activity more informative and memorable. *Relevance* scored only slightly less, helping make an activity enjoyable and memorable, but doing less to make sure students recognised or understood a concept that was being taught. *Novel completion process* ended up with a weight of only slightly more than half that of *student interaction* and *surprising results*, as successfully accomplishing something new increases *student satisfaction*, but isn't integral to it. The weights determined for these criteria can be found in Table E.

Sub-Criterion	Percentage of sub-criteria score	Percentage of total score	Weight Multiplier	Maximum Score
Student Interaction	29.4%	7.4%	14.8	74
Novel Completion Process	16.2%	4.1%	8.2	41
Relevance to Students	25.0%	6.3%	12.6	63
Results are Surprising	29.4%	7.4%	14.8	74
Total	**100%**	**25.7%**	**50.4**	**252**

Table E: Weights of Satisfaction Sub-criteria

Ease of Completion

Ease of completion put the most weight on clarity of instructions and how intuitive an activity was, both being important in making sure a student would perform an activity correctly and thus understand the lesson. Appropriate length was weighted significantly lower, as it did not interfere with a student performing an activity correctly. The weights determined for the ease of completion sub-criteria can be found in Table F.

Sub-Criterion	Percentage of sub-criteria score	Percentage of total score	Weight Multiplier	Maximum Score
Instructions are Clear	38.4%	6.4%	12.8	64
Intuitive to Complete	36.5%	6.1%	12.2	61
Appropriate Length	25%	4.2%	8.3	42
Total	**99.9%**	**16.6%**	**33.3**	**167**

Table F: Weights of Ease of Completion Sub-criteria

4.2 Assessment of Rubric

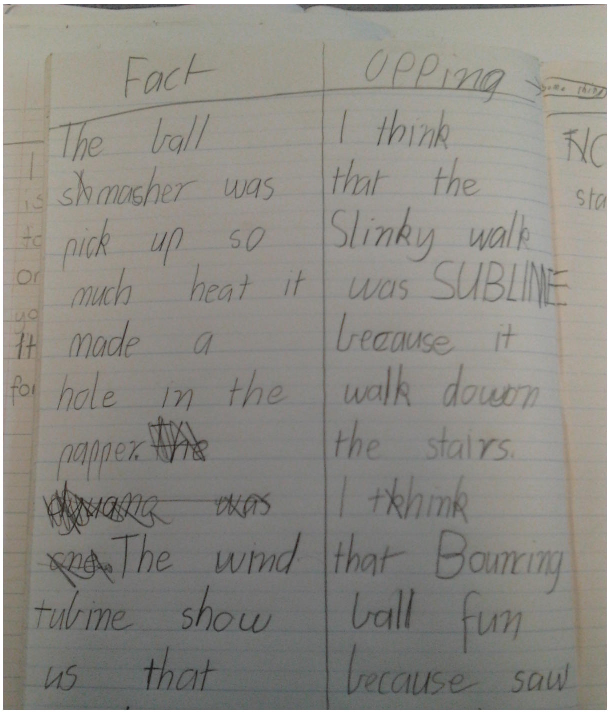
4.2.1 Validity of Rubric

The first step in determining how well our rubric worked was to confirm that it was measuring criteria we actually wanted to measure and representing the evaluated activities as accurately as possible. Through our program observations, conversations with teachers, and discussions with CSIRO educators, the majority of the criteria on our rubric were found to be valid, measurable criteria. However, our rubric is not flawless, and a few discrepancies between rubric scoring results and our qualitative data were found and noted.

During the assessment of our rubric's validity, we noticed a few areas in which the rubric results did not line up with the data collected from observations of programs as they were run. The largest discrepancy found was in the main criteria category of *student satisfaction*. For example, in Energy: Sources and Uses there is an activity that involves sending a Slinky down a small set of stairs. Nearly all of the students recognised the activity and knew what they were meant to do before the program presenter explained the activity. For this reason it received a low *novelty* score and high *relevance and familiarity* score. The *student interaction* score was fairly low, because the only variable students were able to change was whether the slinky had begun "walking" down the stairs. The students also knew exactly what would happen, even so they were still interested by it, so the result wasn't all that unexpected. When we scored it on the rubric with all of the other activities from Energy, it received the second lowest *student satisfaction* score once the individual sub-criteria scores were weighted and summed. This can be seen in Table G. However, all data collected from observations, student and teacher discussions, and a sample of students' written responses to the program contradict this, suggesting that the Slinky Walk activity was a student favourite. Images of these written student responses can be seen in Figure 5.

Activity:	Student Interaction	Novelty of Completion Process	Relevance/ Familiarity to Students	Surprising/ Unexpected Result	Final Weighted Rubric Score
Gyro Rings	4	5	1	5	**187**
Ball Smasher	2	4	3	5	**174**
Bouncing Balls	3	2	4	4	**171**
Pooh Pendulum	2	4	4	3	**157**
Pipe Lines	3	3	2	3	**139**
Wind Turbine	2	3	4	2	**134**
Twirling Tubes	2	4	2	3	**132**
Roll On	4	2	3	1	**128**
Solar Fan	2	2	4	2	**126**
Swinging Balls	2	3	1	4	**126**
Slinky Walk	2	1	4	2	**119**
Dynamo	2	2	2	1	**86**

Table G: Student Satisfaction Rubric Scores from Energy



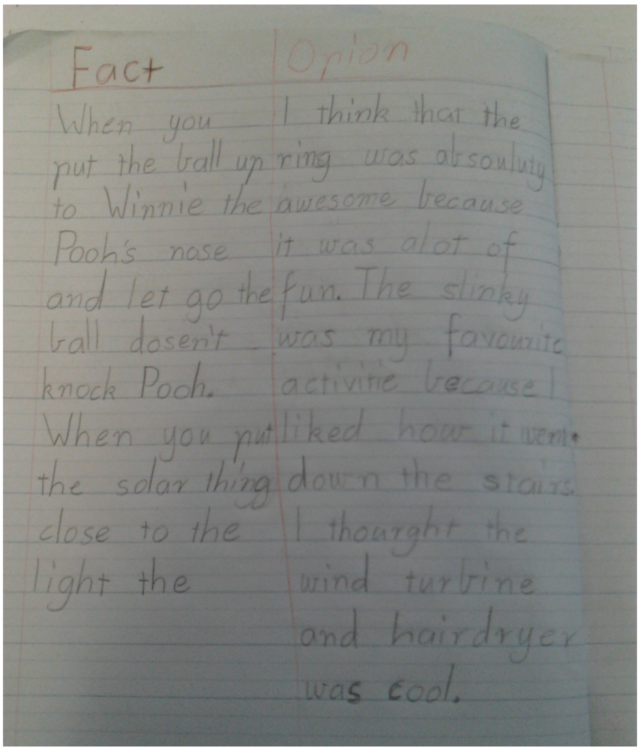


Figure 5: Two students written responses to the Energy Program

We believe the 4 sub-criteria under student satisfaction measure what they are meant to measure and do so consistently, meaning that the score we got was valid for the way the rubric is currently structured. However, during group discussions about this result and those of other activities whose *student satisfaction* rubric scores did not match their observed ability to excite students, we decided that our rubric was missing an element of *student satisfaction* that is difficult to quantify. For the Slinky Walk activity, the students still found the result of sending the slinky down the stairs interesting and "cool" despite knowing it was going to walk down the steps from the beginning. There are some things that are just innately interesting or awesome or cool and never really lose that aspect regardless of how many times a person has seen or done it. It is this "coolness" aspect that our rubric is missing in the *student satisfaction* area. However, the lack of this aspect is only noticeable in the satisfaction scores for a handful of activities across all CSIRO programs that we evaluated. Another notable example is the "Unusual Birds" activity in Forces, Movement and Simple Machines, which students loved but only scored average in *student satisfaction* for the program. For most other activities that students seemed to find exciting and cool, some combination of the four satisfaction sub-criteria in our rubric generally conveyed this sentiment. But these important exceptions indicate that something is missing that plays a role in the satisfaction of several of CSIRO's most popular activities.

Another area in which the validity of the rubric is questioned is in the sub-criterion of *instructions/ expectations are clear*, where questions and concerns were raised by various educators over description wording. The educators felt that the rubric descriptions could be interpreted in two ways, which were mutually exclusive to each other in nearly all cases. This criterion and two of its rubric score descriptions can be seen below in Figure 6, with the blue and green text marking the two possible interpretations.

Rubric Score:	0	5
Instructions/expectations are clear (written/ verbal)	Instructions/ expectations are not communicated at all, students have no idea what to do	Instructions/ expectations are very clearly communicated, all students understand exactly what to do

Figure 6: Instructions/Expectations are Clear Rubric Excerpt

The majority of educators we spoke with raised concerns that the first part of each criterion description (in blue) is highly dependent on the experience of the educator presenting the program and is also distinctly separate from and almost always mutually exclusive of the second part (in green), as CSIRO presenters tend to give clearer and more detailed instructions for activities in which students have less understanding of what to do. For this reason, the educators were unsure of which part of the criterion description to utilise in their evaluations of programs. The fact that the presenters saw the rubric descriptors as measuring two mutually exclusive items clearly calls the validity of the sub-criterion into question. Our team intended for *instructions/expectations are clear* to describe how well the students understood what they were meant to do, as a function of the presenter's oral instructions and the written instruction sheet for each activity. CSIRO education staff raised the legitimate point that nearly all activities have clearly written, easy to follow instruction sheets and the amount of oral instruction given by the presenter is dependent on the presenter's experience with past instances of the activity. The last point made was that some students will pay no attention to the oral or written instructions at all, so regardless of how well they are communicated, those students will never understand what they are supposed to be doing. To address this validity issue, our team believes that separating the two variables contained within the descriptors, and creating another criterion for the students' understanding of instructions would eliminate the concerns that have been raised by this criterion and increase the validity of the rubric substantially.

4.2.2 Reproducibility of Rubric

Based on the comparison of rubric scores given to the same activities by multiple educators, we found certain portions of our rubric to be very reproducible and consistent, while other criteria seem subject to significant discrepancies across the group of evaluators. Aided by the qualitative feedback of these educators with respect to the difficulties that they encountered, we analysed these inconsistent criteria and attributed most of these discrepancies to either a lack of clarity in rubric wording or differences in educator experiences.

One area in which our rubric demonstrated incredibly consistent scoring was the *logistical feasibility* category. For five of the six sub-criteria within this category, the average standard deviations between the scores of the six evaluators (five educators, plus the project team) over the seven activities in the Natural Disasters program were all below 0.75. The average ranges of scores, or the differences between the highest and lowest evaluator scores for each activity-criteria pair, given for each of these five criteria were all less than 2, indicating that all educators scored activities very similarly with regards to these criteria. These standard deviations and ranges for each logistical sub-criterion can be seen in Tables H and I, respectively. The full standard deviation and score range tables for all criteria can be found in Appendices O and P.

