Abstract

In modern societies, science is central to governance and democratic involvement. Without understanding science, leaders and citizens cannot effectively participate in debates on issues ranging from nuclear power to climate change. Scientific literacy is the establishment and assessment of benchmarks in public understanding of science. The dominant model is due to Miller [1983] and has three dimensions: core scientific content like “radiation” and “DNA”; understanding of the scientific process; and consideration of the relationship between science and society.

We have created a new instrument based on existing surveys for public scientific literacy. The instrument covers the three dimensions outlined above. It was administered to 274 first-year Physics students in three classes with differing secondary science backgrounds, and 28 Honours and postgraduate Physics students. The results show marked differences between the classes in the coherence and detail of their responses in the content and process dimensions, indicating that Physics education is effective at improving scientific literacy in those dimensions. However, responses for the society dimension were unexpectedly consistent across all four classes. We present these results, together with the instrument, and analysis techniques newly adapted for assessing scientific literacy in the university setting. These techniques demonstrate greater sophistication and sensitivity than previous analysis methods designed for use in large public studies.

Acknowledgements

Several people deserve plenty of thanks for making this thesis possible. Writing two Honours theses in consecutive years could have been an even more difficult task had it not been for their ongoing support. First, my supervisors Manjula Sharma and Ian Johnston have been tremendously helpful: flexible in accommodating my slightly unconventional topic, enthusiastic about the project's potential, and always wise in their advice along the way. Ian Sefton and Alex Hugman also provided excellent feedback on the survey design. The School of Physics deserves thanks for its strong support of Honours research: my friends in other faculties are always envious of my permanent desk in an office with a fridge and microwave. Even such a well-equipped workplace would have been uninviting, though, without the always-welcoming SUPER students: Christine, Helen, Nigel, Jira and now Simon. Thanks and good luck for your various endeavours next year. Finally, my deepest gratitude is reserved for my family, for their encouragement throughout another busy year, and for Bec, for always being interested and supportive even while working on a thesis of her own – we made it!

Statement of Contribution of Student

The idea for the topic and the formulation of the research questions were my own. In refining the broad topic of scientific literacy into a manageable Honours project, I was assisted by my supervisors, Manjula Sharma and Ian Johnston. I developed the first version of the survey instrument based on my own investigations of the literature. The SUPER research group provided feedback, resulting in some modifications to the wording and layout. I distributed the survey to students, with the cooperation of Manjula, Richard Thompson, Richard Tarrant and various lab supervisors. I carried out all data input.

I made the decision to apply the SOLO and thematic coding techniques, based largely on their successful application in a previous project by Helen Georgiou, who provided helpful advice. In particular, she and Manjula suggested the use of the NVivo software package, in which I received external training. In developing the SOLO analysis, I synthesised three separate draft coding schemes developed by Manjula, Ian and me into a single scheme, before refining it over several rounds of testing during which they provided comments. I performed all implementation of SOLO and thematic coding in NVivo, and all other analysis, including the calculation of the chi-squared statistics suggested by Manjula.

Statement of Originality

I certify that this report contains work carried out by myself except where otherwise acknowledged.

Michael West

Table of Contents

Introduction

Science is of central importance to governing and living in modern societies. Participating in debates on many major civic issues is impossible without an understanding of science and technology. Citizens may require domain-specific factual knowledge, such as where a certain understanding of the concepts of radiation and nuclear decay is needed to effectively discuss nuclear power and storage of radioactive waste. More generally, they must be sufficiently familiar with the scientific process to rigorously evaluate scientific evidence and distinguish it from pseudoscience, as in the debates on climate change. Further, fostering a public that understands and values science is important in addressing the trend in some jurisdictions of removing key scientific results, such as evolution and the age of the universe, from school curricula. This goal of having the public achieve particular benchmarks in scientific understanding is known as scientific literacy.

Formal science education, at both secondary and tertiary levels, is typically considered a key conduit for raising scientific literacy. Other mechanisms, like exposure to science media or visiting science museums, also play a part, but have been shown to be responsible for less of the variation in scientific literacy within populations [Miller 2004]. The ability of science education to contribute to building a scientifically literate public has also been assumed in the Australian government's designation of science, mathematics and statistics as 'national priority' courses. Given this background, it is important to verify more precisely the ways in which tertiary science education affects scientific literacy over time. This is particularly true given the relative flexibility that many university instructors have in setting their courses; many may be unaware of scientific literacy, or how it is defined and assessed, and thus not incorporate these ideas into their teaching. Additionally, much teaching is focused on core science content, while scientific literacy encompasses a broader range of knowledge and understanding, so it may develop unevenly.

Building an evidence base to answer these questions is important if scientific literacy is to be successfully incorporated as a goal of tertiary science education. Applying rigorous standards of investigation to teaching as well as research is vital to the long-term success of the scientific enterprise. This thesis takes that approach in determining how students' scientific literacy changes over the course of their physics education.

Survey of Literature

Scientific literacy as a modern research field was born in the late 1950s in the United States. Public and political consternation over the Soviet launch of Sputnik created a feeling that American science and science education needed revitalisation [Hurd 1958]. Students were encouraged to study science, technology, engineering and mathematics, as a scientifically equipped population was considered vital for long-term prosperity and to maintain public support for funding science and spaceflight programs. For the next two decades, educators developed numerous disparate theories of what level of scientific knowledge would be deemed sufficient in an educated citizen. These conceptual debates, discussed further below, meant that most of the literature from that period focused on definitions and motivations. Because of the theoretical disagreement, attention was not focused on actually measuring scientific literacy in sample populations.

Later, Miller [1983] proposed a multi-pronged framework, with three dimensions of scientific literacy defined as *content* (facts and vocabulary); *process* (knowledge of the scientific method); and an understanding of the relationship between science and *society*. Surveys of the general public followed, to determine their level of literacy under such a framework. This synthesis proved influential and much work since then has adopted its approach, especially national studies commissioned by governments and academies of science in the United States, Europe and Japan [e.g. Miller 2004, Durant et al. 1989]. The results of these and other studies will also be discussed in more depth below. There has never been a national study of adult scientific literacy in Australia, and even at the smaller scale addressed by this study, they remain very few in number.

Theoretical work on scientific literacy is extensive, commensurate with its status as a widespread "goal of contemporary science education" [Kanasa 2008]. Unfortunately, as Miller [1998] observes, "the debate that has occurred has been primarily at the conceptual level with little or no empirical testing of these conceptualizations". Significant discussion of what the public should know about science, how to define 'the public' itself, and how to measure their understanding of the chosen topics, has not been matched by detailed examination of the literacy of particular groups under the different frameworks proposed. The studies that do exist have largely occurred on broad scales, using samples intended to be nationally representative. This has limited researchers' ability to draw fine-grained conclusions about the relative scientific literacy of different groups within society. In particular, although tertiary science education has been established as being broadly positively correlated with scientific literacy [Miller 2004], closer study of how literacy develops within science students while at university has not been undertaken in depth.

Laugksch [1998] offers a useful conceptual overview of the field, endeavouring to explain why it has sometimes been seen as "ill-defined and diffuse" by noting five background factors that influence its meaning. These are: the different interest groups involved, the different conceptual definitions, conflicting views on whether it should be assessed in an absolute or relative manner, different ideas on why literacy is desirable, and different approaches to measurement. The first factor, interest group membership, influences each researcher's perception of their audience and what areas contribute to literacy. Science education researchers generally focus on children and young adults, heavily influenced by the science curriculum; science communicators (such as science museum curators) must consider less structured factors outside of the classroom; and social scientists work more with the general adult population, whose exposure to science is often through the media and built on the scaffolding of their previous formal education.

The second factor, conceptual definitions, entails debate surrounding what level of scientific understanding is required to be considered literate. Citizens fulfil multiple roles—as voters, consumers of media and products, employees, and so on. Each of these requires science to be applied in different ways. Consequently, different types of literacy have been proposed, ranging from 'civic literacy' [Shen 1975, Miller 1987], which emphasises an ability to understand public debate on scientific issues, to 'practical literacy' [Shen 1975], which focuses on using scientific thinking in everyday life, and bears a strong resemblance to the 'minimum science' approach taken in many developing countries [Popli 1999].

In this connection, it is important to note that scientific literacy is closely related to other fields such as public awareness of science (PAS) and public understanding of science (PUS). Some argue that literacy and understanding are essentially synonymous (for example, Laugksch [1998]), but others draw a finer distinction. Burns et al. [2003] frame scientific literacy as an ideal end state that combines both awareness and understanding. They suggest the following distinctions: PAS focuses on developing an initial awareness and positive attitudes; PUS seeks to instil understanding of scientific content and processes; and literacy is the combination of these two areas, such that citizens maintain an interest in science and can form considered, informed opinions on scientific issues. They view it as a "very high objective" that "relatively few people will achieve", considering it to be a desirable goal rather than a practically feasible target. This pessimism is shared by Shen [1975], whose "true scientific literacy" is essentially at undergraduate physics level, such as understanding the laws of thermodynamics. However, although it is clear that the population at large will not achieve university-level science knowledge, it is unlikely that this high level is genuinely required to qualify as scientifically literate for the purpose of general civic understanding of science-related issues. Accordingly, the benchmark for scientific literacy may well be lower, and hence more achievable, than Shen and Burns et al. suggest.

This leads to Laugksch's third point of contention: whether scientific literacy should be assessed in a relative or absolute manner. An absolute conception of literacy would indicate that there is a particular universal standard of knowledge that must be acquired by all people in all societies. In contrast, a relative literacy framework would suggest that since the point of literacy is to function effectively within a particular role in a particular society, it must be assessed in that context. The level of science needed for democratic participation is different to that needed for employment in a technical occupation, and they are both different in an agrarian society to one with an advanced information economy. This concept of relative scientific literacy seems to be a more useful definition. Even under a relative framework, the content that people need to know might still be decided by experts in a prescriptive manner. However, other approaches do exist to help determine what content the public should be familiar with, such as extracting the most commonly used scientific terms in the media [Brossard 2006].

The fourth contentious factor surrounding scientific literacy is the set of different purposes for advocating it as an educational goal. Given their background as scientists and educators, there is an intuitive link between researchers measuring literacy and desiring to improve it. Their particular motives for valuing improved literacy often influence the way it is defined. Laugksch categorises them into two schools. The 'macro' view is that literacy is needed for the good of the nation; most economies depend heavily on science and technology, so a supply of scientifically-literate employees is required. This view also includes the belief that a population that is more knowledgeable about science will continue to support its funding at higher levels than a scientifically illiterate one (for example, [Durant et al. 1989]). As a result, their definition is more likely to focus on aspects of science required for civic engagement and cultural appreciation. The 'micro' perspective values scientific literacy for its ability to improve the lives of individual citizens, by helping them to make more rational choices (for example, when evaluating products or medical treatments [Miller and Kimmel 2001]) and accessing science as a cultural achievement of humanity. Consequently, supporters of this view are more likely to define scientific literacy in terms of a level of understanding for using science practically in everyday life. There is, however, significant overlap: for example, improving scientific literacy to enable citizens to take highly skilled, technology-intensive jobs benefits both the nation and those individuals.

The final and perhaps most significant conflict identified by Laugksch, and also discussed by others, is on ways to measure scientific literacy. Multiple instruments have been developed, with correspondingly many schemes for assessing answers and setting thresholds for scientific literacy and illiteracy. Knowledge of content may be assessed by multiple-choice questions with straightforward answers (the approach taken by Durant et al. [1989]), but open-ended questions may also be used to probe understanding more deeply. Multiple-choice questions are also typically employed in assessing the process and society dimensions. Open-ended questions require a carefully designed coding strategy to determine how to allocate a degree of correctness to each answer, and how to achieve this reliably and repeatably between coders. Expert opinion is frequently employed to set the scope of content and prescribe the benchmark levels that constitute literacy. Miller [2004], for example, sets a base score of 70% on a combined content-process test, and argues that anybody below this mark will have difficulty assessing scientific information and should thus be classified as scientifically illiterate. The theoretical basis of this exact figure is unclear.