Standard Deviations	Health, Safety, and Environment	Ease of Setup / Takedown	Ease of Transport / Handling	Durability	Cost	Ease of Repair
Animated Earth	0.00	0.00	0.41	0.82	0.82	1.72
How Dense are You?	0.52	1.05	0.84	0.52	0.52	0.75
All Together Now	0.41	0.49	0.55	0.98	1.05	1.22
Round the Twist	0.52	0.63	0.41	0.41	0.41	1.60
Slip Sliding Away	0.52	0.52	0.75	0.52	0.80	1.52
It's Runny, Honey	0.52	0.00	0.76	1.03	0.89	0.82
Shaking All Over	0.82	0.41	0.75	0.41	0.42	1.08
Average	**0.47**	**0.44**	**0.64**	**0.67**	**0.70**	**1.25**

Table H: Standard Deviations of Educator Evaluations for Logistical Feasibility

Ranges of Scores	Health, Safety, and Environment	Ease of Setup / Takedown	Ease of Transport / Handling	Durability	Cost	Ease of Repair
Animated Earth	0.00	0.00	1.00	2.00	2.00	4.00
How Dense are You?	1.00	3.00	2.00	1.00	1.00	2.00
All Together Now	1.00	1.00	1.00	2.00	3.00	3.00
Round the Twist	1.00	2.00	1.00	1.00	1.00	4.00
Slip Sliding Away	1.00	1.00	2.00	1.00	2.00	4.00
It's Runny, Honey	1.00	0.00	2.00	3.00	2.00	2.00
Shaking All Over	2.00	1.00	2.00	1.00	1.00	3.00
Average	**1.00**	**1.14**	**1.57**	**1.57**	**1.71**	**3.14**

Table I: Ranges of Educator Evaluation Scores for Logistical Feasibility

This consistency is most likely a result of the precision and detail conveyed in the rubric descriptions for many of these criteria. For example, the *ease of setup and takedown* criterion, which received the most consistent scores across evaluators, specifies scores in terms of specific time limits. The rubric clearly states that an activity that takes longer than 10 minutes to set up or take down receives a score of 0, an activity that takes 6-10 minutes receives a score of 1, and so on. Thus, the subjectivity is removed from the evaluation of this criterion, leaving differences in presenter skill as the only factor that can cause scores to vary. Similarly, the *ease of transport* criterion specifies exact weights and sizes for activities to earn each score, and the *health, safety, and environment* criterion clarifies the type of preparation that must be taken for an activity receiving each score.

The only exception to this pattern is the logistical sub-criterion *ease of repair*. As Table I clearly shows, the average range of scores given to activities for this criterion was far larger than that of any other criterion that contributes to the *logistical feasibility* category. In fact, the scores for *ease of repair* varied significantly more widely than any other criterion on the rubric. Based on our discussions with multiple educators, we identified several issues that likely contribute to this variation. Firstly, educators brought to our attention the fact that some of the wording on this row of the rubric was somewhat misleading. One educator that we spoke with commented that many components of activities that become damaged or broken cannot be repaired but can be very easily replaced. The rubric, however, only mentions replacement under the lowest possible *ease of repair* score, while all other score descriptions focus on how long components will take to repair. Fortunately, this issue can be fixed very simply by modifying every mention of repair to include replacement as well.

More concerning is the inherent vagueness of the *ease of repair* criterion as to which components are being referred to. Multiple educators discussed with us the difficulty of assigning a single *ease of repair* score to an activity. Many activities have some components that break frequently, but can be easily repaired or replaced, along with others that are rarely damaged, but require an enormous amount of time and effort to replace when necessary. In such a situation, it is very difficult for an evaluator to assign a single score for an activity, as the rubric is unclear about which of these components should be taken into account. When educators evaluate the ease of repairing different parts of the same activity, it is very easy to obtain the wildly varying results that our comparisons show. The final impediment to the reproducibility of the *ease of repair* criterion is the variation in experience of each educator. Multiple educators are trained in each CSIRO program, and many of the activities in these programs only require repair or replacement every once in a while. When an educator has never experienced a certain type of damage to a given activity, it is very difficult for him or her to estimate the ease with which the activity can be repaired. These three issues combine to explain the low reproducibility of the *ease of repair* criterion on our rubric. With concerted future efforts to improve the clarity of this specific criterion, the *logistical feasibility* category as a whole will provide very consistent scores across evaluators.

Other reproducibility issues that our educator evaluation comparisons brought to light lay with scoring the educational value of an activity. In particular, the evaluators gave somewhat disparate scores to many activities with regards to the *curriculum relevance* and *potential for teacher extension/application* sub-criteria. For both of these criteria, the average standard deviation was greater than 0.95 and the average score range was greater than 2.25. Based on discussion with some of the educators who provided these scores, we feel that a large portion of this discrepancy is not due to excessive ambiguity in the rubric, but rather to lack of relevant knowledge among many educators. Every CSIRO educational program comes with teacher notes that highlight the concepts behind the program's activities, and list the links to the national or state curriculum that these activities contain. When a CSIRO educator is trained in presenting a new program, he or she is expected to study the program's teacher notes, ensuring that he or she understands everything about the program well enough to explain to students and teachers. Theoretically, this experience should allow educators to know exactly how much relevance each activity has to the curriculum, allowing them to score activities easily for this criterion. One problem lies with the fact that the national and state science curriculums have recently undergoing drastic changes, and though CSIRO has updated its teacher notes to match the new topics of study, it is generally not necessary for educators to reread these updated notes once they understand the activities themselves. Thus, many educators may not know how the activities that they present relate to the prescribed curriculum, creating problems when they wish to compare programs based on their *educational value*. Fortunately, if an evaluator has the time to perform a thorough evaluation of a program, taking a quick look through the most recent teacher notes for the curriculum links that each activity contains can largely rectify this issue.

Similarly, it can be very difficult to determine how easily a teacher can connect a particular activity to his or her lesson plans based solely on observing the activity as students perform it. For this criterion, examination of the teacher notes and curriculum objectives can help educators get a better idea of the kinds of lessons in which the activity could play an important role. Even with this research, it is difficult to predict what kinds of applications will actually work for teachers without asking them directly, and these applications can be so wide-ranging that it is difficult to make the rubric very specific on the matter. Thus, there will inherently be a certain amount of uncertainty in this criterion, but we feel that it is important enough to the success of an activity that it must be included.

One final area in which there was significant variation in the scores given by the six evaluators was the category of *satisfaction*. As Tables J and K show, each of the four sub-criteria that combine to form the satisfaction score had an average standard deviation greater than 0.85 and an average score range greater than 2.2.

Standard Deviations	Student Interaction	Novelty of Completion Process	Relevance to Students / Familiarity of Concepts	Results are Surprising / Unexpected
Animated Earth	0.55	1.17	0.89	1.22
How Dense are You?	1.72	1.03	0.52	1.37
All Together Now	1.41	0.75	0.75	0.75
Round the Twist	0.82	1.11	1.02	0.84
Slip Sliding Away	0.55	1.63	0.75	1.03
It's Runny, Honey	1.22	1.21	1.10	1.03
Shaking All Over	0.75	0.75	1.21	0.89
Average	**1.00**	**1.10**	**0.89**	**1.02**

Table J: Standard Deviations of Educator Evaluations for Satisfaction

Ranges	Student Interaction	Novelty of Completion Process	Relevance to Students / Familiarity of Concepts	Results are Surprising / Unexpected
Animated Earth	1.00	3.00	2.00	3.00
How Dense are You?	4.00	2.00	1.00	4.00
All Together Now	4.00	2.00	2.00	2.00
Round the Twist	2.00	3.00	2.50	2.00
Slip Sliding Away	1.00	4.00	2.00	3.00
It's Runny, Honey	3.00	3.00	3.00	3.00
Shaking All Over	2.00	2.00	3.00	2.00
Average	**2.43**	**2.71**	**2.21**	**2.71**

Table K: Ranges of Educator Evaluation Scores for Satisfaction

Of the five primary categories of criteria, the *satisfaction* that students get out of completing an activity is possibly the most difficult to measure, and thus inherently subject to a large degree of subjectivity in evaluation. For example, regardless of the scale used to quantify the information, it is difficult to determine how surprised students are by an activity or how novel the experience is for them. Unlike with many of the logistical considerations, most of the rubric descriptions for these criteria do not use any absolute units of measure, relying instead on relative qualifiers such as "mildly surprised" or "significant differences." Clearly these cannot perfectly describe a consistent method of scoring for all educators, but as there are no units of measure for satisfaction, the rubric settles for outlining clear distinctions between each score so that educators can easily, if not reproducibly, determine the differences between them.

Along with this general property of the satisfaction category as a whole, our discussions with educators raised a few points in which specific satisfaction criteria were found to be unclear and in need of improvement. One educator mentioned an issue with the scope of the *student interaction* criterion. As the rubric descriptions for this criterion were entirely focused on student modification of experimental variables, the criterion seemed ill equipped to deal with activities that do not involve experimentation. For example, the How Dense Are You activity in the Natural Disasters program was an involved multi-step process, in which students weighed several different volcanic rocks, submerged them in water to determine their volumes, and calculated their densities from these values. As this activity involved a large number of steps and a great deal of hands-on physical manipulation of a variety of materials and equipment, all things that generally characterise interaction, 2 of the 6 evaluators gave the activity a 5 out of 5 for student interaction. But if one interprets the rubric differently and more strictly, it appears that the density activity has few variables that students can actually experiment with. With this argument, 2 other evaluators gave the activity a 1 out of 5. This stark contrast in results clearly indicates that the criterion needs to be reworded to be clearer about what exactly it covers if the reproducibility of the rubric is to be increased.

Another interpretation issue was raised with regards to the *relevance to students / familiarity of concepts* criterion. One of the educators that we spoke to about the rubric experience was unsure about what type of relevance that this was referring to, noting that relevance to what students learn in class is very different from relevance to their life outside of school. Furthermore, he found it confusing that the criterion referred to both relevance and familiarity, which also measure two different things, making it difficult to give a single score for the entire criterion with any degree of consistency with other evaluators. This issue as well is one that can be addressed by modifying the wording of the criterion so that it is clear what constitutes the types of familiarity or relevance of interest to the rubric.

While we have highlighted a number of points at which the results of the developed rubric are not as reproducible as they could be, whether due to clarity issues or differences in the experiences of the educators, on the whole, the results are consistent enough among evaluators to provide a decent comparison between programs and activities with regards to various aspects and objectives. While a difference in score between programs or activities evaluated by different educators will not necessarily be a perfectly accurate comparison of their overall value, it will certainly serve as a very useful starting point for discussion on the strengths and weaknesses of each.

4.2.3 Universality of Rubric

One of the main reasons for observing several different programs was to test whether the rubric would be able to give consistently accurate results with a variety of scenarios and contexts. While scores stayed consistent amongst certain categories, differences in scores for various programs would need to be carefully analysed to determine whether different target audiences (i.e., age groups) have different priorities for sub-criteria, and thus should have a different set of weights for those sub-criteria. Table L, below, shows the overall category scores given to each observed program, and Figure 7 shows some of the trends in these scores across age groups.