Rather than the binary literate-illiterate approach, scientific literacy may be represented as a more continuous spectrum. Koballa et al. [1997] recommend a seven-level framework. The levels range from total scientific illiteracy (failure to recognise words as science-related), through correct definitions of scientific terms, to an understanding of the scientific process, and finally a broad knowledge of the history of a scientific discipline and the connections between its major ideas. They argue that attaining the higher levels is both possible and desirable, but suggest that secondary teaching is currently focused too heavily on "vocabulary and isolated facts", hampering students' ability to develop a cohesive framework linking ideas within a discipline, such as physics.

Koballa et al. also argue for the need to consider that scientific literacy is highly domain-specific, with literacy in one discipline not guaranteed to provide expertise elsewhere, due to the specialised nature of modern science. An expert biologist may not have an above-average understanding of the content or history of physics, for example, although their understanding of the scientific process will typically exceed that of the general population. In this regard they are joined by Levy-Leblond [1992]: "Indeed, given the current state of scientific specialization, ignorance about a particular domain of science is almost as great among scientists working in another domain as it is among lay people." Experimental support has been provided by Rennie and Stocklmayer [2003], who report on a survey of 193 practising scientists. The survey used was originally developed by Durant et al. [1989] for use on the general public. It revealed that scientists themselves do not achieve universal scientific literacy across all disciplines: "Not one of the scientists felt completely confident of their answers to every question and many admitted frankly to ignorance of several questions outside their discipline. There was no question on which all the scientists were correct."

Miller has long argued that the definition of scientific literacy should be the level of knowledge required to read and comprehend science sections in high-quality media, such as the New York Times in the United States or similar international equivalents [Miller 2004]. As a result, the content dimension figures strongly; he suggests that because fundamental scientific terms are often used without explanation in such media, the reading public must "rely on previous formal and informal education for their basic inventory" of such concepts. This inventory includes simple items from many disciplines, such as 'radiation', 'molecule', and 'DNA'. Without a correct understanding of these terms, it is easy to see how the public's consumption of science from media would be strongly hindered. For example, consider 'radiation'. Miller cites a 2000 national study showing only 11% of US adults were able to give a scientifically correct general definition of radiation, in terms of emission of energy or particles from a source. A further 26% can name a source, and 11% can name effects of some types of radiation, but neither of those groups could explain further. The large proportion of adults not able to explain radiation even in basic scientific terms would be prone to severe misunderstandings when analysing media that referred to it; for example, it is plausible that they might mistake the electromagnetic radiation emitted from a new mobile phone tower as being identical to ionising radiation generated by nuclear decay. This illustrates the importance of scientific content knowledge in civic participation.

Miller's work also covers the process dimension, examining levels of understanding about the nature of scientific inquiry in the United States. In this area he builds on the early work of Davis [1958], who found that only 14% of US adults could correctly explain the distinguishing features of scientific methodology, such as recourse to experiment, hypothesis constructing and testing, and objective evaluation of results. Half of respondents indicated that science is thorough and careful but did not add specifics. Repeating the question, "Some things are studied scientifically; some things are studied in other ways. From your point of view, what does it mean to study something scientifically?" in periodic studies over the course of the past half-century has shown that the proportion of adequate responses had risen to about 21% by 1999. However, using an open-ended question like this requires careful design of a coding strategy for responses, and some simplistic but technically correct answers (such as merely writing "an experiment") were coded as adequate. To probe public understanding of the conduct of experiments, a further question was added in several studies; it offered various choices for an experimental design, with the correct answer being the one that used a control group. Between 12% and 25% of people were successfully able to pick and justify their choice of that design, indicating that understanding of exactly what constitutes a reliably designed experiment is similarly limited.

Durant et al. [1989] conducted similar studies of adults in the United Kingdom. They showed that a lack of understanding is not due to simple disinterest. Topics from science, technology and medicine were rated highly (above sport, films and politics) when people were asked which of a set of sample headlines they would be likely to read, even before knowing that the study was focused on science. Respondents also rated themselves as highly interested in scientific topics, but only moderately informed about them. When asked the same question about scientific process ("In your own words, what does it mean to study something scientifically?"), the results were similar to the American findings. 3% made mention of theory construction, 11% referred to conducting experiments, while 86% had no answer or an unsatisfactory one. However, Durant et al. caution that asking such an open-ended question "ignores the possibility of tacit public understanding of the processes of science" that they are not able to effectively communicate. In a multiple-choice question, just over half of respondents did indeed select an experimental approach over alternate options. Unfortunately, public understanding of important terms in describing the scientific process, like 'theory', is still low. Respectively 33% and 45% of people were able to describe the theories of relativity and evolution as "a well established explanation" rather than "a hunch or idea" or "a proven fact".

Many studies have focused on the process and society dimensions of scientific literacy outside the context of a prescriptive literacy framework. Instead, studies have been situated in related fields such as ‘Nature of Science' (NOS) where the emphasis is on capturing different possible understandings of the scientific process rather than making a summative judgement about whether they are correct on the whole (see, for example, [Martin-Dunlop 2004]). Methodological debates have resulted in a proliferation of instruments. Many early instruments (e.g. the 1966 Science Process Inventory) used only Likert scales or multiple-choice questions, which may have returned data that was "more likely an artifact of the instrument in use than a faithful representation," of respondents' views, as Lederman et al. [2002] argue. As a result, more recent studies have "embraced highly interpretive qualitative studies involving small sample sizes" [Martin-Dunlop 2004], including techniques like follow-up interviews. However, such time-consuming data collection measures are not always practical, and compromise approaches have been developed. These include the Views on Science-Technology-Society survey (VOSTS) [Aikenhead et al. 1992], a set of multiple-choice questions with a large number of possible responses for each, which were empirically derived by synthesising the views presented in open-ended paragraphs written by secondary students.

The most significant study to date of the effect of tertiary education on scientific literacy as a whole was conducted by Miller [2004]. By adding a demographic question to a public adult survey of civic scientific literacy, he was able to group all respondents into three broad categories based on how many tertiary science courses they had taken while at university. This parameter was found to be the strongest predictor of literacy levels, having nearly double the positive correlation compared to age, general level of educational attainment, and self-rated use of informal resources like museums and science magazines. Miller thus attributes the United States' relatively strong performance in scientific literacy at an adult level to its imposition of general education requirements that mandate some science even for non-science majors. However, because this study was conducted with adults rather than university students themselves, there was no opportunity to investigate finer distinctions, such as on the effect of studying one science discipline in depth versus the same number of courses distributed over a range of areas. Scientific literacy at the tertiary level remains an under-studied area.

Norris et al. [2003] investigated Canadian students' ability to engage with science-focused media articles, by asking them to assess the truth and certainty of various claims and identify causal relationships after reading. The sample of first-year psychology students answered less than half of these interpretive questions correctly. This performance was very similar to a previous study of high school students, despite the university students having received significantly more science education. Norris et al. concluded that such education "seemed to have very little to do with these important tasks associated with life-long learning of science and democratic citizenship". They also found that students are likely to overestimate their ability to analyse these media reports; Norris et al. thus advocate more specific training in reading and interpretation of scientific media articles as part of science curricula. This does not currently occur to any significant degree; Hobson [2008] has remarked that "Scientific literacy has the lowest possible priority in most college physics departments, if indeed it is taught at all."

Robinson and Crowther [2001] studied a subset of scientific literacy, focusing on the environmental domain only. Their instrument was a short survey, with questions that were essentially content-based with unambiguous correct or incorrect answers, but ranging from pure science (e.g. the definition of biodiversity) to science in society (e.g. identifying the prevailing method of nuclear waste disposal). They found that US students studying to become science teachers performed better than either biology or chemistry majors, although the difference was quite small and only statistically significant for the comparison with chemistry. Results from the same survey had also been obtained from a poll of the general adult population, and no statistically significant difference was found between science students and this broader average. These findings cast doubt on the intuitive notion that science study leads directly to higher scientific literacy, and reinforce the domain-specific nature of literacy.

Oliver's [2008] study of a single group of first-year humanities students is the closest approach to an assessment of adult or university-age scientific literacy that has been conducted in the Australian system. These students, none of whom were science majors, were asked a mix of content and process questions. The process component was scored by counting the number of terms relevant to the scientific process, such as 'experiment' or 'unbiased', in open-ended responses to the typical question, "What does it mean to study something scientifically?" The mean result was 1.3; using Oliver's criterion that three or more terms constituted a scientifically literate answer, 15% of the group were literate by this measure. The content dimension was assessed by multiple-choice questions on topics such as tectonics, probability, and buoyancy. No open-ended or definitional questions were asked. The mean result was 12 on a scale from 5 to 15, but Oliver does not make a judgement about the level that would constitute literacy within her framework. This analysis, while worthwhile, was limited in two ways: the word-counting method is a relatively simplistic way of assessing understanding, and the study did not include any science students in order to test the effectiveness of science education. There is room for a more detailed investigation of the scientific literacy of science students themselves.

Research Questions

The above discussion has identified that while there is increasing consensus on a viable definition of scientific literacy, in terms of Miller's three dimensions of content, process and society, there remain significant gaps in our knowledge of literacy levels. Consequently, we have limited understanding of the effectiveness of many efforts to raise literacy made by scientists and science faculties. If scientific literacy is to be a goal of science education, as is well accepted [Kanasa 2008], it is important for science faculties to engage in rigorous evidence-based analysis to determining whether they are succeeding. In that context, the goals of the present work are two-fold.

First, we aim to develop an instrument suitable for assessing scientific literacy within a university student demographic. This instrument must be sensitive to the constraints of data collection for that group, where contact with survey recipients is limited, in comparison to the detailed sets of questions that may be asked of adults via phone polls or longer interviews. It will focus primarily on the first two dimensions of literacy, content and scientific process, as these have the strongest degree of consensus as constitutive elements of literacy; but given the current prominence of the relationship between science and society, some parts of the instrument may also assess this third dimension. The instrument will be developed in accordance with established methodology for science education research, such as data-driven refinement and validation for self-consistency.

Second, we will employ this instrument to obtain new data on scientific literacy within various classes of university students, for the first time. This data will be used to study the extent to which science education in Australia affects scientific literacy, in terms of each of its dimensions and as an integrated whole. By conducting this analysis on data drawn from students, rather than simply sorting the general adult population into university graduates and non-graduates, we will be able to analyse more closely the evolution of scientific literacy throughout the educational process.

This may be expressed as the following two research questions:

1. ***What features are required for an instrument that can effectively assess scientific literacy in a university setting?***
2. ***How do students with differing levels of secondary and tertiary physics perform across the three dimensions of Miller-type scientific literacy?***

Methodology

Instrument development

Existing instruments for measuring public scientific literacy have typically been administered by phone interview to large cross-sections of the general adult population. There are also a range of lengthy paper-based surveys, such as the Views of the Nature of Science (VNOS-C) instrument, which may be administered in combination with interviews [Lederman et al. 2001]. Administering one of these long surveys was not feasible in this study, due to the time constraints imposed on data collection in this university setting. Accordingly, a new instrument was devised, with reference to the literature but remaining sensitive to the need for brevity. It was intended primarily to address the content and process dimensions of scientific literacy, which are more widely recognised in previous work as requisite components of any measurement. The society dimension, which measures the respondent's attitudes towards the place of science within society, was included in a more limited context, as it is relatively difficult to assess whether such attitudes are objectively 'correct'.

The full instrument as distributed to students is reproduced in Appendix A.

Content (radiation), open-ended

The content dimension is typically assessed by determining respondents' familiarity with a set of core scientific terms. The set of such terms varies from instrument to instrument, but Miller [2004] lists several terms that have been examined in many previous surveys. They are generally chosen by expert opinion (see Rutherford and Ahlgren [1990]), although alternative approaches have been explored, such as finding scientific terms that appear frequently in the media [Brossard 2006]. Understanding of the terms is tested by asking for a definition or by requiring the respondents to answer various multiple-choice or true-false questions about the term. Definitional questions may be open-ended, where the term is given and the respondent asked to define it, or closed, where the definition itself is given, and the respondent must recognise and provide the word being defined. Open-ended definitions generally reveal more accurate information about the respondent's depth of knowledge [Bell and Lederman 2003], so this form was employed in the instrument.

The term chosen for assessment was "radiation", as it is a key concept in the academic study of physics. It also appears frequently in the media; a Factiva news database search returned almost 100,000 references to radiation for each of the last four years. It is also relevant to public policy issues and civic engagement with them: for example, it could be argued that citizens should be able to distinguish between the different types of radiation emitted from a nuclear power plant and a mobile phone tower. These differences are reflected in the multiple ways that radiation can be understood. The broad physical definition is in terms of the transfer of energy away from a source via the emission of particles or waves. Most everyday usage refers to the context-specific case of ionising radiation produced by nuclear activity. The broader "energy-type" definition might be considered to be more expert than the "nuclear-type" definition, as it indicates a greater ability to explain the underlying concept in the way a practising physicist might, rather than simply naming or describing a source of radioactivity. The wording used for this question was simply: "In your own words, how would you define radiation?"

Process, open-ended

To assess the second dimension, understanding of the scientific process, another open-ended question was used. This question asked for a description of what it means to "study something scientifically", drawing a contrast with topics that are "studied in other ways", such as philosophically. Any particular specific definition of scientific study would be highly contested and it is not our aim to explore that topic in depth. Indeed, it is a detailed focus on these aspects that makes some prior instruments (such as the previously mentioned VNOS-C) too unwieldy to deploy in our university classroom setting. Nevertheless, the scientific process has several commonly accepted elements, such as the need for experimental verification of hypotheses, a focus on objectivity, and so on. The question in our instrument anticipates that these concepts would appear in a scientifically literate response. Some responses might also include other features, such as peer review and reference to theory, that are not necessarily unique to science but are closely linked with it. Less sophisticated respondents may focus on factors such as hard work or detailed analysis, which might form part of science but certainly do not define it with respect to other fields.

The wording of the question was: "Some things are studied scientifically; some things are studied in other ways. In your own words, what does it mean to study something scientifically?". This was derived directly from the literature. It has appeared almost unchanged in multiple surveys since 1957 (see Miller [2004] for a full listing), and has gained authority as a benchmark through which new results from different populations may be compared. These studies have employed diverse methods of analysing the correctness of answers and hence the scientific literacy of respondents; we apply their methods to our dataset below as well as our own new method of analysis.

Society, Likert scales

The remaining four questions addressed the society dimension. Each question addressed a situation in the practice of science, and allowed respondents to indicate when they thought pure scientific ideals would be upheld, and when they would instead be disrupted by social influences, such as funding or the media. This theme was particularly prominent in the last question, which directly asked respondents to reflect on political distortions of science.

All four questions were adapted from the widely-used Views on Science, Technology and Society (VOSTS) instrument of Aikenhead and Ryan [1989]. This is a set of dozens of multiple-choice questions, each having as many as nine possible choices. The available choices were synthesised from themes commonly expressed in the open-ended responses of a cohort of Canadian secondary students initially presented with the questions. This instrument required significant modification for two reasons. Its length makes it impractical to administer, so a subset of questions were chosen for incorporation in our instrument. These dealt with: objective reporting of scientific results; scientific decision-making in the acceptance of a new theory; disagreement among scientists on a current theory; and the previously mentioned question about political influence on science.

The second disadvantage of VOSTS is that multiple-choice surveys inevitably constrain respondents [Lederman et al. 1998], forcing them to choose one description when their true opinion might be a combination of several of the responses available. To improve the quality of our data, the multiple-choice questions were adapted into a set of grouped Likert scale response questions. For each question, respondents were presented with a situation and asked to rate the relative importance of various factors affecting scientists' behaviour. For example, one question presents the situation of scientists disagreeing on a theory, and asks respondents to weigh the relative influence of incomplete evidence, differing scientific interpretations, scientists' personal beliefs, and external funding. Each of these factors roughly corresponded to the main themes in the multiple-choice options from the original VOSTS question. In this way, respondents were able to indicate if more than one factor was important, rather than selecting only a single choice.

Demographic information

Several pieces of demographic information were solicited to enable the causes of any observed differences in scientific literacy to be investigated. These were: level of tertiary physics education, level and areas of secondary science education, age, and gender.

Development and data collection

An initial instrument was created, with reference to previous tests presented in the literature. This draft instrument was subjected to validation by expert opinion, through a panel of science education researchers with physics expertise, and revised accordingly. After collecting one set of data, from a class of first-year Fundamentals Physics students, some minor revisions to wording were made for clarity. The effect of these was judged to be sufficiently minor that the results from both versions of the instrument remain comparable to each other. The revised instrument was distributed to first-year Regular and Advanced Physics streams and a group of Honours and PhD Physics students, whose demographics are summarised in Table 1. Human ethics approval was received for all survey distributions and is reproduced in Appendix B.

Class	**N**	**Year**	**Prior physics study**	**Typical degrees**
Fundamentals	91	1	Minimal secondary	BE, BMedSci
Regular	132	1	Some secondary	BSc, BE, BMedSci
Advanced	50	1	Strong secondary	BSc (Adv)
Postgraduate	28	4+	Completed undergraduate	Honours, PhD

Table 1: Sample demographics. Note that the Postgraduate grouping also includes Honours students.

Methods of data analysis

The open-ended definitional questions were analysed in several ways. We developed two main schemes, and also applied methods from the literature for comparative purposes. Our approaches were based on two schemas: the Structure of Observed Learning Outcomes (SOLO) taxonomy, and thematic coding.

Structure of Observed Learning Outcomes (SOLO)

An important feature of an open-ended response is the extent to which it demonstrates understanding through the use of fluent expression and connected ideas. A response that simply lists terms in minimal detail, without showing how they are linked, demonstrates a lower level of understanding than a response that integrates the same concepts into a cohesive overall definition. This is particularly true when aiming to distinguish between responses that use a particular specific situation to illustrate a general definition (e.g. gamma rays as a type of energy emitted from a source, in the radiation question) and responses that conflate the two and use the specific case as the definition itself (e.g. describing gamma rays only).

To obtain this information, we adapted the Structure of Observed Learning Outcomes (SOLO) taxonomy [Biggs and Collis 1982]. This is a five-band system used to describe learning progress in a particular field. It has previously been applied to studies in biology [Lake 1999] and chemistry [Hodges and Harvey 2003]. Responses classified in higher bands have demonstrated improved ability to express and connect scientific ideas. A generic SOLO taxonomy is outlined in Table 2. Our study focused on specific definitions of concepts rather than student learning progress in a whole field. For this reason, we removed the "generalising to a new domain" criterion for a response to be classified in the uppermost band, “extended”. Instead, responses that were more complex and detailed than “relational” responses were placed in the “extended” band. Note that from here onwards, the quotation marks around the band labels will be dropped for simplicity.

Band	Name	Sorting criteria [Boulton-Lewis 1994]
5	Extended	Integrated knowledge is generalised to a new domain
4	Relational	Relevant aspects are integrated into an overall structure
3	Multistructural	Several relevant independent aspects ... not integrated into [a] structure
2	Unistructural	One relevant aspect is understood and focused on
1	Prestructural	No evidence of any knowledge of the processes involved

Table 2: The generic SOLO taxonomy used to quantify educational achievement in a domain. The prestructural band contains the least developed responses; the extended band the most developed ones.

The bands for ranking responses to our open-ended questions, and their correspondence to the SOLO system, were determined by having the researcher and two experts independently attempt to sort sets of responses into categories, without an *a priori* ranking scheme. Three slightly different schemes were developed, and reconciled into a single ranking system that was then applied to all the data. During this process, distinguishing features emerged from the data—such as the distinction between a single and multiple ideas, or the inclusion or lack of connecting explanations—that mapped well to the SOLO bands, indicating that SOLO could be adapted into a descriptive theoretical framework. Examples of how the band criteria are applied and sample responses, for both the radiation and process questions, are shown in Table 3.

The SOLO taxonomy also has the advantage of providing a simple numerical rating of a response's general quality. This is useful in our cross-sectional study, because it enables the direct comparison of the level of answers provided by different demographic groups. It is also well suited to use under a relative scientific literacy framework, as a different SOLO band might be deemed acceptable for groups with different backgrounds or literacy requirements. However, it is more difficult to use unambiguously than approaches based on searching for references to specific ideas in the response (thematic coding, discussed below). This is because SOLO requires the researcher to make a judgement about each response, rather than referring to a more sharply defined list of keywords. This issue can be mitigated by the use of an inter-coder review, ensuring that several independent judges allocate the same SOLO rankings to a given set of answers. For this reason, our SOLO sorting criteria were examined multiple times in a group setting to check that they were transparent and logical.

Thematic coding

Different respondents' definitions of the same term will often repeatedly refer to one or more particular themes. For example, many definitions of the scientific method will refer to the themes of "experiment", "hypothesis", and so on. Identifying these themes, especially when one or more of them are repeatedly observed within a particular group of responses, gives insight into how respondents understand the term in question. This process is known as thematic coding. By unambiguously assessing the use of particular thematic terms that appear in a longer definition, it enables rigorous quantitative analysis of qualitative data.

This approach was applied to both of the open-ended questions. Each response was analysed to see if it contained any of several thematic clusters known as “nodes”. The list of nodes was a combination of those that commonly appeared in responses (generative coding), and those that might be expected to appear based on expert definitions. A particular node might incorporate multiple words with very similar meanings. For example, a response would contain the "emission" node if it contained related words such as "given off" or "release", even without specifically mentioning the word “emission”; while the "experiment" node would also include references to "empirical", "observations", and so on.

Table 3: SOLO sorting criteria for open-ended questions, with sample responses for each band

Level	**Sorting criteria**	**Q1 (Radiation) samples**	**Q2 (Scientific process) samples**
Prestructural	No answer / Don't know / Very little or no scientific content	• *Dangerous. K-19* • *X-ray, computer*	• *Looking at the origins of waste and also how things are essentially created* • *To consider it is specific scientific terms, values and point of views*
Unistructural	Single idea, possibly in combination with other incoherent statements	• *Energy produced from the sun* • *Radiation is the spontaneous emission of radical atoms* • *Radiation, involves any substance 'projecting' or emitting particles. e.g. an oven radiates heat.*	• *On the basis of experiments and surveys* • *To have a more accurate prediction (theory)* • *To study something at its simplest terms and determine problems and theories to an exactness.*
Multistructural	Two or three ideas, with little explanation of each	• *Energy which is emitted from a body* • *The emission of energy from decaying elements.* • *Electro-magnetic wave / particle that is given off*	• *To analyse in depth, to apply precision, to provide a description, mathematical evidence* • *To have a written purpose without bias judgement. It is factual and evidence can be used to support claims. It is relatively rational.*
Relational	Cohesive definition with links between ideas, but perhaps partially incorrect or incomplete	• *Radiation is an electromagnetic energy given off by atoms of an element. Electrons give off radiation when they fall back to ground state / stable state from excited state.* • *The travel of energy or sub-atomic particles, usually uncontrollable and/or unwanted. Most commonly associated with nuclear activity.*	• *To perform an experiment to find results in a manner that can be repeated and verified.* • *To study something empirically and reliably to support theories.*
Extended	Cohesive packet of linked ideas described in greater depth, possibly using examples	• *The process of dissipation of internal energy via expulsion of electromagnetic waves, or, in the case of nuclear radiation, small particles (or e/m waves, which indeed themselves are made of photons).* • *Radiation refers to particles (photons, leptons or baryons) emitted by a source due to nuclear or atomic reactions or events.*	• *Scientifically, usually associated with a defined structure ie. aim, method plan, hypothesis, repetition of trials, comparison to known results and an evaluation.* • *To approach a problem by theorising, gathering info & data, experimenting, and modelling all these into a final theory or conclusion.*

While there may be minor distinctions between these related terms grouped within each single node, for the purposes of this study they are similar enough to be classified together when considering the respondent's level of understanding. For the radiation content question, the nodes were "emission", "nuclear", and "energy transfer". The "nuclear" node was applied to responses that defined radiation solely as ionising radiation, by making reference specifically to unstable nuclei, alpha / beta / gamma radiation, and so on, without also including a more general definition. The "emission" and "energy transfer" concepts covered those terms and close variations thereof.