Program	Ages of observed students	Overall Score	Educational Value	Logistical Feasibility	Initial Draw	Satisfaction	Ease of Completion
Forces, Movement and Simple Machines*	6-8	609	103	168	95	125	118
Energy: Sources and Uses	8-10	644	137	165	87	140	114
Natural Disasters	11-12	617	157	150	78	117	115
Chemistry: Actions and Reactions	13-14	628	136	136	77	153	126
VCE Physics: Materials and Structures	17-18	661	194	122	74	163	107
Maximum Possible Score		**1000**	**248**	**214**	**119**	**252**	**167**

Table L: Scores From All Observed Programs for Each Main Category

***Only enough data to score 7 out of 14 activities**

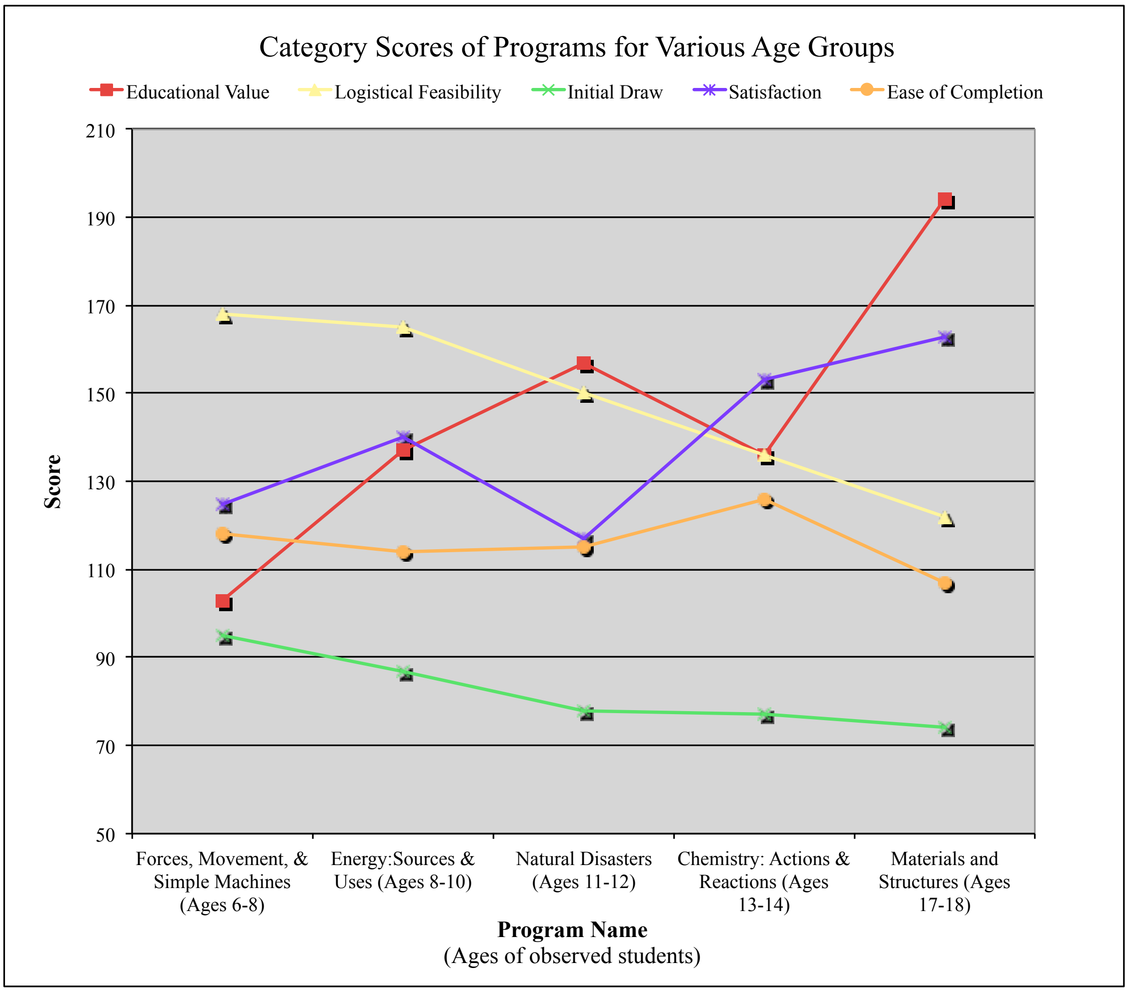


Figure 7: Category Scores of Programs for Various Age Groups

Logistical Feasibility

Logistical feasibility had one of the most noticeable trends based on program audience out of all criteria. Generally, the older the group of students a program was intended for, the worse the *logistical feasibility* scores were. These scores can be seen above in Table L. Almost every sub-criterion has a strong justification for earning lower scores with older students as a result of how older students will treat a program. *Health, safety, and environment* scores can generally be lower, since an older student is generally expected to be more aware of their surroundings and be careful around potentially unsafe activities. *Durability and ease of repair* become less essential for similar reasons, with older students less likely to carelessly damage CSIRO equipment. *Ease of setup and takedown, along with transport and handling*, are also justified in scoring lower, since the specialised nature of science classes in Australian schools lead to specialised programs, which are booked differently. Since these programs would be required by smaller number of students at specific points in the year, programs for older students were often booked in the same location for several days, meaning that any difficulties moving the program would occur less frequently. Lastly, the need for higher-end equipment requires more flexibility in budgeting; advanced equipment to better demonstrate practical applications with better precision is often expensive.

Educational Value

During data collection, we noticed that programs that were targeted at students taking higher levels of science classes tended to score higher in *educational value*, generally as a result of having more activities tied in to the curriculum. This was especially so in the VCE physics program, where every activity was intended to fit into either a lesson plan or a subject covered by the standardised test required at the end of the unit. In addition, older students who have begun taking classes in specific parts of science are generally better trained to use equipment, and activities can become more similar to experiments, increasing the likelihood that students will think through what they are doing and why they are doing it.

Initial Draw

Programs with a greater emphasis on answering questions in a workbook tended to score lower in *initial draw* overall, while programs without workbooks for students to fill out, such as Energy: Sources and Uses and Forces, Movement and Simple Machines, scored somewhat lower for this criterion. It is unclear whether the workbooks act as an alternative motivation and activities are designed with less emphasis on *initial draw*, or whether students tend to make interest less visible due to their desire to efficiently complete schoolwork, but in both cases *initial draw* may warrant a lower weight, since students now have other motivations to correctly complete activities.

Satisfaction

The main criterion of *student satisfaction* showed no discernable data trend in terms of higher scores being seen in programs for younger students or older students. However, the observation data for *student satisfaction* became more difficult for the team to collect the older the students were. This is likely a consequence of either the older students having more control in expressing their emotions towards activities, the fact that the students were focused on collecting data to fill out worksheets rather than just being able to do cool science activities as in most primary level programs, or both. Despite this difficulty, of the programs we observed, the ones for the year 12 and then the year 8 students had the highest and second highest overall satisfaction scores, respectively. This is likely due to the secondary school and VCE programs having a much greater level of student interactivity than the primary school programs, as the activities were generally significantly more complex. Sub-criterion within *student satisfaction* may need to be weighted different for primary school programs and secondary school programs to account for the inherent differences in what students are capable of doing at those ages.

Ease of Completion

Ease of completion rubric scores were fairly consistent amongst all observed programs, ranging from a 107 for VCE Physics: Materials & Structures to 126 for Chemistry: Actions & Reactions, out of a maximum possible score of 167. This relative consistency among programs indicates that the level of complexity of the activities rises consistently as the students increase in age and become capable of taking on more difficult tasks.

4.3 Activity Design Recommendations

4.3.1 Energy: Sources and Uses

The flexible solar cell activity serves two specific roles in the Energy: Sources and Uses program once it replaces the Solar Fan activity. The first role is to display the conversion of light energy to electrical energy, as no other activities in the program display a conversion *from* light energy, only a conversion *to* light energy. The second role the activity has is to display an example of new technology CSIRO has developed, changing how students think about solar energy and solar panels, and demonstrating what kinds of new technologies scientists can and are developing.

In order to properly serve these two roles, and to avoid weaknesses innate to any activity involving a solar cell, we suggest these recommendations and considerations for a flexible solar cell activity design:

- Activity needs to make clear the transfer of light energy to electrical energy, and display it in an obvious way. [Curriculum Relevance]
    - This is generally done through a further transfer from electrical energy to mechanical or light energy, which both provide very obvious sensory stimulus.
    - The way the transfer is displayed should be highly sensitive, clearly showing when the amount of energy being collected and transformed changes. [Encourages Student Thought, Student Interaction, Intuitive to Complete, Appropriate Length]
    - Because current flexible solar cells are generally inefficient at energy transformation, the way the transferred energy is output should also be extremely efficient and generate a larger, clearly noticeable effect with less energy. [Initial Draw, Results are Surprising/ Unexpected]
- The way the flexible solar cell is used should be easily recognisable and familiar to students, to make the connection between solar energy and other forms of energy more clear, and to make the activity more recognisable. [Potential for Teacher Extension; Relevance to Students]
- A solar cell activity will require a source of light to be provided by CSIRO, currently in the form of a bright light bulb that is very hot to the touch. The presence of this light bulb creates safety concerns, and any designs should seek to minimise these concerns to help compensate [Health, Safety, and Environment]
    - If possible, a design should also seek to minimise the amount of electricity needed, so that when the current light bulbs need to be replaced, lower heat alternatives can be used.
        - In addition, if the activity requires little enough electricity that a self-contained source of light can be used, this would mitigate concerns involving the use of the power strip and the requirement of a power point present in the current solar panel activity [Ease of Setup/Takedown; Ease of Transport/Handling]
- The flexible nature of the solar cells is key to their appeal over traditional solar panels, and any activity should in some way stress that difference and

how flexible solar cells can be used in different ways [Novel Completion Process; Student Interaction]

    - Current generation flexible solar cells are designed to be mounted to a specific shape rather than constantly being adjusted, and repeated bending can damage them. Because bending the solar cells is crucial to the role of displaying new technology, any activity design needs to have a quick and simple way of replacing a no longer functioning solar cell to prevent excessive downtime or repair expenses, and needs to have as few other points of failure as possible in order to compensate [Ease of Repair; Cost; Durability]
        - While unlikely to be possible, a design that allows the solar cell to be mounted in a fixed, though curved, shape without detracting from the demonstration of the new technology's capabilities would be very effective. [Durability]

4.3.2 VCE Chemistry: Polymers and Nano Chemistry

For the polymer program, the solar cell mainly serves to demonstrate new applications of polymers. It should focus on having clear calculations and demonstrating the chemistry behind the energy conversion. Because it is unlikely for a new technology to be directly related to the VCE Chemistry exam, the activity should also avoid detracting from the rest of the program without a strong link to the curriculum.

- Because the activity has weak curriculum relevance, effort should be made to keep the activity short, so that a group that begins it will not be held up completing it, preventing them from visiting other activities. [Curriculum Relevance, Appropriate Length]
- The activity should emphasise the unique properties of the solar cell, to emphasise why this example of new technology is important to the world. [Encourages Student Thought, Novel Completion Process, Surprising/Unexpected Results]

- The special properties of flexible solar cells can be magnified by comparing the flexible solar cell with traditional solar panels (which students are likely more familiar with) [Relevance to Students]
- The program mostly consists of small but numerous pieces of equipment. Having fewer parts reduces the confusion during setup, and having small parts keeps the program easy to transport [Ease of Transport/Handling, Ease of Setup/Takedown]
- Current flexible solar cells are designed to be mounted to a specific shape rather than constantly being adjusted, and repeated bending can damage them. Because bending the solar cells is crucial to the role of displaying new technology, any activity design needs to have a quick and simple way of replacing a no longer functioning solar cell to prevent excessive downtime or repair expenses, and needs to have as few other points of failure as possible in order to compensate [Ease of Repair; Cost; Durability]
- The activity should have ways to clearly and precisely measure any changes in energy, in order to calculate subtle effects on electricity generated, and ways to manipulate that generated electricity on measurable scales [Student Interaction, Potential for Teacher Extension]

5.0 Conclusions and Recommendations

5.1 Project Summary

During our seven-week project, we created a rubric for the evaluation of hands-on science activities. The design of this rubric was informed by data collected from a literature review, interviews and discussions with CSIRO educators, and observations of CSIRO's educational programs. The rubric evaluates activities based on a number of criteria, each falling into one of the following five categories:

1. Educational value
2. Logistical feasibility
3. Initial draw
4. Student satisfaction
5. Ease of completion

Through observation and evaluation of a variety of CSIRO educational programs, the project group tested the validity of the rubric's scoring of activities and the universality with which the rubric could be applied to programs covering a wide range of topics and age groups. The reproducibility of the rubric was tested by comparison of results between six CSIRO educators evaluating the same set of activities. Through these methods, we were able to conclude that our rubric was usable universally across programs, and that the results from most criteria were valid and reproducible.