For the scientific process question, the nodes were "experiment", "objectivity", "hypothesis", "falsifiability" (of hypotheses), "theory" (in the sense of making reference to a body of prior work), and "repetition" (of experiments). Additionally, a "tools" node was applied to responses that superficially described typical activities a scientist might engage in, such as using special equipment or mathematics, without actually defining the scientific process or how those activities contributed to it.

In summary, the application of thematic coding to this study allowed us to recognise the key themes that respondents used in defining both the scientific process and an example of scientific content, radiation. By examining which themes are particularly prominent in the explanations given by each demographic group within the dataset, insight about their scientific literacy levels was gained.

External comparisons

To compare our methods to those used in prior studies, we applied four other analysis schemes from the literature to our dataset: one for the content question and three for the process question. All four schemes were based on identifying key nodes in the manner of thematic coding. However, where thematic coding records all of the nodes present in each response for descriptive purposes, these schemes from the literature would use them to make a value judgement about the response's quality, either by counting the number of concepts used (Oliver's approach), or by deciding that referring to certain concepts showed superior understanding (the approach taken by Miller, and Durant et al.).

Content scheme, Miller

Miller's two analysis schemes [2004] utilised a ranked hierarchy of concepts, with some concepts being seen as more correct than others. The first scheme assessed definitions of radiation on a similar open-ended question to ours. Miller grouped those responses into four bands, as shown in Table 4. The implication was that each band demonstrates superior scientific content literacy to the ones below it. This is in contrast to our approach to thematic coding, where key concepts are identified solely for descriptive purposes, with the task of assessing the overall quality of an answer being answered by the integrative SOLO taxonomy.

Band	Sorting criteria [Miller 2004]
4 (4a,4b)	Originally "Able to provide an explanation that involved the emission of energy as particles or waves from a material or source", this band was split into two for clarity: a) "Emission of energy, including as waves or particles"; b) "Emission of waves or particles, with no reference to energy".
3	"Able to name a source of radiation but not explain it".
2	"Able to mention the effect of radiation but were unable to name a source or explain the meaning".
1	All remaining answers.

Table 4: Miller's analysis scheme for open-ended definitions of radiation.

Process scheme, Miller

The hierarchical approach was even more apparent in Miller's second scheme, applied to the same scientific process question as used in our instrument. As reproduced by Bauer and Schoon [1993], this scheme had five bands, shown in Table 5.

Band	Sorting criteria [Miller, as reproduced by Bauer and Schoon 1993]
5	"Theory construction and testing" … "studied in the context of a theory about the problem/phenomenon being examined, and/or that the study is an attempt to disprove a hypothesis about the nature of the problem/phenomenon being studied"
4	"To undertake experiments / tests" … "to carry out experiments or tests in a strictly controlled way (but this may be implied rather than specifically stated)"
3	"Open, in-depth exploration of phenomena/problem to be examined" … "evaluating a problem in an unbiased/open-minded way, taking into account all possible information"
2	"To measure or classify/no mention of any rigour" … "may describe a study in terms of concrete actions used by scientists (e.g. use a microscope or telescope)"
1	All remaining answers, with subgroups for blank responses and so on.

Table 5: Miller's analysis scheme for open-ended definitions of studying something scientifically.

There are obvious similarities between Miller's set of concepts and ours. However, because of his requirement that no response be classified into more than one band, this scheme is a hybrid between thematic coding and a ranking system. The bands were assessed in descending order from the top. For example, if a response referred to "theory" or "hypothesis", it would be placed in band 5 (theory construction and testing), and the assessor would move on to the next response. Whether the response referred to using experiments to test those hypotheses, or that the enterprise must be objective, would never be considered. This approach was therefore limited because it could not demonstrate how all of the different concepts which constitute the scientific process are perceived by the respondents. In contrast, our scheme allows any given response to be linked to all of the concept nodes to which it refers. This still allows us to prioritise and decide that responses mentioning "theory" or "hypothesis" may demonstrate deeper understanding, but without losing reference to mentions of "experiment", "objectivity", and any other terms the response included.

Process scheme, Durant et al.

The third reference scheme drawn from the literature was that applied by Durant et al. [1989] for analysing the scientific process question; again, the question wording was identical to ours. Their scheme, shown in Table 6, was very similar to Miller's, with each response being sorted into only one concept-based band. It differed primarily in removing Miller's band 3 on unbiased, open investigation, and was included mainly to facilitate comparison of our results with their dataset in the UK adult population. Durant et al. generated an overall process score by combining this ranking with the respondent's results on a set of multiple choice and true-false questions, which we were unable to include due to space and time constraints in the survey.

Band	Sorting criteria [Durant et al. 1989]
4	"Answers referring to science as a process of theory construction and hypothesis testing".
3	"Answers referring to the notion of experimentation, but not mentioning the testing of theories or hypotheses".
2	"Other, often rather vague answers referring to science as a process of fact gathering, or mentioning concrete scientific procedures (looking down a microscope, for example)".
1	All remaining answers.

Table 6: Durant et al.'s analysis scheme for open-ended definitions of studying something scientifically.

Process scheme, Oliver

The final external scheme was that used by Oliver [2008] for the scientific process question, which once again was identically worded to that used in our instrument. She defined a list of key terms such as "experiment" and "hypothesis", and then counted the number of such terms appearing in each response. This method provided a relatively coarse measure of each response's quality, as it took no account of whether the respondent demonstrated the linkages between the concepts (e.g., using an experiment to test a hypothesis), and would privilege answers that were a scattergun list of many terms over those that carefully explained the scientific process with reference to fewer concepts. Oliver's method did have the advantage of being unambiguous to use and simple to calculate, especially in comparison to the SOLO analysis.

Likert scale questions

The Likert scale questions provided quantitative, categorical data. A simple descriptor was provided by allocating 1 to responses marked "Less" on the scale, and 5 to responses marked "More", and then determining the mean and standard deviation of all the responses for each class on each question. This showed quickly whether responses were tending towards rating a particular factor as being of high or low importance.

Comparison and analysis

All of the above schemes were implemented in software, using QSR International's NVivo and Microsoft Excel. This enabled us to investigate the associations between any particular pair of attributes that responses had been allocated: a demographic characteristic of the respondent (e.g. class or level of secondary science); their answers to the open-ended questions, in terms of thematic coding, the SOLO band, or bands under the external comparison schemes; and their answers to the Likert scale questions. To assess whether the differences between responses from different classes were statistically significant, χ^2 tests were used. This was appropriate because the bands in SOLO and the external comparison schemes do not represent an exact continuum and the data do not necessarily fit a normal distribution, so a nonparametric test like χ^2 must be used.

Results

The results of this study are presented as follows. For each of the open-ended questions, the SOLO and thematic coding analyses are explored, followed by the external comparison schemes. Next, further investigations on the influence of secondary science study and gender are provided. Finally, the results of the Likert scale questions are analysed.

Question 1: Content dimension

SOLO bands

Figure 1 illustrates the distribution of SOLO bands, representing the sophistication of responses to the question about radiation, testing the content dimension. The results are grouped by class; a class skewed to the left indicates more incoherent (prestructural) or very simple (unistructural) responses, while a class skewed to the right indicates a higher prevalence of connected and coherent definitions of radiation. Strong differences between the distributions for most classes are apparent. The Fundamentals and Regular students are relatively similar, with no extended responses. Advanced students are more sophisticated, with far less vague answers–a strong baseline of multistructural answers holds most of the cohort, demonstrating that most Advanced students can give a reasonable definition of radiation in terms of more than just one of its features. The prevalence of the more integrated answers, in the relational and extended bands, also increases for the Advanced students. The postgraduate students score better still, with no prestructural answers at all and a high level of relational and extended answers, illustrating that they are able to synthesise their ideas about radiation into a coherent definition.

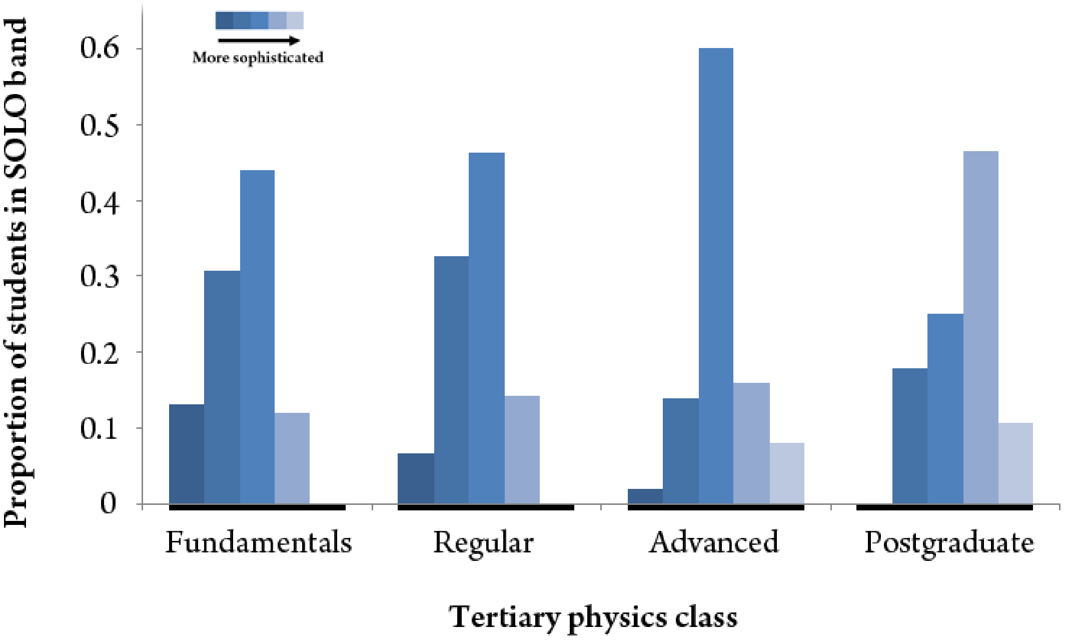


Figure 1: SOLO bands for Question 1, testing the content dimension*. Responses for each class are grouped by a black line. Within each class, the least sophisticated (prestructural) responses are on the left with dark shading, with the most sophisticated (extended) responses on the right with light shading. The Fundamentals and Regular classes exhibit no extended answers at all, while the Postgraduate class performs better, with many extended answers and no prestructural*

ones.

As displayed in Table 7, χ^2 tests show that the changes in SOLO band distribution between the Fundamentals and Advanced and Advanced and Postgraduate classes were statistically significant. These results illustrate the positive contribution of physics education to scientific content literacy.

Tested transition	**χ^2 value**	***d.o.f.***	***N***	***p* value**	**Significant**
Fundamentals to Regular	4.28	4	224	0.45	No
Regular to Advanced	7.53	4	182	0.057	No
Fundamentals to Advanced	10.1	4	141	0.018	Yes
Advanced to Postgraduate	11.5	4	78	0.0091	Yes

Table 7: χ^2 tests for class distributions of SOLO bands in the content dimension.