5.2 Recommendations for Rubric Implementation

5.2.1 Use of Rubric

We recommend CSIRO Education staff utilise the rubric as a tool for evaluating existing activities and as a design guideline for creating new activities and programs. The rubric largely provides standardisation of subjective judgments, rather than completely objective, measurable criteria. For this reason, educators should use it as a tool for starting a discussion about the strengths and weaknesses of activities, rather than as concrete, non-debatable scores for which no discussion is necessary. The rubric should also be used a guide to design new activities and programs, ensuring that all of the important aspects of activities are considered during the design process.

5.2.2 Suggested Rubric Improvements

As previously discussed, the rubric has a few specific areas in which it can be improved by CSIRO staff to make it a more effective evaluation, design, and discussion tool. These areas for improvement are:

1. Creation of an additional student satisfaction sub-criterion that measures how cool students find an activity
2. Separation of "Instructions/ expectations are clear" into two sub-criterion: one which measures how clear the given instructions are, and one which measures how well students understand what they are meant to do
3. Changing the wording of the rubric score descriptors for "Ease of repair" to include replacement of damaged parts as well as the option to repair them
4. Rewording of the "Student interaction" sub-criterion in order to shift the focus away from experimental variables and include other types of physical interactivity.
5. Clarification and possible separation of the "relevance to students/ familiarity of concepts" sub-criterion

In addition, we recommend that CSIRO consider the following shifts in weightings based on target audiences for programs as they find it appropriate:

1. Reduce weight of the "Logistical Feasibility" criterion, or the weights of the " "Health, Safety, and Environment" and "Cost" sub-criteria for programs intended for older students, especially VCE programs, to compensate for the higher maturity and care demonstrated by VCE students as well as the increased program budgets for older students.
2. Increase weight of "Educational Value" criteria, or inter-criteria weight of the "Curriculum Relevance" sub-criterion for programs intended for more specialised classes, especially VCE programs, to magnify the weaknesses of low-education activities and compensate for a more specific curriculum with higher rates of standardised testing.

References

Ainley, J., Kos, J., & Nicholas, M. (2008). Participation in science, mathematics and technology in Australian education. *ACER Research Monographs*, 4.
Australian Curriculum Assessment and Reporting Authority. (2012). *The Australian Curriculum: Science*. Australia.
Carney, K., Hyman, A., Mello, E., & Snieckus, K. (2011). Teachers' Insights into Informal Science Education Programs: WPI.
Christensen, L., Nielsen, J. E., Rogers, C. M., & Volkov, B. (2005). Creative data collection in nonformal settings. *New Directions for Evaluation, 2005*(108), 73-79. doi: 10.1002/ev.172
CSIRO. (2009). What is CSIRO? Retrieved 18 January, 2013, from http://www.csiro.au/en/Portals/About-CSIRO/What-is-CSIRO.aspx
CSIRO. (2011). About CSIRO. Retrieved January 27, 2013, from http://www.csiro.au/Portals/About-CSIRO/About-CSIRO-overview.aspx
CSIRO. (2012). CSIRO Education, Victoria Primary Programs. Retrieved 1/20/2013, 2013, from http://www.csiro.au/Portals/Education/Teachers/Incursions-and-excursions/education-centres/Education-VIC/Primary-programs.aspx
Fraas, L. M., & Partain, L. D. (2010). *Solar cells and their applications* (Vol. 217): Wiley.
Green, M. A., Emery, K., Hishikawa, Y., Warta, W., & Dunlop, E. D. (2012). Solar cell efficiency tables (version 39). *Progress in Photovoltaics: Research and Applications, 20*(1), 12-20.
Harris, K. L. (2012). A Background in Science: What Science Means for Australian Society. *Centre for the Study of Higher Education*.
Satterthwait, D. (2010). Why Are" Hands-On" Science Activities so Effective for Student Learning? *Teaching Science, 56*(2), 7-10.
Sjoberg, S., & Schreiner, C. (2005). How do learners in different cultures relate to science and technology? *Asia-Pacific Forum on Science Learning and Teaching, 6*(2), 1-17.
Watkins, S. (2012). 'Printed power'to light the world's darkness? *ECOS, 2012*(172).
Xue, J. G. (2010). Perspectives on Organic Photovoltaics. *POLYMER REVIEWS, 50*(4), 411-419. doi: 10.1080/15583724.2010.515766

Appendices

Appendix A: Successful Activity Criteria from Literature Review

Communication of educational goals:

- Overarching Australian Curriculum goals
    - Identifying and posing questions
    - Planning, conducting and reflecting on investigations
    - Processing, analysing, and interpreting evidence
    - Communicating findings
- Curriculum links for Flexible Solar Cell Activities:
    - Foundation
        - Objects are made of materials that have observable properties
        - Science involves exploring and observing the world using the senses
    - Year 1
        - Everyday materials can be physically changed in a variety of ways
        - Participate in different types of guided investigations to explore and answer questions, such as manipulating materials, testing ideas, and accessing information sources
    - Year 2
        - Participate in different types of guided investigations to explore and answer questions, such as manipulating materials, testing ideas, and accessing information sources
    - Year 4
        - Natural and processed materials have a range of physical properties; these properties can influence their use
    - Year 5
        - Light from a source forms shadows and can be absorbed, reflected and refracted
        - Scientific understandings, discoveries and inventions are used to solve problems that directly affect peoples' lives
        - Scientific knowledge is used to inform personal and community decisions
    - Year 6
        - Electrical circuits provide a means of transferring and transforming electricity
        - Energy from a variety of sources can be used to generate electricity
        - Scientific understandings, discoveries and inventions are used to solve problems that directly affect peoples' lives
        - Scientific knowledge is used to inform personal and community decisions

Wow factor: Activities should contain/ be:

- Hands-on
- Responsive to student inputs
- Unexpected use of everyday items

- Unanticipated results
- Bright and/or multi coloured lights
- Sudden noises
- Students can try things and come to their own conclusions
- Must show off the properties of flexible solar cells
- Applications that clearly relate to students' daily lives
- Clear about which science concepts are involved

Logistics: Activities need to be:

- Portable: can fit on a classroom table
- Lightweight
- Durable: must not break if students are rough with it
- Safe for students to do
- Can be set up or dismantled in 5-10 minutes.
- Easy for students to understand
- Easy for students to interact with
- 3-5 minutes long
- Expense within reason

Appendix B: Elements of Successful Activities Mentioned in Educator Interviews

Concept	Portion of Interviewees	Number of Interviewees
Elements of a successful activity		
Activity itself must be interesting and fun to hold younger kids' attention	0.6	3
Links to applications that kids can relate to	0.6	3
Must be durable	0.6	3
Needs to relate to the curriculum and syllabus	0.6	3
Short and sweet (1-2 minutes for Primary Schools, 5-10 minutes for VCE)	0.6	3
Balance of fun and information	0.4	2
Calculations in VCE programs, but not in Primary School programs	0.4	2
Keep it simple	0.4	2
Makes or reiterates a point	0.4	2
Must be intuitive (kids probably won't read instruction sheet)	0.4	2
Older kids more interested in science and applications	0.4	2
Reasonably easy to put together	0.4	2
Requires students to think	0.4	2
Also good to have some longer activities	0.2	1
Breaking things is popular	0.2	1
Buttons help, but knobs are better	0.2	1
Changing a variable and observing the effect	0.2	1
Cheap enough (including maintenance and consumables)	0.2	1
Colour	0.2	1
Doing unusual things or getting unexpected results, especially for older kids	0.2	1
Don't let the activity get in the way of the science	0.2	1
Easy to fix if something goes wrong	0.2	1
Hands-on tends to be better	0.2	1
Kids are intrinsically interested but easily turned off	0.2	1
Look "sciency"	0.2	1
Multiple steps	0.2	1
Offer access to things teachers can't get, especially for older kids	0.2	1
Physically Active Activities are popular, especially for primary schools	0.2	1
Pouring things, things that look like chemistry	0.2	1
Safe enough	0.2	1
Some activities really interesting, but less intuitive: need more explanation	0.2	1
Something moving is more interesting: don't just generate energy: use it	0.2	1
Something new	0.2	1
Something that students have to work at	0.2	1

Teacher notes are very important	0.2	1
Teachers love measuring in VCE	0.2	1
Variety of activities lets everyone take away something	0.2	1

Appendix C: Suggestions for Uses of Flexible Solar Cells in Educator Interviews

Suggestions for Use of Flexible Solar Cells	Portion of Interviewees	Number of Interviewees
Compare flexible and traditional solar cells	0.6	3
Highlight advantages of flexible solar cells	0.6	3
Control bending so that solar cells don't break	0.4	2
More interesting applications details for older kids	0.4	2
(Only one activity in Polymers program demonstrates realistic applications)	0.2	1
A building with solar windows	0.2	1
Demo on soda bottle: solar cells power colour change or screen	0.2	1
Flexible solar cells work better at an angle	0.2	1
Human endeavour: alternatives to coal / impact of clean energy	0.2	1
Improve world by making life easier as well	0.2	1
Mock up solar cell on a physical object (bags, hats, model buildings)	0.2	1
Older kids can look at angular variation and conduct mini experiment	0.2	1
Pictures of polymer structure and how it comes off the printer would help	0.2	1
Put solar cells on the blades of the fan	0.2	1
Research aspect - modern context	0.2	1
Seeing current change as they mess with the solar cell	0.2	1
Show them lots of shapes for the solar cells	0.2	1
Things like bag show lightweight property, but the solar cells are still flat	0.2	1
Understanding: Transformation of light energy into electrical energy	0.2	1
VCE program focus is on comparing materials	0.2	1
VCE: Measure amount of electricity produced	0.2	1
Demo / discussion: charge phone from car covered in solar panels	0.2	1