Thematic coding

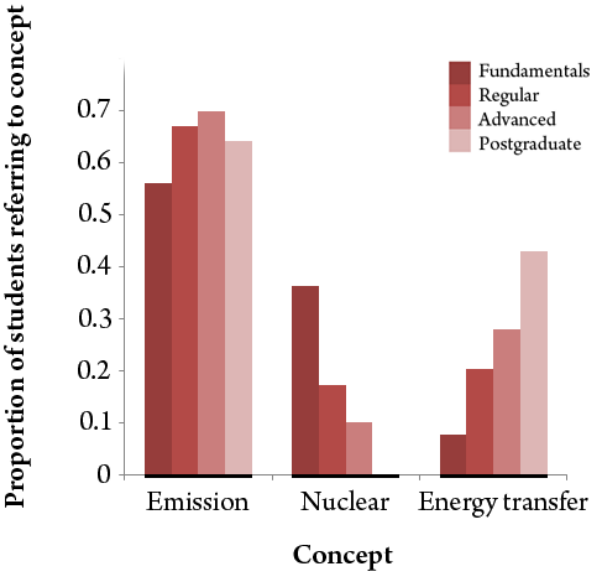


Figure 2: Thematic coding for Question 1, testing the content dimension. *Students with further physics education place less emphasis on nuclear aspects of radiation, instead focusing on energy.*

Three major thematic nodes were identified, and their incidence in the different student cohorts is illustrated in Figure 2. The major conclusion of interest to be drawn is the reduced prevalence of nuclear-focused answers with increasing physics education. While a full third of responses from the Fundamentals students give definitions that were restricted to the ideas involved in ionising radiation from nuclear decay, this proportion decreases in the Regular and Advanced classes, and no such answers were recorded in the Postgraduate class. This indicates that physics education at both secondary and tertiary levels is effective in shifting students' descriptions of a value-laden term like radiation towards more scientific definitions.

Simultaneously with the decrease in nuclear-type answers, an increase in answers referring to the transfer of energy is observed. This reinforces the evidence of a shift towards a more generic, scientific definition and thus increased scientific literacy. Lastly, we note that the concept of emission, focused on by Miller in his scheme, receives a relatively strong response at all study levels, with a slight upward trend with more physics study. However, even the highest proportional response (from the Advanced class, at 70%) still falls well short of universal identification of emission as a key feature of radiation.

External comparison: Miller's scheme

Figure 3 shows responses from the Fundamentals, Advanced and Postgraduate cohorts assessed with Miller's bands, as described in Table 4. The figure indicates a progression to more expert-type answers with increasing physics experience from Fundamentals to Postgraduate. Notably, the Advanced and Postgraduate cohorts have very few of the band 2 and 3 responses, which do not give a general definition, but instead only provide an example of a source or effect of radiation. All classes perform better than the US adult population sample surveyed by Miller, where only 11% of responses were able to give a definition in terms of emission, and thus be placed in the highest band.

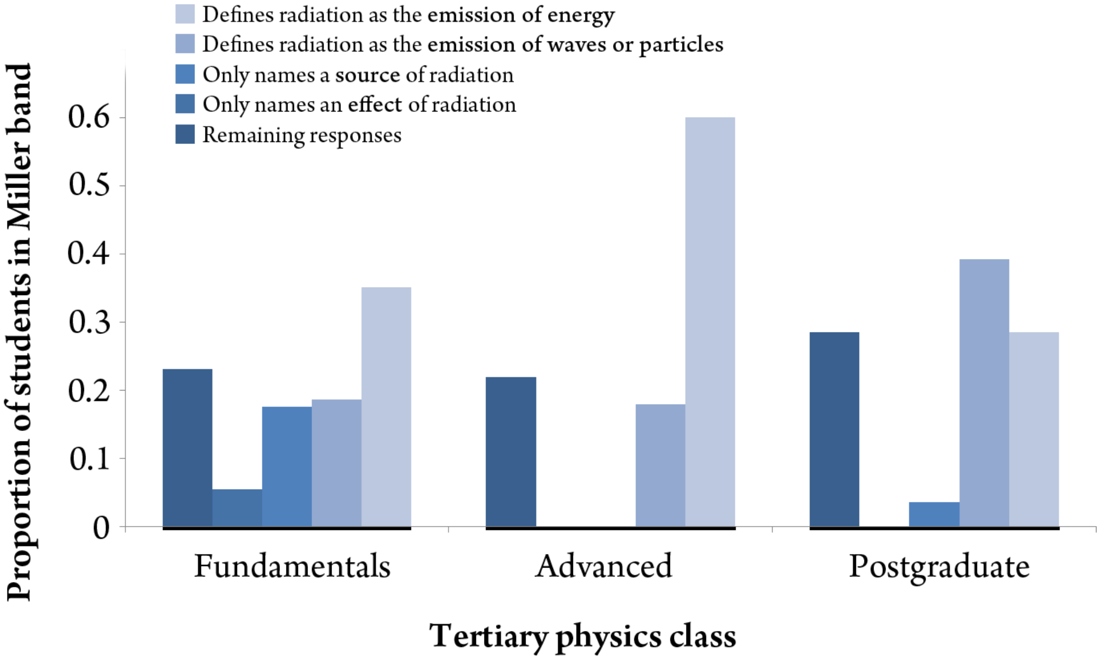


Figure 3: Miller's sorting scheme applied to Question 1, testing the content dimension. *The bands and sorting criteria for this scheme can be found in Table 4. Improvements with increasing physics education are evident, but the bands are somewhat erratic.*

However, it is obvious that the scheme is relatively erratic in comparison to the others previously applied, with a more bipolar rather than normal band distribution. This is largely because Miller's description of his levels is quite vague, making it difficult to sort responses with certainty. Particularly problematic is the fact that the catch-all lowest band might contain answers of very different sophistication. It holds any completely blank or incoherent responses, but is also the only available level for a well-formed response that nonetheless does not mention emission and is therefore ineligible for inclusion in bands 4a or 4b, such as a definition that refers to radiation as the propagation of energy. The criteria are too rigid to be used effectively without further guidance on how to code borderline responses. For this reason, Miller's framework is less useful than our SOLO taxonomy, which admits of the many different ways that answers might be slightly deficient, and hence more accurately captures their level of sophistication.

Question 2: Process dimension

SOLO bands

Figure 4 illustrates the differences in the distribution of SOLO bands by class. It shows a general trend of improvement with increasing physics education. Moving from left to right (from Fundamentals to Postgraduate), at every stage there is a decrease in the number of prestructural and unistructural responses and an increase in the number of responses in the upper bands. It is difficult and somewhat arbitrary to define the exact level at which the respondent can be said to have demonstrated a scientifically literate understanding of the scientific process. Most previous work (such as that of Miller and Durant et al. as discussed above) studying adult populations has been lenient, accepting answers as correct if they contained even a single accurate idea about the scientific process, which corresponds to our unistructural band and above. Miller [2004] judges that approximately 21% of US adults had at least a "minimal level of understanding of the meaning of scientific study," based on analysis of the same question. Taking the unistructural band as this level of understanding, 80% of Fundamentals students achieve scientific literacy in the process dimension, with 84% of Regular students, 94% of Advanced students, and 100% of Postgraduate students reaching the benchmark.

The χ^2 tests in Table 8 show the changes in SOLO band distribution between each student class. The shifts from Fundamentals to Advanced and Advanced to Postgraduate were statistically significant. These transitions essentially correspond to having studied secondary and tertiary physics, respectively. The transition between the Regular and Advanced classes, both containing students who had studied secondary physics with differing levels of ability, was not statistically significant.

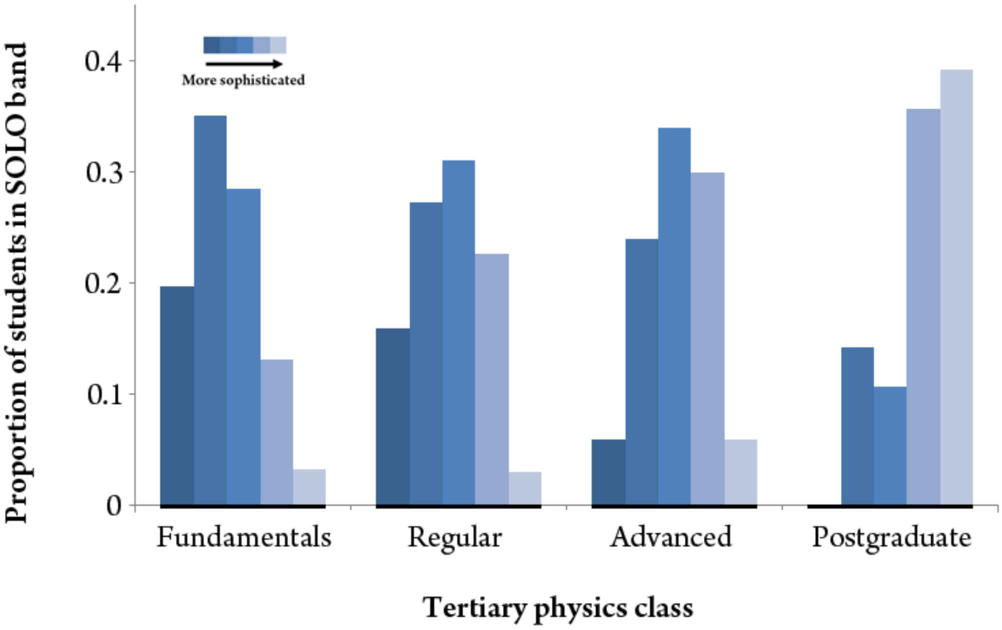


***Figure 4: SOLO bands for Question 2, testing the process dimension**. Responses for each class are grouped by a black line. Within each class, the least sophisticated (prestructural) responses are on the left with dark shading, with the most sophisticated (extended) responses on the right with light shading. The results are similar to those for the content dimension. The Fundamentals and Regular classes exhibit few extended answers, while the Postgraduate class features very large proportions of relational and extended answers, and no prestructural ones.*

These results therefore illustrate that science education, at both secondary and tertiary levels, also makes a positive contribution to scientific process literacy. Determining the exact basis of this improvement is beyond the scope of this study, as multiple effects are likely to be at work. These may include: explicit instruction in the scientific method as part of coursework, particularly in the secondary curriculum; greater exposure to the practising culture of science, particularly in tertiary laboratory and research project work; and a 'correlation rather than causation' effect, where students who are more interested in science and aware of its practices may be more likely to pursue continued study, and thus generate improved scores for the higher classes.

Tested transition	χ^2 value	*d.o.f.*	*N*	*p* value	Significant
Fundamentals to Regular	4.3	4	224	0.37	No
Regular to Advanced	4.6	4	182	0.34	No
Fundamentals to Advanced	11.0	4	141	0.026	Yes
Advanced to Postgraduate	17.6	4	78	0.0015	Yes

Table 8: χ^2 tests for class distributions of SOLO bands in the process dimension.

Thematic coding

Certain key elements of the scientific process are cited with very different frequency in the different classes. Figure 5 shows the seven nodes that were assessed as part of the process dimension. Each cluster shows the differing prevalence of references to the node across the four classes.