Appendix D: Activity Observation Sheet

Observation	0	1	2	3	4	5
Initial Draw	Most students clearly do not want to begin the activity. They go because it is all that is free or because they are forced to	Students seem generally uninterested in beginning activity and bored by activity's appearance	Students begin activity without much obvious interest or reluctance	Some students appear interested in beginning activity, while others appear uninterested	Students seem very interested in beginning the activity	Students are openly very enthusiastic about beginning activity
Excitement/ Interest	Students are visibly bored or frustrated with the activity	Students appear confused about or somewhat disinterested with the activity	Student attitude is neutral toward the activity	Some students are smiling and otherwise showing excitement relating to the activity, while others are not	Students seem quite excited about and pleased with the activity; smiling is common	Students are very visibly excited about the activity. There is a lot of smiling and laughing
Referenced Instructions	Students never looked at instruction sheet	Students glanced only briefly at instruction sheet for only long enough to look at the picture	Students read some of the instruction sheet	Students referenced the instruction sheet more than once	Students referenced the instruction sheet repeatedly	Students clearly read through every step of the instructions as they completed the activity
Asked for Help	RE=Asked for and received help from educator	GE=Received un-asked-for educator help	RS=Asked for and received help from another student	GS=Received un-asked-for student help	RT=Asked for and received help from the teacher	GT=Received un-asked-for teacher help
Observation	**0**	**1**	**2**	**3**	**4**	**5**
Performed Correctly	Students did not complete any part of the activity whatsoever	Students completed only a minor portion of the activity, while missing most of the point	Students completed significant portions of the activity	Students completed most of the activity as intended	Students completed almost all of the activity as intended, missing only a minor part	Students completed all parts of the actvity as intended
Unexpected Result	Students seemed entirely unsurprised and even bored by the results of the activity	Students are not very surprised or interested by the results of the activity	Students are somewhat intrigued by results of the activity, though not very surprised	Students are mildly surprised by aspects of results	The results of the activity are surprising and unexpected for students	Students are clearly stunned and intrigued by the results of the activity
Student Thought	Students put the absolute minimum amount of thought into completing the activity	A few students put a small amount of thought into completing the activity	Activity completion process provokes thinking among some students	Activity completion process regularly prompts some thought from students	Activity completion process prompts significant thought from the majority of students	Activity completion process has nearly all students thinking critically
Apparent Novelty	Students give impression that activity is an everyday occurance	Students seem to mostly see activity as an everyday occurrence, with only minor, trivial differences	Students seem to find some part of the activity new and novel	Students seem to find significant portions of activity new and interesting	Students seem to find activity new and interesting	Students seem to find activity very different from anything that they have done before
Observation	**0**	**1**	**2**	**3**	**4**	**5**
Apparent Relevance/ Familiarity (make comments)						
Health, safety and environment (record presenter warning, if any)						

Appendix E: Presenter-led Discussion Questions

Presenters were told to use these question formats as a guideline for leading the end of program discussion with the students

Did anyone have a favourite activity? Which activity was it? Did anyone else like that activity? Have you ever done anything like that before?

Did anyone have a different favourite activity?   Which activity was it? Did anyone else like that activity? Have you ever done anything like that before?

Did you like X activity?

Why do you think X happened during activity Y?

How could Z apply to something you do? / How could/would you use Z technology?

Appendix F: Guidelines for Teacher Conversations

During our conversations with teachers, we tried to work in some of the following questions in order to obtain additional qualitative data.

1. How are you using these programs?
2. Is the program an introduction to a topic, or a wrap up?
3. If the program is an introduction to a unit, do you see yourself being able to refer back to any of these activities while you teach? If so, which ones?

Gauging student interest
4. How do you think your students are doing?
5. Are students acting/behaving like they normally do?
6. Is there any activity you've noticed that students seem drawn to? Why do you think that is?
7. Are there any activities that directly relate to what you are currently teaching?
8. Do you think there are any activities with a high relevance to students' lives?

Appendix G: Educator Criteria Importance Survey

Educator Criteria Importance Survey

We are this term's group of WPI students and we are working on the development of a rubric to evaluate the effectiveness of CSIRO hands-on activities. We have developed a list of criteria that we have found to contribute to the overall success of an activity, but we would like your informed input on the relative importance of these criteria. If you could take a few minutes to answer the following questions based on your own experience with these programs, it would help us a great deal, and we would very much appreciate it. Thank you for your time.

1. For how many years have you been working with informal education programs?

2. For how many years have you been working as an educator at CSIRO?

3. Do you primarily present programs for primary school students or secondary school students?

- I primarily present primary school programs
- I primarily present secondary school programs
- I present primary and secondary programs about equally

4. Based on your experience with CSIRO educational programs, please rate the following criteria based on how important they are to the overall success of a hands-on activity.

	1 (Not at all important)	2	3	4	5	6	7	8	9	10 (Critically important)
Student Satisfaction										
Ease of Completion										
Initial Draw										
Logistical Feasibility										
Educational Value										

5. Please rate the following criteria based on how important they are to the educational value of a hands-on activity.

	1 (Not at all important)	2	3	4	5	6	7	8	9	10 (Critically important)
Relevance to Curriculum										
Encourages Student Thought										
Potential for Teacher Extension or Application										

6. Please rate the following criteria based on how important they are to the ease of completion of a hands-on activity.

	1 (Not at all important)	2	3	4	5	6	7	8	9	10 (Critically important)
Intuitive to Complete Correctly										
Clarity or Instructions / Expectations										
Appropriate Length										

7. Please rate the following criteria based on how important they are to the satisfaction that students obtain from a hands-on activity.

	1 (Not at all important)	2	3	4	5	6	7	8	9	10 (Critically important)
Student Interaction and Interactiveness	○	○	○	○	○	○	○	○	○	○
Relevance to Student Lives / Real-world Implications	○	○	○	○	○	○	○	○	○	○
Novelty of Completion Process	○	○	○	○	○	○	○	○	○	○
Surprising / Unexpected Results	○	○	○	○	○	○	○	○	○	○

8. Please rate the following criteria based on how important they are to the logistical feasibility of a hands-on activity.

	1 (Not at all important)	2	3	4	5	6	7	8	9	10 (Critically important)
Ease of Repair	○	○	○	○	○	○	○	○	○	○
Durability	○	○	○	○	○	○	○	○	○	○
Health, Safety, and Environment	○	○	○	○	○	○	○	○	○	○
Ease of Transport and Handling	○	○	○	○	○	○	○	○	○	○
Cost	○	○	○	○	○	○	○	○	○	○
Ease of Setup and Takedown	○	○	○	○	○	○	○	○	○	○

Appendix H: Full Evaluation Rubric

Evaluator Score

Educational Value	**0**	**1**	**2**	**3**	**4**	**5**	**Notes:**
Science Understanding **Curriculum Relevance**	Activity does not cover any scientific concepts	The activity covers a scientific concept but that concept is not relevant to the curriculum of the appropriate year level	An aspect of the activity is tangentially related to a scientific concept in the curriculum	Some aspects of activity are directly related to scientific concept(s) in the curriculum	Main topic of activity is directly related to scientific concept(s) in the curriculum	Main scientific concept(s) in activity are ones that are covered directly by the curriculum	
Potential for teacher extension or application	Teachers cannot build upon concepts covered by the activity	Teachers will have extreme difficulty building upon the concepts covered by the activity	Teachers will have some trouble building upon the concepts covered by the activity	Teachers can build upon the concepts covered by the activity	Teachers can easily build upon the concepts covered by the activity	Teachers can build upon the concepts covered by the activity extremely easily/ lesson plan based on activity basically writes itself	
Encourages student thought	Students put the absolute minimum amount of thought into completing the activity	A few students put a small amount of thought into completing the activity	Activity completion process provokes thinking among some students	Activity completion process regularly prompts some thought from students	Activity completion process prompts significant thought from the majority of students	Activity completion process has nearly all students thinking critically	
Logistical Feasibility	**0**	**1**	**2**	**3**	**4**	**5**	**Notes:**
Health, safety and environment	The activity is very dangerous to perform, and presents a serious risk to the health and safety of students and others	The activity can be quite dangerous if significant care is not taken. Presenter must keep careful watch.	The activity presents a risk to students if not used carefully. A strict warning is required.	The activity presents a risk to students if used improperly, but a light warning is sufficient	The activity presents minimal health, safety and environmental risks that preparation can account for	The activity presents minimal/negligible safety, health and environmental risks.	
Ease of setup and takedown	Activity takes more than 10 minutes to set up and/or take down	Activity takes 6-10 minutes to set up and/or take down	Activity takes 4-6 minutes to set up and/or take down	Activity takes 2-4 minutes to set up and/or take down	Activity can be set up and taken down in less than two minutes, and is almost entirely self-contained when in storage	Activity set up and takedown consists solely of taking it out of the container, and putting it back in	
Ease of transport and handling	Activity weighs >10Kg, **and/or** is too large to fit into a box and is very difficult to handle	Activity weighs 7-10 Kg, **and/or** takes up most of the space in a box **OR** Doesn't fit in a box and is awkward to carry	Activity weighs 3-7 Kg **and/or** takes up 40-70% of a single box **OR** Activity shouldn't be carried in box but is easily carried separately	Activity weighs 1.5-3 Kg, **and/or** takes up 25-40% of a single box	Activity weighs 0.5-1.5Kg **and/or** takes up 10-25% of the space in a single box	Activity weighs less than 500g **and/or** takes up less than 10% of space in a single box	

Logistical Feasibility	0	1	2	3	4	5	Notes:
Durability	Activity is very fragile and/or is expected to be damaged frequently	Activity only lasts for a few sessions before being damaged (approx. monthly)	Activity is damaged on a fairly regular basis (approx. once per term)	Activity is damaged once in a while	Activity is damaged only on very rare occasions	Activity is never damaged and doesn't ever need to be replaced or repaired within its expected time of use	
Cost	Unfeasible without a large additional outside funding source	Uses up a significant component of program budget	Cost is high enough to make budgeting more challenging	Cost is reasonable for the available budget	Cost is low enough to make budgeting noticably easier	Cost of activity is negligible over the short and long term	
Ease of Repair	Cannot be fixed by CSIRO staff, time and cost to have it repaired or replaced is excessive	Cannot be fixed at program site and repairs could take a week or more	Cannot be fixed at program site, fixable once back at CSIRO	If broken probably fixable during a half-hour break between program sessions without special tools/or materials	All likely points of failure fixable mid-program	All likely points of failure easily fixable in a few minutes at most	
Initial Draw	**0**	**1**	**2**	**3**	**4**	**5**	**Notes:**
Initial Draw	Activity can be expected to be avoided by most students	Activity does not stand out or appears uninteresting	Activity has few or ineffective ways to draw student interest	Activity draws interest of some students	Activity consistently draws student interest, students want to begin	Activity is started enthusiastically by nearly all students	
Satisfaction	**0**	**1**	**2**	**3**	**4**	**5**	**Notes:**
Student Interaction	Student has no control over any variables (reading a book)	Student has minimal control of one variable (pressing a button)	Student has some control of one variable (turning knobs/cranks at different speeds)	Student has a significant amount of control over one variable OR minimal amounts of control over a few variables	Student has some amount of control over a few variables	Student has a significant amount of control over several meaningful variables	
Novel Completion Process	Activity is something students do everyday with no change in outcome	Activity has only minor, trivial differences from everyday experiences for students	Activity puts a small, but noticable twist on an otherwise everyday activity for students	Activity is conceptually similar to something that students do on a regular basis, but with significant differences in completion process	Activity is something that most students don't normally have the opportunity to do	Activity is very different from anything that the large majority of students have done before	

Satisfaction	**0**	**1**	**2**	**3**	**4**	**5**	**Notes:**
Relevance to students/ familiarity of concepts	Activity has no relevance or familiarity for nearly all students	Activity has minimal relevance or familiarity for the large majority of students	Activity has some relevance or familiarity for some students, but not for others	Activity has relevance or familiarity for most students	Activity has high relevance or familiarity to the large majority of students	Activity has extremely high relevance or familiarity to nearly all students	
Results are surprising/ unexpected	Students seemed entirely unsurprised and even bored by the results of the activity	Students are not very surprised or interested by the results of the activity	Students are somewhat intrigued by results of the activity, though not very surprised	Students are mildly surprised by aspects of results	The results of the activity are surprising and unexpected for students	Students are clearly stunned and intrigued by the results of the activity	
Ease of Completion	**0**	**1**	**2**	**3**	**4**	**5**	**Notes:**
Instructions/expectations are clear (written/ verbal)	Instructions/ expectations are not communicated at all, students have no idea what to do	Instructions/ expectations are communicated very poorly, nearly all students are unsure of what to do	Instructions/ expectations are not communicated well, a majority of students do not understand what to do	Instructions/ expectations are communicated decently, some students are unsure of what to do	Instructions/ expectations are well communicated and most students understand what to do	Instructions/ expectations are very clearly communicated, all students understand exactly what to do	
Intuitive to Complete	Even with a strong emphasis on correct instructions, students are still likely to perform activity incorrectly	Presenter instructions need to be made very clear in order for students to not complete incorrectly	Some students may make false assumptions about activity, but are easily corrected by presenter instructions	Students do not know intuitively how activity works, but are unlikely to make false assumptions either	Presenter instructions largely confirm correct assumptions on the nature of the activity	Presenter instructions are entirely unnecessary; merely seeing the activity as it is set up makes how to do the activity clear	
Appropriate length (amount of time activity takes to complete correctly)	Activity consumes a very large portion of "hands-on" time. Few groups do it, and those that do don't get to do much else	Activity usually cannot be completed in a reasonable amount of time	Activity takes longer to complete than would be ideal for the program	Activity can be completed in a reasonable amount of time	Activity can be performed relatively quickly, and very rarely holds groups up	Activity can be performed very quickly and can be done while waiting for other activities	

Appendix I: Pairwise Comparisons used for Criteria Weighting

Main Criteria	Educational Value	Initial Draw	Student Satisfaction	Ease of Completion	Logistical Feasibility
Educational Value					
Initial Draw					
Student Satisfaction					
Ease of Completion					
Logistical Feasibility					

Educational Value	Curriculum Relevance	Potential for teacher extension	Encourages student thought
Curriculum Relevance			
Potential for teacher extension			
Encourages student thought			

Student Satisfaction	Student interaction	Novel completion process	Relevance/ familiarity to students	Unexpected/ surprising results
Student interaction				
Novel completion process				
Relevance/ familiarity to students				
Unexpected/ surprising results				

Ease of Completion	Instructions/ expectations are clear	Intuitive to complete	Appropriate length
Instructions/ expectations are clear			
Intuitive to complete			
Appropriate length			

Logistical Feasibility	Health, Safety & Environment	Cost	Ease of setup and takedown	Ease of transport and handling	Durability	Ease of repair
Health, Safety & Environment						
Cost						
Ease of setup and takedown						
Ease of transport and handling						
Durability						
Ease of repair						Dependent on durability

Key:

More Important	Equally Important	Less Important

Appendix J: Activity Scoring Sheet

Activity 1 Name	Evaluator Score	Weighting Multiplier	Weighted Score	Max Possible Score
Educational Value				
Science Understanding Curriculum Relevance		14.0	0	70
Potential for teacher extension or application		17.8	0	89
Encourages student thought		17.8	0	89
Educational Value Total Score			0	248
	Evaluator Score	Weighting Multiplier	Weighted Score	Max Possible Score
Logistical Feasibility				
Health, safety and environment		9.9	0	49
Ease of setup and takedown		4.4	0	22
Ease of transport and handling		4.4	0	22
Durability		8.9	0	44
Cost		8.9	0	44
Ease of Repair		6.4	0	32
Logistical Feasibility Total Score			0	214
	Evaluator Score	Weighting Multiplier	Weighted Score	Max Possible Score
Initial Draw				
Initial Draw		23.8	0	119
Initial Draw Total Score			0	119
	Evaluator Score	Weighting Multiplier	Weighted Score	Max Possible Score
Satisfaction				
Student Interaction		14.8	0	74
Novel Completion Process		8.2	0	41
Relevance to students/ familiarity of concepts		12.6	0	63
Results are surprising/ unexpected		14.8	0	74
Satisfaction Total Score			0	252
	Evaluator Score	Weighting Multiplier	Weighted Score	Max Possible Score
Ease of Completion				
Instructions/expectations are clear (written/ verbal)		12.8	0	64
Intuitive to Complete		12.2	0	61
Appropriate length		8.3	0	42
Ease of Completion Total Score		33.3	0	167
			Weighted Score	Max Possible Score
Total Score			**0**	1000

Appendix K: Rubric Weighting Sheet

Overall Categories	Relative Weight	Percent of Total	Category Weight		Max Possible Score
Educational Value	260	24.8	49.5		248
Logistical Feasibility	225	21.4	42.9		214
Initial Draw	125	11.9	23.8		119
Satisfaction	265	25.2	50.5		252
Ease of Completion	175	16.7	33.3		167
Educational Value	Relative Weight	Percent of Category	Percent of Total	Weight	Max Possible Score
Curriculum Relevance	75	28.3	7.0	14.0	70
Potential for Teacher Extension / Application	95	35.8	8.9	17.8	89
Encourages Student Thought	95	35.8	8.9	17.8	89
Logistical Feasibility	Relative Weight	Percent of Category	Percent of Total	Weight	Max Possible Score
Health, safety and environment	100	23.0	4.9	9.9	49
Ease of setup and takedown	45	10.3	2.2	4.4	22
Ease of transport and handling	45	10.3	2.2	4.4	22
Durability	90	20.7	4.4	8.9	44
Cost	90	20.7	4.4	8.9	44
Ease of Repair	65	14.9	3.2	6.4	32
Initial Draw	Relative Weight	Percent of Category	Percent of Total	Weight	Max Possible Score
Initial Draw	1	100.0	11.9	23.8	119
Satisfaction	Relative Weight	Percent of Category	Percent of Total	Weight	Max Possible Score
Student Interaction	100	29.4	7.4	14.8	74
Novel Completion Process	55	16.2	4.1	8.2	41
Relevance to students/ familiarity of concepts	85	25.0	6.3	12.6	63
Results are surprising/ unexpected	100	29.4	7.4	14.8	74
Ease of Completion	Relative Weight	Percent of Category	Percent of Total	Weight	Max Possible Score
Instructions/expectations are clear (written/ verbal)	100	38.5	6.4	12.8	64
Intuitive to Complete	95	36.5	6.1	12.2	61
Appropriate length	65	25.0	4.2	8.3	42
Maximum Possible Score	1000				

Appendix L: Sub-Criteria Rubric Scores for Observed Activities

Chemistry: Actions & Reactions

Criteria	Mystery Muddle	Reversible Reactions	Flocculation	Chemistry catches the cheater	heat sensitive materials	Slime	Compression	Centrifuge
Science Understanding Curriculum Relevance	3	4	4	2	1	4	5	4
Potential for teacher extension or application	3	4	3	1	3	0	4	3
Encourages student thought	3	2	3	2	4	1	1	3
Health, safety and environment	2	3	2	5	3	4	5	4
Ease of setup and takedown	2	3	1	4	3	2	5	2
Ease of transport and handling	3	4	3	4	3	4	5	4
Durability	4	4	4	4	5	4	3	4
Cost	3	3	3	2	3	3	4	1
Ease of Repair	2	2	2	1	3	3	2	1
Initial Draw	3	3	3	4	4	5	1	3
Student Interaction	4	4	3	1	4	4	2	2.5
Novel Completion Process	2.5	2	3	4	5	5	2	4
Relevance to students/ familiarity of concepts	3	3	4	3	2	3	4	2
Results are surprising/ unexpected	2	2	3	2	5	4	1	3
Instructions/expectations are clear	3	4	3	5	4	5	5	4
Intuitive to complete	3	3.5	3	3	4	3.5	4	2
Appropriate length (amount of time activity takes to complete)	3	4	3	4	4	3	5	3
TOTAL:	**48.5**	**54.5**	**50**	**51**	**60**	**57.5**	**58**	**49.5**
Notes:								

students had finished their chemistry unit in term 1, this was a wrap up

Energy: Sources and Uses

Criteria	Bouncing Balls	Slinky Walk	Ball Smasher	Pooh Pendulum	Twirling tubes	Pipe Lines	Swinging Balls	Roll on	Dynamo	Gyro Rings	Solar Fan	Wind Turbine
Science Understanding Curriculum Relevance	4	3	4	5	2	4	3	3	4	2	5	5
Potential for teacher extension or application	3	2	4	4	1	3	2	3	4	1	5	4
Encourages student thought	4	1	2	1	1	3	2	2	1	2	1	1
Health, safety and environment	5	5	3	5	3	4	5	5	3	5	3	3
Ease of setup and takedown	5	5	3	4	5	3	5	4	5	5	3	3
Ease of transport and handling	4	4	3	5	5	4	5	4	3	5	3	3
Durability	5	3	5	5	5	4	5	5	4	5	4	4
Cost	4	3	2	3	5	2	4	3	3	4	3	3
Ease of Repair	4	1	2	4	2	2	4	4	2	5	4	4
Initial Draw	5	4	4	4	4	3	3	3	3	5	3	3
Student Interaction	3	2	2	2	2	3	2	4	2	4	2	2
Novel Completion Process	2	1	4	4	4	3	3	2	2	5	2	3
Relevance to students/ familiarity of concepts	4	4	3	4	2	2	1	3	2	1	4	4
Results are surprising/ unexpected	4	2	5	3	3	3	4	1	1	5	2	2
Instructions/expectations are clear	4	5	4	2	4	4	3	4	5	2	4	4
Intuitive to complete	3	5	3	1	4	4	1	2	4	0	4	2
Appropriate length (amount of time activity takes to complete)	3	5	4	4	5	4	3	3.5	5	2	4	4
TOTAL:	**66**	**55**	**57**	**60**	**57**	**55**	**55**	**55.5**	**53**	**58**	**56**	**54**
Notes:	Activity was done outside		Students really, really liked this one			time cost	Students have no patience			students generally require some adult help to get started	Ease of setup dependant on how fast power can be found	

Forces, Movement and Simple Machines

Criteria	Unusual Birds	Obedient Can	Magnetic Marbles	Pick-up cones	Trebuchet	records	Big Lever
Science Understanding Curriculum Relevance	1	2	4	3	2	1	1
Potential for teacher extension or application	2	1	1.5	1	3	2	4
Encourages student thought	2	1	3	2	2	2	3
Health, safety and environment	3	5	5	5	4	5	2
Ease of setup and takedown	5	5	4	5	4	5	5
Ease of transport and handling	5	5	5	5	4	4	1
Durability	3	3	5	5	4	4	5
Cost	4	3	4	3	3	5	3
Ease of Repair	3	2	5	3	3	3	2
Initial Draw	4	4	4	3	4	4	5
Student Interaction	2	2	4	2	2.5	3	4
Novel Completion Process	3	4	2	1	4	4	3
Relevance to students/ familiarity of concepts	1	0	1	2	1	1	2
Results are surprising/ unexpected	4	4	1	1	3	4	4
Instructions/expectations are clear	4	4	1	4	3	4	5
Intuitive to complete	4	2	5	4	3	2	4
Appropriate length (amount of time activity takes to complete)	4	5	3	4	3	4	3
TOTAL:	**54**	**52**	**57.5**	**53**	**52.5**	**57**	**56**
Notes:							