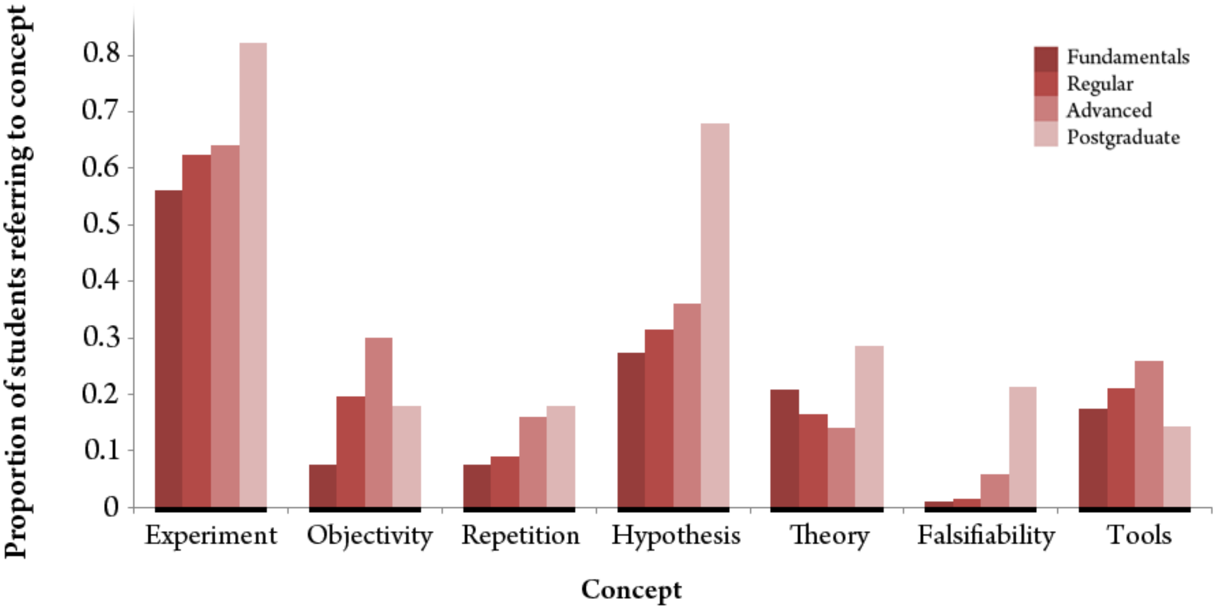


Figure 5: Thematic coding for Question 2, testing the process dimension. *Postgraduates make more frequent references to theory-construction nodes like hypothesis, falsifiability, and theory.*

Several features are apparent. First, references to the empirical or experimental basis of science are strongly prevalent. Close to 60% of all three first-year classes, and over 80% of postgraduate students, identified this as part of the definition of scientific enquiry. This is generally encouraging as the comparable figure for the UK adult population [Durant et al. 1989] is about 10%, and in 1999 was no higher than about 25% for the US adult population [Miller 2004]. Somewhat more disappointing is the low proportion of responses, at all levels, identifying objectivity or impartiality as an important feature of science. No class had more than 30% of responses mentioning this, and the Postgraduate figure was less than 20%. However, it is possible (as with other terms in the list) that respondents might refer to this in a deeper probe such as an interview, which was not feasible to conduct in this study.

Next, note the set of three terms—"hypothesis", "theory" and "falsifiability"—which describe a central feature of science, the construction of theory through the successful testing of falsifiable hypotheses, which are themselves based in prior theory. Postgraduate students demonstrated higher response rates in all of these categories, especially "falsifiability", which is arguably the most nuanced. This is likely to be an outcome of their firsthand exposure to investigative research work, where hypothesis development is central; first-year students have experienced less of the practice of science in this way.

External comparison: Oliver's scheme

Figure 6 illustrates the results of assessing the responses with Oliver's word-counting framework. The number of scientific terms from Oliver's list used in the definition increases from left to right in each cluster. A cluster skewed right indicates that a higher proportion of responses used more scientific terms in their responses, while a cluster skewed left would indicate most responses used few terms.

The figure demonstrates a clear progression from Fundamentals through Advanced to Postgraduate. This is a pleasing result as it helps to independently validate the conclusion reached by the application of our own framework. Also, because Oliver's system is relatively simple to apply, requiring only counting of particular words rather than assessment of a response's sophistication, it could be applied to larger datasets more easily than our taxonomy.

However, this speed does seem to sacrifice a degree of accuracy, for two reasons. First, it can privilege answers that happen to repeat a concept multiple times with slightly different variations of the same word (e.g. "experiment", "observation", "empirical") over those that consistently apply the same word. Second, because it draws from a very specifically defined list of terms, certain synonyms are unfairly excluded—so an answer that used "measurement" instead of "observation", for example, would not be rewarded. In future work, it may be useful to modify the system, including more such synonyms in the list (or allowing the coder flexibility in counting such terms) and restricting the number of times a given concept may be counted if the response refers to it several times. Nevertheless, it provides a useful approximation of the quality of responses.

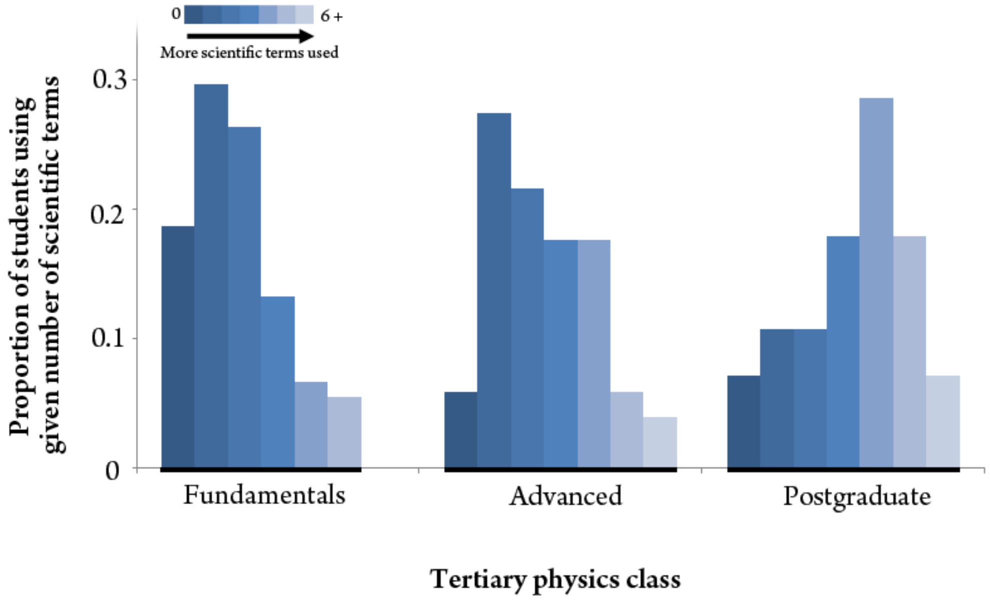


Figure 6: Oliver's word-counting scheme applied to Question 2, testing the process dimension. *Oliver's framework counts the number of relevant scientific terms used in each response. Improvements in the number of terms used with increasing physics education are evident, but the framework cannot supply any more detailed information.*

External comparison: Miller's and Durant et al.'s schemes

Figures 7 and 8 show the band distributions produced by applying the similar frameworks of Miller and Durant et al. respectively. Both schemes show Postgraduates achieving significant numbers of responses in the highest band. This is largely due to these frameworks' emphasis that to be placed in the highest band, a response must mention the ideas of theory and hypothesis construction. Because Postgraduates were much more likely to include these terms in their answers (as seen in Figure 5), they achieve high bands here. However, it is debatable whether such strong emphasis should really be attached to a single concept. While hypothesis forming is undoubtedly important, the question of whether it is *most* central to a scientifically literate definition of the scientific process, and thus deserves to be prioritised over other features that respondents might identify, is a philosophical one. It is not particularly relevant in assessing the level of understanding achieved by the public, and is perhaps better suited to an allied field of study such as the Nature of Science (NOS). This problem is amplified because Miller's and Durant et al.'s coding schemes do not allow responses to be sorted into more than one category. For this reason, allocating response quality via SOLO and separately analysing thematic coding, as in our approach, introduces less artificial distinctions.

In Tables 9 and 10, the results of χ^2 tests to determine whether the differences in distributions between adjacent classes are statistically significant. We find that Miller's scheme is not able to detect a statistically significant difference between any of the adjacent classes. Durant et al.'s scheme finds only a marginal distinction (p = 0.045) between the Advanced and Postgraduate classes.

These both compare relatively poorly with the SOLO taxonomy's clear detection of differences between the Fundamentals and Advanced classes (p = 0.026) and the Advanced and Postgraduate classes (p = 0.0015). Taken together, these results demonstrate that the analysis methods of Miller and Durant et al., based on ranking responses by the concepts that they refer to, are less suitable than the SOLO taxonomy for assessing scientific literacy within a university setting.

Tested transition	**χ^2 value**	***d.o.f.***	***N***	***p* value**	**Significant**
Fundamentals to Regular	4.2	4	224	0.38	No
Regular to Advanced	2.0	4	182	0.72	No
Fundamentals to Advanced	4.8	4	141	0.31	No
Advanced to Postgraduate	8.3	4	78	0.083	No

Table 9: χ^2 tests for class distributions under Miller's scheme (Figure 7) for the process dimension.

Tested transition	**χ^2 value**	***d.o.f.***	***N***	***p* value**	**Significant**
Fundamentals to Regular	1.2	3	224	0.76	No
Regular to Advanced	2.1	3	182	0.56	No
Fundamentals to Advanced	3.9	3	141	0.28	No
Advanced to Postgraduate	8.0	3	78	0.045	Yes

Table 10: χ^2 tests for class distributions under Durant et al.'s scheme (Figure 8) for the process dimension.

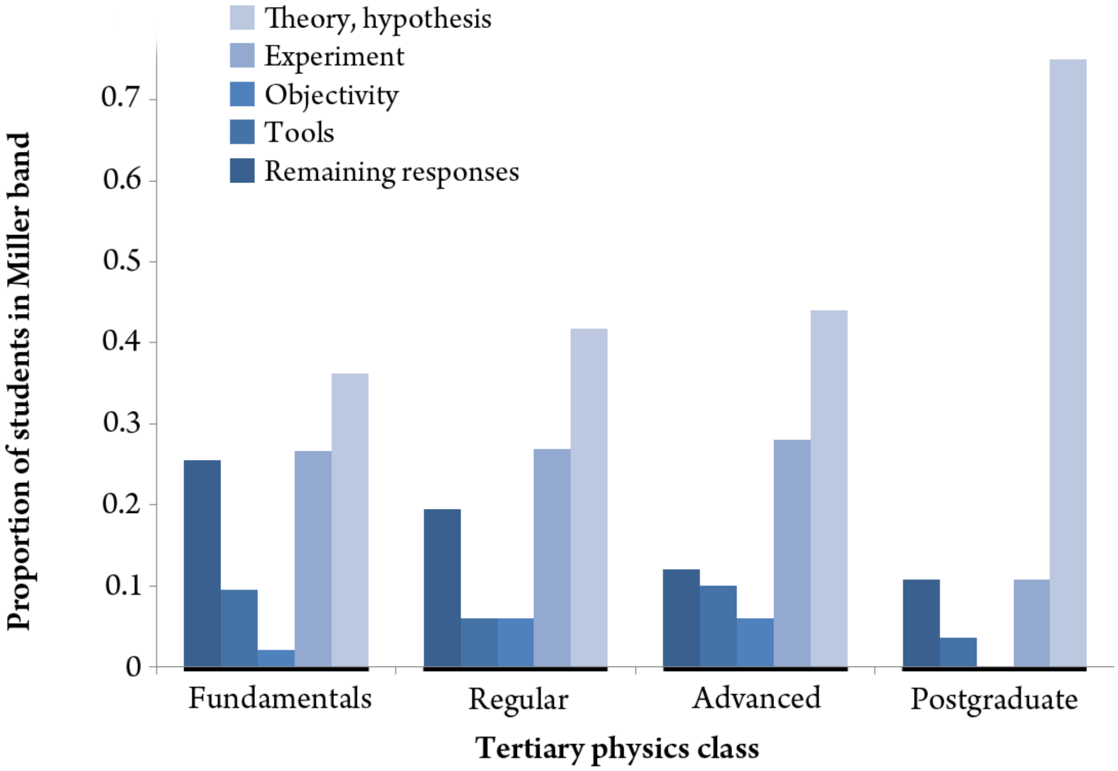


Figure 7: Miller's sorting scheme applied to Question 2, testing the process dimension. *Postgraduates perform much more strongly than first-year students, because Miller's sorting criteria prioritise answers that refer to "theory" or "hypothesis".*