Using Old Rubric (V3.0)
Only 2 observers
Only activities with 2 data points or more scored

Natural Disasters

Criteria	Animated Earth	How Dense are you?	All Together Now	Tornado	landslide	Viscosity	Shaking all over
Science Understanding Curriculum Relevance	5	2	5	3	3	2	4
Potential for teacher extension or application	2	4	5	3	3	3	3
Encourages student thought	2	3	3	3	2	3	4
Health, safety and environment	5	4	5	4	4	5	4
Ease of setup and takedown	5	1	4	3	4	5	4
Ease of transport and handling	5	2	5	4	3	4	2
Durability	3	4	3	4	5	3	4
Cost	5	3	4	2	3	3	3
Ease of Repair	2	2	2	1	2	2	2
Initial Draw	2	3	3	5	3	3	4
Student Interaction	0	1	3	2	2	1	3
Novel Completion Process	1	3	1	5	1	3	3
Relevance to students/ familiarity of concepts	1	2	4	4	2	4	4
Results are surprising/ unexpected	3	3	1	4	1	2	2
Instructions/expectations are clear	5	2	5	5	4	4	3
Intuitive to complete	4	1	3	4	4	3	3
Appropriate length (amount of time activity takes to complete)	4	1	3	4	4	2	3
TOTAL:	**54**	**41**	**59**	**60**	**50**	**52**	**55**
Notes:	flip book did break at one point	takedown is a pain	cost= man hours	will take time to fix		issues with stopwatch	

VCE Physics: Materials & Structures

Criteria	Strength of Materials 1	Strength of Materials 2	Truss Behavior	Forces in a Truss	Heat Treatment	Reaction Forces	Temperature Vs. Toughness	Composites
Science Understanding Curriculum Relevance	5	5	5	5	3	5	4	4
Potential for teacher extension or application	5	5	5	4	3	4	4	3
Encourages student thought	4	4	3	3	3	2	4	3
Health, safety and environment	4	3	5	5	2	5	2	3
Ease of setup and takedown	1	1	0	5	3	3	1	4
Ease of transport and handling	0	0	0	1	4	1	1	0
Durability	4	3	3	4	5	4	4	5
Cost	1	1	3	4	4	3	3	3
Ease of Repair	1	1	1	2	4	2	2	3
Initial Draw	3	3	3	2	4	2	4	4
Student Interaction	5	5	4	1	4	2	4	2
Novel Completion Process	5	5	4	1	4	1.5	4.5	4
Relevance to students/ familiarity of concepts	3	3	4	4	4	2	3	3
Results are surprising/ unexpected	3	3	3	1	4	1	5	3
Instructions/expectations are clear	5	3	2	5	3	5	4	4
Intuitive to complete	1	1	0	5	2	5	4	4
Appropriate length (amount of time activity takes to complete)	1	1	1	5	3	5	3	4
TOTAL:	**51**	**47**	**46**	**57**	**59**	**52.5**	**56.5**	**56**
Notes:	Program normally lives at Highett							

Appendix M: Rubric Summary Sheets for Observed Programs

Chemistry: Actions & Reactions

Overall Score	
Program Average	**624**
Heat Sensistive Materials	713
Slime	669
Reversible Reactions	640
Compression	617
Flocculation	610
Centrifuge	591
Mystery Muddle	584
Chem Catches Cheater	570

Educational Value	
Program Average	**136**
Reversible Reactions	163
Flocculation	163
Centrifuge	163
Compression	159
Mystery Muddle	149
Heat Sensistive Materials	138
Chem Catches Cheater	81
Slime	74

Logistical Feasibility	
Program Average	**136**
Compression	168
Slime	147
Heat Sensistive Materials	146
Chem Catches Cheater	144
Reversible Reactions	135
Mystery Muddle	117
Centrifuge	117
Flocculation	112

Initial Draw	
Program Average	**77**
Slime	119
Chem Catches Cheater	95
Heat Sensistive Materials	95
Mystery Muddle	71
Reversible Reactions	71
Flocculation	71
Centrifuge	71
Compression	24

Satisfaction	
Program Average	**152**
Heat Sensistive Materials	200
Slime	197
Flocculation	164
Mystery Muddle	147
Reversible Reactions	143
Centrifuge	140
Chem Catches Cheater	115
Compression	111

Ease of Completion	
Program Average	**123**
Compression	154
Chem Catches Cheater	134
Heat Sensistive Materials	133
Slime	132
Reversible Reactions	127
Centrifuge	101
Mystery Muddle	100
Flocculation	100

Energy: Sources and Uses

Overall Score	
Program Average	**642**
Bouncing Balls	778
Ball Smasher	684
Pooh Pendulum	668
Solar Fan	651
Pipe Lines	643
Gyro Rings	635
Slinky Walk	624
Wind Turbine	617
Roll On	617
Twirling Tubes	608
Dynamo	597
Swinging Balls	585

Educational Value	
Program Average	**136**
Bouncing Balls	180
Solar Fan	177
Ball Smasher	163
Pipe Lines	163
Pooh Pendulum	159
Wind Turbine	159
Dynamo	145
Roll On	131
Swinging Balls	113
Slinky Walk	95
Gyro Rings	81
Twirling Tubes	64

Logistical Feasibility	
Program Average	**165**
Gyro Rings	205
Swinging Balls	199
Bouncing Balls	195
Pooh Pendulum	186
Roll On	181
Twirling Tubes	175
Slinky Walk	149
Solar Fan	144
Wind Turbine	144
Dynamo	140
Pipe Lines	136
Ball Smasher	131

Initial Draw	
Program Average	**87**
Bouncing Balls	119
Gyro Rings	119
Slinky Walk	95
Ball Smasher	95
Pooh Pendulum	95
Twirling Tubes	95
Pipe Lines	71
Swinging Balls	71
Roll On	71
Dynamo	71
Solar Fan	71
Wind Turbine	71

Satisfaction	
Program Average	**140**
Gyro Rings	187
Ball Smasher	174
Bouncing Balls	171
Pooh Pendulum	157
Pipe Lines	139
Wind Turbine	134
Twirling Tubes	132
Roll On	128
Solar Fan	126
Swinging Balls	126
Slinky Walk	118
Dynamo	86

Ease of Completion	
Program Average	**114**
Slinky Walk	167
Dynamo	154
Twirling Tubes	142
Pipe Lines	133
Solar Fan	133
Ball Smasher	121
Bouncing Balls	113
Wind Turbine	109
Roll On	105
Swinging Balls	76
Pooh Pendulum	71
Gyro Rings	42

Forces, Movement & Simple Machines

Only 7 of 14 Activities scored

Overall Score

Program Average	**609**
Big Lever	694
Magnetic Marbles	634
Records	627
Trebuchet	595
Unusual Birds	595
Pick-up Cones	562
Obedient Can	557

Educational Value

Program Average	**103**
Big Lever	138
Magnetic Marbles	136
Trebuchet	117
Pick-up Cones	95
Unusual Birds	85
Records	85
Obedient Can	64

Logistical Feasibility

Program Average	**168**
Magnetic Marbles	201
Records	188
Pick-up Cones	184
Obedient Can	160
Trebuchet	156
Unusual Birds	155
Big Lever	130

Initial Draw

Program Average	**95**
Big Lever	119
Unusual Birds	95
Obedient Can	95
Magnetic Marbles	95
Trebuchet	95
Records	95
Pick-up Cones	71

Satisfaction

Program Average	**125**
Big Lever	169
Records	149
Trebuchet	127
Unusual Birds	126
Obedient Can	122
Magnetic Marbles	103
Pick-up Cones	78

Ease of Completion

Program Average	**118**
Big Lever	138
Unusual Birds	133
Pick-up Cones	133
Obedient Can	117
Records	109
Trebuchet	100
Magnetic Marbles	99

Natural Disasters

Overall Score

Program Average	**617**
Tornado	724
All Together Now	691
Shaking All Over	666
It's Runny, Honey	585
Animated Earth	578
Landslide	568
How Dense Are You?	507

Educational Value

Program Average	**157**
All Together Now	212
Shaking All Over	180
How Dense Are You?	152
Tornado	149
Animated Earth	141
It's Runny, Honey	135
Landslide	131

Logistical Feasibility

Program Average	**150**
Animated Earth	177
All Together Now	164
It's Runny, Honey	155
Landslide	154
Shaking All Over	141
Tornado	130
How Dense Are You?	128

Initial Draw

Program Average	**78**
Tornado	119
Shaking All Over	95
How Dense Are You?	71
All Together Now	71
Landslide	71
It's Runny, Honey	71
Animated Earth	48

Satisfaction

Program Average	**117**
Tornado	180
Shaking All Over	149
It's Runny, Honey	120
All Together Now	118
How Dense Are You?	109
Landslide	78
Animated Earth	65

Ease of Completion

Program Average	**115**
Animated Earth	146
Tornado	146
Landslide	133
All Together Now	126
It's Runny, Honey	104
Shaking All Over	100
How Dense Are You?	46

VCE Physics: Materials & Structures

Overall Score	
Program Average	**661**
Temperature vs. Toughness	730
Heat Treatment	690
Strength of Materials 1	678
Composites	673
Forces in a Truss	657
Strength of Materials 2	634
Reaction Forces	615
Truss Behavior	613

Educational Value	
Program Average	**194**
Strength of Materials 1	230
Strength of Materials 2	230
Truss Behavior	212
Temperature vs. Toughness	198
Forces in a Truss	194
Reaction Forces	177
Composites	163
Heat Treatment	149

Logistical Feasibility	
Program Average	**122**
Forces in a Truss	160
Heat Treatment	156
Reaction Forces	142
Composites	137
Truss Behavior	109
Temperature vs. Toughness	103
Strength of Materials 1	95
Strength of Materials 2	76

Initial Draw	
Program Average	**74**
Heat Treatment	95
Temperature vs. Toughness	95
Composites	95
Strength of Materials 1	71
Strength of Materials 2	71
Truss Behavior	71
Forces in a Truss	48
Reaction Forces	48

Satisfaction	
Program Average	**163**
Temperature vs. Toughness	208
Heat Treatment	202
Strength of Materials 1	197
Strength of Materials 2	197
Truss Behavior	187
Composites	145
Forces in a Truss	88
Reaction Forces	82

Ease of Completion	
Program Average	**107**
Forces in a Truss	167
Reaction Forces	167
Composites	133
Temperature vs. Toughness	125
Heat Treatment	88
Strength of Materials 1	85
Strength of Materials 2	59
Truss Behavior	34

Appendix N: Sub-Criteria Rubric Scores for Educator Program Evaluations

Curriculum Relevence	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	5	5	?	3	5	5	4.6
How Dense are You?	2	5	?	3	2	2	2.8
All Together Now	5	4	?	4	5	5	4.6
Round the Twist	3	4	?	2.5	2	4	3.1
Slip Sliding Away	3	4	?	3	1	5	3.2
It's Runny, Honey	2	4	?	4	1	2	2.6
Shaking All Over	4	4	?	3	4	4	3.8

Potential for Teacher Extension or Application	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	2	1	2	3	4	3	2.5
How Dense are You?	4	2	3	4	1	3	2.8
All Together Now	5	2	4	3	4	4	3.7
Round the Twist	3	3	1	3	2	4	2.7
Slip Sliding Away	3	4	3	3	1	4	3.0
It's Runny, Honey	3	4	3	3	4	3	3.3
Shaking All Over	3	4	2	3	4	4	3.3

Encourages Student Thought	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	2	1	1	1	0	3	1.3
How Dense are You?	3	3	3	2	4	4	3.2
All Together Now	3	1	3	2	3	3	2.5
Round the Twist	3	1	0	3	1	3	1.8
Slip Sliding Away	2	1	2	2	2	2	1.8
It's Runny, Honey	3	2	3	3.5	3	4	3.1
Shaking All Over	4	2	2	2	3	3	2.7

Health, Safety, and Environment	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	5	5	5	5	5	5	5.0
How Dense are You?	4	4	4	4	5	5	4.3
All Together Now	5	5	4	5	5	5	4.8
Round the Twist	4	5	4	5	5	5	4.7
Slip Sliding Away	4	5	4	5	5	5	4.7
It's Runny, Honey	5	5	4	5	4	5	4.7
Shaking All Over	4	3	3	2	4	4	3.3

Ease of Setup and Takedown	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	5	5	5	5	5	5	5.0
How Dense are You?	1	2	4	3	2	3	2.5
All Together Now	4	4	5	4.5	5	5	4.6
Round the Twist	3	2	4	3	3	3	3.0
Slip Sliding Away	4	5	5	5	5	4	4.7
It's Runny, Honey	5	5	5	5	5	5	5.0
Shaking All Over	4	4	4	4	5	4	4.2

Ease of Transport and Handling	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	5	4	5	5	5	5	4.8
How Dense are You?	2	4	4	4	4	3	3.5
All Together Now	5	4	5	4	5	4	4.5
Round the Twist	4	4	4	5	4	4	4.2
Slip Sliding Away	3	4	4	4	5	3	3.8
It's Runny, Honey	4	4	5	4.5	5	3	4.3
Shaking All Over	2	3	2	4	3	3	2.8

Durability	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	3	1	3	2	2	3	2.3
How Dense are You?	4	4	4	4	3	3	3.7
All Together Now	3	2	4	2	4	4	3.2
Round the Twist	4	4	4	4	4	5	4.2
Slip Sliding Away	5	5	4	4	4	4	4.3
It's Runny, Honey	3	4	5	4	2	4	3.7
Shaking All Over	4	3	3	3	3	3	3.2

Cost	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	5	5	5	3	5	5	4.7
How Dense are You?	3	4	4	3	3	3	3.3
All Together Now	4	3	4	2	5	3	3.5
Round the Twist	2	3	3	3	3	3	2.8
Slip Sliding Away	3	4	3	3.5	5	3	3.6
It's Runny, Honey	3	5	5	4	4	3	4.0
Shaking All Over	3	3	2	2.5	3	3	2.8

Ease of Repair	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	2	2	5	1	2	5	2.8
How Dense are You?	2	2	3	1	1	2	1.8
All Together Now	2	2	4	2	1	4	2.5
Round the Twist	1	1	1	1	2	5	1.8
Slip Sliding Away	2	1	3	1	3	5	2.5
It's Runny, Honey	2	2	4	2	2	2	2.3
Shaking All Over	2	1	2	1.5	3	4	2.3

Initial Draw	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	2	2	2	0	2	3	1.8
How Dense are You?	3	4	3	5	5	5	4.2
All Together Now	3	2	3	4	4	3	3.2
Round the Twist	5	4	5	5	4	5	4.7
Slip Sliding Away	3	2	3	3.5	3	3	2.9
It's Runny, Honey	3	2	4	4	4	5	3.7
Shaking All Over	4	5	3	4	4	4	4.0

Student Interaction	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	0	1	0	1	1	0	0.5
How Dense are You?	1	5	2	4	2	5	3.2
All Together Now	3	1	0	4	2	2	2.0
Round the Twist	2	2	2	2	4	2	2.3
Slip Sliding Away	2	2	3	2	3	3	2.5
It's Runny, Honey	1	3	1	4	3	3	2.5
Shaking All Over	3	2	2	3	4	3	2.8

Novel Completion Process	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	1	2	2	1	4	1	1.8
How Dense are You?	3	3	3	3	5	5	3.7
All Together Now	1	2	3	2	1	2	1.8
Round the Twist	5	4	2	3.5	4	5	3.9
Slip Sliding Away	1	4	3	2	1	5	2.7
It's Runny, Honey	3	4	2	4	2	5	3.3
Shaking All Over	3	5	4	4	4	5	4.2

Relevance to Students / Familiarity of Concepts	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	1	3	1	2	3	2	2.0
How Dense are You?	2	3	3	2	2	2	2.3
All Together Now	4	3	3	3	2	4	3.2
Round the Twist	4	2	3	1.5	3	4	2.9
Slip Sliding Away	2	2	2	1	1	3	1.8
It's Runny, Honey	4	2	2	1	1	2	2.0
Shaking All Over	4	2	3	2	4	5	3.3

Results are Surprising / Unexpected	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	3	1	1	3	4	3	2.5
How Dense are You?	3	4	3	1	4	5	3.3
All Together Now	1	3	3	2	2	2	2.2
Round the Twist	4	3	2	4	4	4	3.5
Slip Sliding Away	1	2	2	2	0	3	1.7
It's Runny, Honey	2	3	2	2	1	4	2.3
Shaking All Over	2	3	1	2	1	3	2.0

Instructions / Expectations are Clear	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	5	5	4	4	4	4	4.3
How Dense are You?	2	4	3	3		4	3.2
All Together Now	5	5	4	4	5	5	4.7
Round the Twist	5	5	5	5	5	4	4.8
Slip Sliding Away	4	5	4	4	3	4	4.0
It's Runny, Honey	4	5	5	4	4	4	4.3
Shaking All Over	3	5	4	3	3	4	3.7

Intuitive to Complete	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	4	4	4	1	2	4	3.2
How Dense are You?	1	1	3	1	1	2	1.5
All Together Now	3	4	3	4	5	4	3.8
Round the Twist	4	4	5	4.5	4	4	4.3
Slip Sliding Away	4	4	4	4	4	4	4.0
It's Runny, Honey	3	3	4	4	5	3	3.7
Shaking All Over	3	4	4	2.5	3	4	3.4

Appropriate Length	Project Team	Educator 1	Educator 2	Educator 3	Educator 4	Educator 5	Average
Animated Earth	4	5	5	3	5	5	4.5
How Dense are You?	1	0	2	2	2	3	1.7
All Together Now	3	4	4	3	4	4	3.7
Round the Twist	4	5	5	4	5	4	4.5
Slip Sliding Away	4	5	4	3	4	4	4.0
It's Runny, Honey	2	4	3	2.5	5	3	3.3
Shaking All Over	3	4	4	3	4	4	3.7

Appendix O: Standard Deviations in Educator Program Evaluations

Standard Deviations	Animated Earth	How Dense are You?	All Together Now	Round the Twist	Slip Sliding Away	It's Runny, Honey	Shaking All Over	**Average**
Curriculum Relevance	0.89	1.30	0.55	0.89	1.48	1.34	0.45	**0.99**
Potential for Teacher Extension or Application	1.05	1.17	1.03	1.03	1.10	0.52	0.82	**0.96**
Encourages Student Thought	1.03	0.75	0.84	1.33	0.41	0.66	0.82	**0.83**
Health, Safety, and Environment	0.00	0.52	0.41	0.52	0.52	0.52	0.82	**0.47**
Ease of Setup / Takedown	0.00	1.05	0.49	0.63	0.52	0.00	0.41	**0.44**
Ease of Transport / Handling	0.41	0.84	0.55	0.41	0.75	0.76	0.75	**0.64**
Durability	0.82	0.52	0.98	0.41	0.52	1.03	0.41	**0.67**
Cost	0.82	0.52	1.05	0.41	0.80	0.89	0.42	**0.70**
Ease of Repair	1.72	0.75	1.22	1.60	1.52	0.82	1.08	**1.25**
Initial Draw	0.98	0.98	0.75	0.52	0.49	1.03	0.63	**0.77**
Student Interaction	0.55	1.72	1.41	0.82	0.55	1.22	0.75	**1.00**
Novelty of Completion Process	1.17	1.03	0.75	1.11	1.63	1.21	0.75	**1.10**
Relevence to Students / Familiarity of Concepts	0.89	0.52	0.75	1.02	0.75	1.10	1.21	**0.89**
Results are Surprising / Unexpected	1.22	1.37	0.75	0.84	1.03	1.03	0.89	**1.02**
Instructions / Expectations are Clear	0.52	0.84	0.52	0.41	0.63	0.52	0.82	**0.61**
Intuitive to Complete	1.33	0.84	0.75	0.42	0.00	0.82	0.66	**0.69**
Appropriate Length	0.84	1.03	0.52	0.55	0.63	1.08	0.52	**0.74**
Average	**0.84**	**0.93**	**0.78**	**0.76**	**0.78**	**0.86**	**0.72**	**0.81**

Appendix P: Score Ranges in Educator Program Evaluations

Ranges	Animated Earth	How Dense are You?	All Together Now	Round the Twist	Slip Sliding Away	It's Runny, Honey	Shaking All Over	**Average**
Curriculum Relevance	2.00	3.00	1.00	2.00	4.00	3.00	1.00	**2.29**
Potential for Teacher Extension or Application	3.00	3.00	3.00	3.00	3.00	1.00	2.00	**2.57**
Encourages Student Thought	3.00	2.00	2.00	3.00	1.00	2.00	2.00	**2.14**
Health, Safety, and Environment	0.00	1.00	1.00	1.00	1.00	1.00	2.00	**1.00**
Ease of Setup / Takedown	0.00	3.00	1.00	2.00	1.00	0.00	1.00	**1.14**
Ease of Transport / Handling	1.00	2.00	1.00	1.00	2.00	2.00	2.00	**1.57**
Durability	2.00	1.00	2.00	1.00	1.00	3.00	1.00	**1.57**
Cost	2.00	1.00	3.00	1.00	2.00	2.00	1.00	**1.71**
Ease of Repair	4.00	2.00	3.00	4.00	4.00	2.00	3.00	**3.14**
Initial Draw	3.00	2.00	2.00	1.00	1.50	3.00	2.00	**2.07**
Student Interaction	1.00	4.00	4.00	2.00	1.00	3.00	2.00	**2.43**
Novelty of Completion Process	3.00	2.00	2.00	3.00	4.00	3.00	2.00	**2.71**
Relevence to Students / Familiarity of Concepts	2.00	1.00	2.00	2.50	2.00	3.00	3.00	**2.21**
Results are Surprising / Unexpected	3.00	4.00	2.00	2.00	3.00	3.00	2.00	**2.71**
Instructions / Expectations are Clear	1.00	2.00	1.00	1.00	2.00	1.00	2.00	**1.43**
Intuitive to Complete	3.00	2.00	2.00	1.00	0.00	2.00	1.50	**1.64**
Appropriate Length	2.00	3.00	1.00	1.00	2.00	3.00	1.00	**1.86**
Average	**2.06**	**2.24**	**1.94**	**1.85**	**2.03**	**2.18**	**1.79**	**2.01**