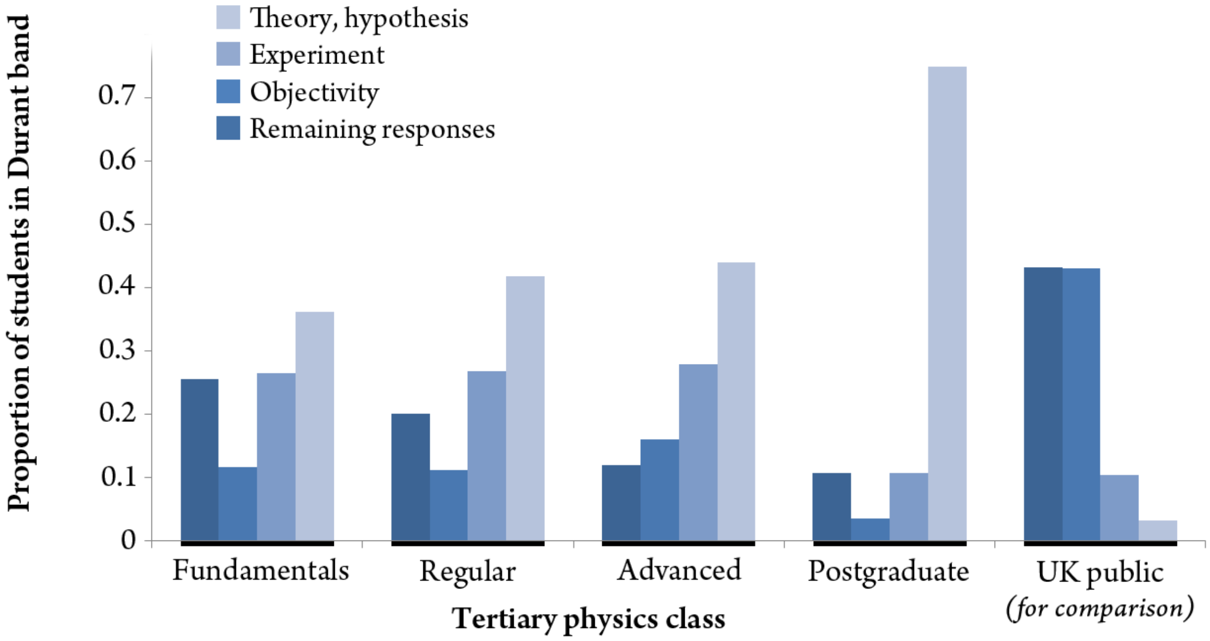


Figure 8: Durant et al.'s sorting scheme applied to Question 2, testing the process dimension. *Postgraduates perform much more strongly than first-year students, for the same reason as in Miller's scheme Figure 7. Here, we also note that all four student classes perform markedly better than Durant et al.'s dataset from the UK adult public.*

Further analysis

Influence of secondary science on SOLO band distributions

Another interesting result is shown in Figure 9, which compares the SOLO band distributions for first-year students who did and did not study science in their final year of secondary school. Postgraduates are excluded from this data-set, as their more recent undergraduate education would introduce a bias. Figure 9(a) applies to Question 1, testing the content dimension, and Figure 9(b) applies to Question 2, testing the process dimension.

The graphs demonstrate a strong improvement in the content dimension of scientific literacy for students who studied secondary science ($\chi^2(4, N = 274) = 19.3, p < 0.001$). Although there appears to be a shift towards more sophisticated responses in the band distribution for the process dimension, it is not quite statistically significant ($\chi^2(4, N = 274) = 8.8, p = 0.06$). These results indicate that secondary science education is most effective at improving scientific content literacy, where tertiary education was more effective at improving process literacy.

This analysis also provides an interesting positive assessment of Australian secondary science education. Miller [2004] indicates that the major determining factor in US adult scientific literacy is the number of tertiary science courses taken, which "compensate in part for inadequate middle school and high school science". Our results match the comparatively strong performance of Australian students in TIMSS, an international survey of scientific literacy at the 'middle school' (Year 9) level, where the US typically performs relatively poorly [Forgione 1998]. The fact that our instrument reveals this difference in secondary-based achievement, which is borne out in another major instrument, helps validate our survey and assessment model in answering the first research question.

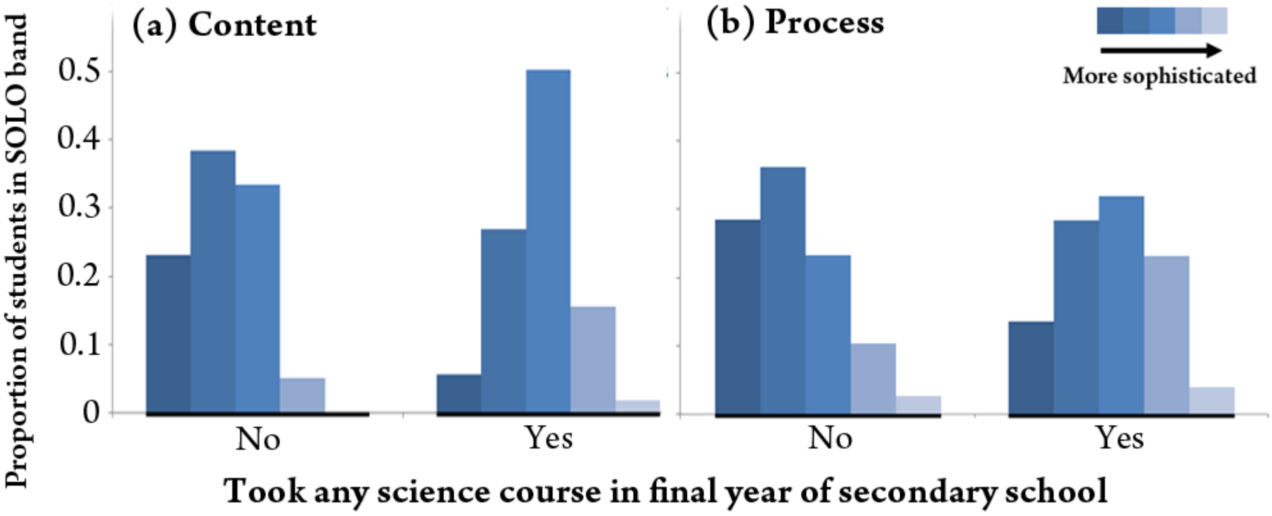


Figure 9: SOLO band distributions, vs. secondary science*. The graph shows the three first-year classes only, with Postgraduates excluded to avoid biasing the sample. Students who took any science course in their final year of secondary school tend to achieve higher SOLO bands than students who did not. The difference is stronger for process than content, because process is accessible through any science course, while radiation is more physics-specific, so other students (taking biology, for example) may not improve.*

Influence of gender on SOLO band distributions and type of science studied

Figure 10 compares the SOLO band distributions on both open-ended questions with the gender of respondents for the first-year classes. Figure 10(a), for the question on radiation, demonstrates that males achieve statistically significantly higher bands than females ($\chi^2(4, N = 274) = 20.6, p < 0.001$). In Figure 10(b), which applies to the process question, there is no statistical difference ($\chi^2(4, N = 274) = 1.3, p = 0.85$). To explain this, we note the different distribution of secondary science subjects by gender, with males more likely to have studied physics and females more likely to have studied biology, as shown in Figure 11. The overall proportions of males and females having studied science at secondary level are very similar. This may explain the observed results because the content question, being about radiation, is more physics-focused, and may therefore slightly advantage the males who are more likely to have studied physics. The process question, being related to generic features of science, is essentially equally accessible to anyone who has studied any course in secondary science, leading to the very similar distribution between genders.

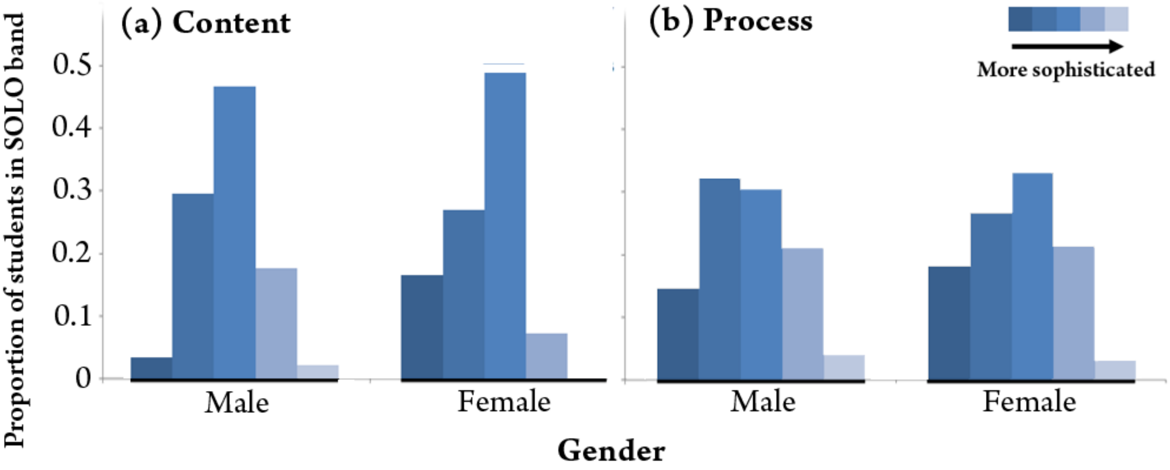


Figure 10: SOLO band distributions, vs. gender*. The graph shows the three first-year classes only, with the male-dominated Postgraduates excluded to avoid biasing the sample. Males tend to achieve higher SOLO bands than females for content (10a), but not for process (10b).*

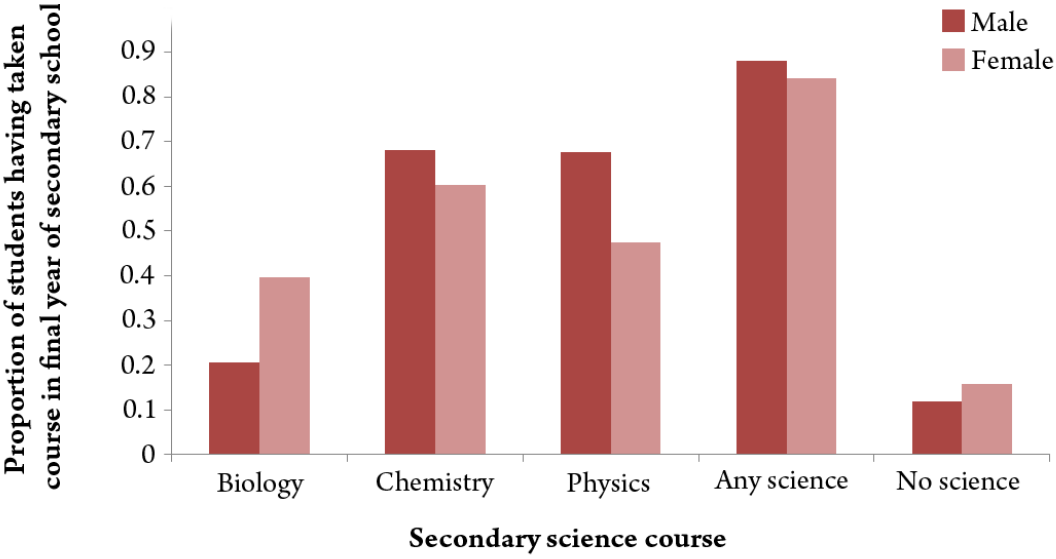


Figure 11: Secondary science, vs. gender, *illustrating the proportion of students who took each course. The bars are not mutually exclusive; students may have studied multiple courses.*

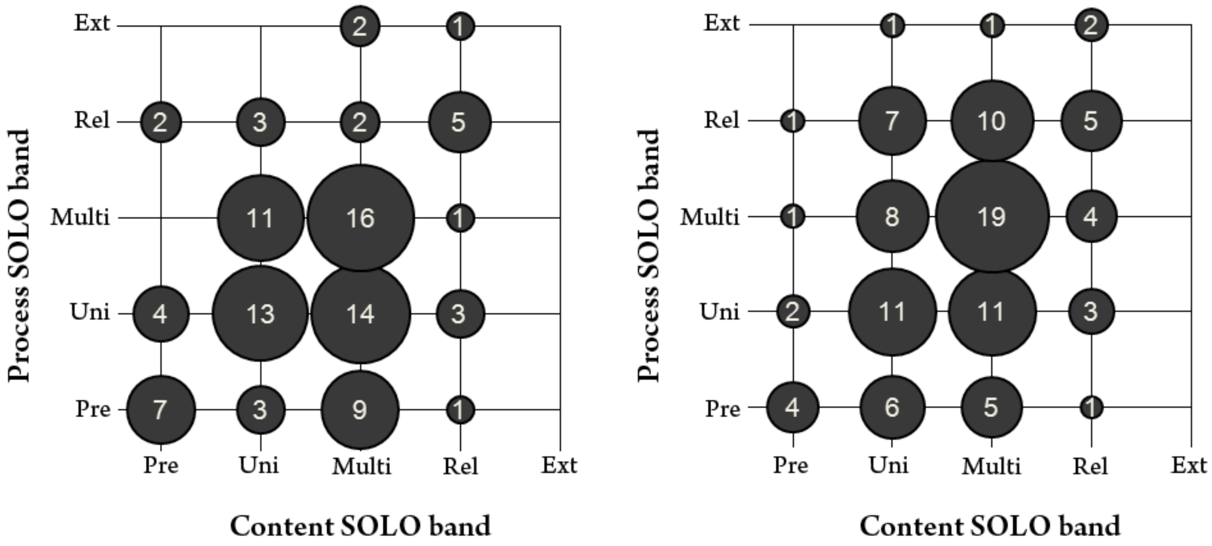
Comparing the content and process dimensions

We conclude the analysis by simultaneously examining the SOLO band distributions achieved by each class on both open-ended questions. This comparison between the content and process dimensions is useful in determining where the physics education process is having its greatest effect. The four graphs in Figure 12 illustrate this point. For each graph, the axes represent the SOLO bands for the two questions. The area of the circle at each intersection of two bands represents the proportion of students in the class who fell into both of those corresponding bands, expressed as a percentage.

Differences are easily apparent between the various classes. The Fundamentals students in Figure 12(a) primarily cluster in the low-middle ratings on both questions. On average lying closer to the bottom than the left, their scientific literacy is more deficient in the process dimension. This might be expected from their lower level of previous science study, meaning they have had less firsthand exposure to the practice of the scientific method. The Regular class in Figure 12(b) begins to expand towards the upper right quadrant of the graph, where students are scoring well in both dimensions, but retains a strong component of responses in the lower ratings. This lagging set is diminished in the Advanced class in Figure 12(c), with very few prestructural responses for either dimension, and a reduced number in the unistructural band. Also, there is now a relatively even distribution of students across the vertical process dimension, indicating the content-biased literacy of Fundamentals students has been replaced with a more unified understanding at the Advanced level. This trend continues into the Postgraduate students in Figure 12(d), who are heavily biased towards the top of the graph; it seems likely that their exposure to scientific research work provides strong literacy in the process dimension. Their responses are also generally clustered in the upper right, indicating high scientific literacy overall.

(a) Fundamentals

(b) Regular



(c) Advanced

(d) Postgraduate

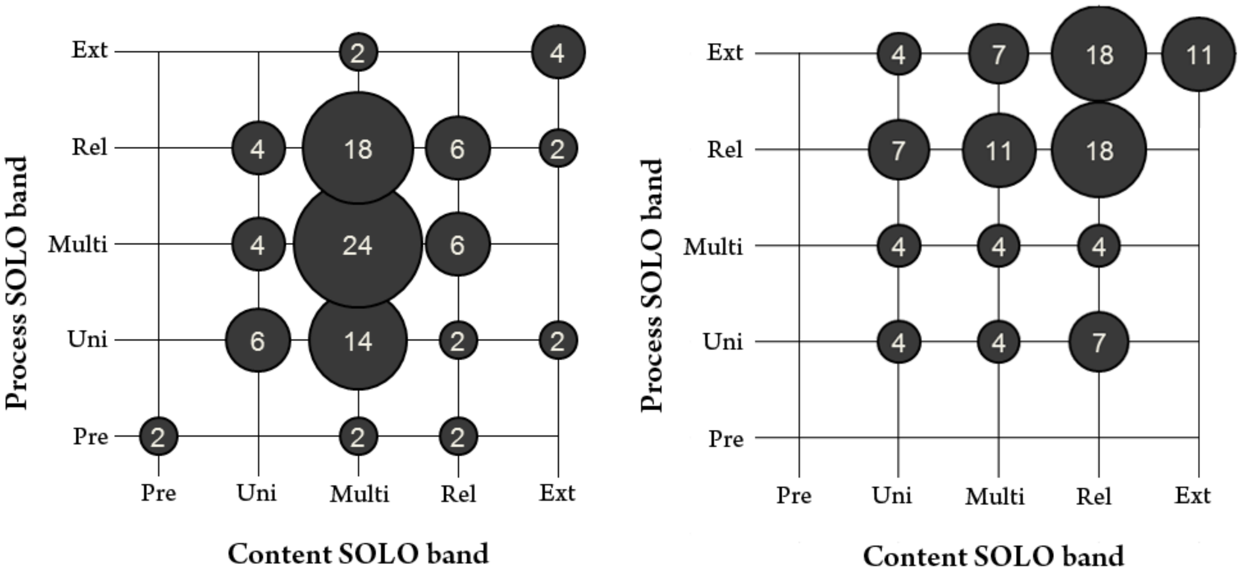


Figure 12: SOLO band distributions for content and process, by class. *Each matrix illustrates one class's performance on the content and process dimensions of scientific literacy. The horizontal axis represents SOLO bands along the content dimension, and the vertical axis, along the process dimension. The axes each increase from the prestructural band (Pre) through unistructural (Uni), multistructural (Multi) and relational (Rel) to the extended band (Ext). The area of the circle at each intersection of two bands represents the percentage of students in the class receiving those two bands. Thus, large circles in the top right of a matrix would indicate strong scientific literacy in both dimensions, while smaller circles across the matrix would indicate a spread of literacy levels.*

Society dimension: Likert questions

The four Likert questions largely tested the society dimension, with some overlap into process. The students' responses were analysed to search for differences based on their level of secondary science and tertiary physics education. Recall that in all the following tables, a higher score indicates that students have rated a particular factor as being more important. The scale ranges from 1 to 5.

Question 3: Scientific objectivity

This question asked: *Which factors do you think are important in determining how objectively scientists report on their results?* The factors were: *Scientific ideals, like reporting results accurately and without bias; The knowledge that their results will be checked by other scientists; The knowledge that their work affects the wellbeing of society;* and *A desire to please governments and corporations that fund them.*

Factor:	*Scientific ideals*		*Peer review*		*Social wellbeing*		*Funding sources*	
Class:	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Fundamentals (n=91)	4.5	0.8	3.3	1.2	3.8	1.0	2.3	1.2
Regular (n=133)	4.5	0.7	3.9	1.1	3.6	1.2	2.4	1.2
Advanced (n=50)	4.6	0.6	3.2	1.2	3.4	1.2	2.8	1.2
Postgraduate (n=28)	4.5	0.9	4.2	0.9	2.6	1.0	2.5	1.1

Three major trends are evident in the table. First, all classes allocate high importance to the upholding of scientific ideals, indicating a strong belief that these are not only abstractions but are also functional and relevant guides of behaviour for practising scientists (all have $\bar{x} \geq 4.5$, with $0.6 \leq \sigma \leq 0.9$). Second, all classes allocate comparatively low importance to the influence of funding sources such as governments and corporations (all have $\bar{x} \leq 2.8$, with $1.1 \leq \sigma \leq 1.2$). This insistence that scientists do remain impartial despite funding pressures is an interesting finding. It demonstrates faith in the integrity of the scientific enterprise despite media attention to the objectivity of scientists in controversial areas such as climate change. Third, Postgraduate students attach relatively more importance to peer review, and less to the social impacts of science, as motivating factors. This weighting may indicate that closer experience with the human factors of scientific research gives them a deeper understanding of how it operates, replacing the more naïve view that motivates students of science who have not yet practised research. This outlook arguably demonstrates a higher level of realism in Postgraduates' scientific literacy, matching the results from the open-ended questions.

Question 4: Theory construction and acceptance

This question asked: *Which factors do you think are important in determining how scientists decide whether to accept a proposed new theory?* The available factors were: *The body of evidence and testing that supports the theory*; *Whether the theory is elegant, explaining the evidence simply*; *Personal views or attachments to the theory or opposing theories*; and *Factors like job security, fame and success.* These factors were designed to approximately align with those from Question 3: the first one in each question relating to a pure scientific ideal; the second relating to an accepted part of the practise of science; the third relating to a less scientific but still legitimate influence; and the fourth relating to a potential external disruptor of the scientific process.

Factor:	*Body of evidence*		*Elegance*		*Personal views*		*Fame, success*	
Class:	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Fundamentals (n=91)	4.6	0.7	3.0	1.2	2.4	1.2	2.2	1.2
Regular (n=133)	4.5	0.7	2.9	1.1	2.2	1.2	1.9	1.2
Advanced (n=50)	4.7	0.5	3.1	1.1	2.2	1.2	2.1	1.1
Postgraduate (n=28)	4.7	0.8	3.8	1.0	2.7	1.1	2.1	1.1

Again, three major trends are evident, and they correspond well to the results from the previous question. First, all classes attach a high rating to the need for testing and evidence to be the primary arbiter of a new theory's acceptability (all have $\bar{x} \geq 4.5$, with $0.5 \leq \sigma \leq 0.8$). Second, all classes give a low average rating to the role of fame, success and job security, rejecting the suggestion that scientists may support a less sound theory if it would lead to personal gain (all have $\bar{x} \leq 2.2$, with $1.1 \leq \sigma \leq 1.2$). The combined result of these trends is to reinforce the students' previous strong focus on scientific ideals above economic and other influences. Finally, Postgraduate students attach more importance to the two other factors: the theory's elegance and simplicity, and the personal view of the scientist about the theory's value. These responses demonstrate that Postgraduate students recognise that the evolving nature of science necessitates other heuristics beyond factual analysis of the body of current evidence, when seeking to make judgements about theory. This recognition reveals a more nuanced understanding than the undergraduates' dominant focus on evidence alone. The responses to this question therefore once again illustrate stronger scientific literacy in the Postgraduate class.

Question 5: Scientific disagreement

This question asked: *Which factors do you think are important in determining why scientists disagree on some scientific theories?* The available factors were: *Not all of the relevant facts have been discovered; Different scientists interpret the available facts differently; Different scientists have different personal opinions or moral values; Different scientists have different ties to companies and governments.* These factors are again intentionally aligned in style with those from the previous questions, as described above.

Factor:	*Need more facts*		*Interpretations*		*Personal values*		*External ties*	
Class:	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Fundamentals (n=91)	4.0	0.9	3.8	1.0	3.1	1.3	2.4	1.2
Regular (n=133)	4.0	0.9	3.8	0.9	3.1	1.2	2.6	1.3
Advanced (n=50)	4.3	0.8	3.9	0.9	3.1	1.2	2.8	1.3
Postgraduate (n=28)	4.1	0.9	4.0	1.0	2.7	1.0	2.3	1.2

Two major features of the table are worth noting. First, all classes give relatively high scores to the role of factual evidence (and differing interpretation of available evidence (all have $\bar{x} \geq 3.8$, with $0.9 \leq \sigma \leq 1.0$). These scores are not as high as those given to comparable factors in previous questions, and have higher standard deviations. This shift indicates that students are less sure of how the scientific method applies when scientists are disagreeing, rather than moving in a straightforward way from finding evidence to creating a theory. However, the correct emphasis on the possibility of legitimately divergent interpretations of the same evidence, from both undergraduates and Postgraduates, indicates a good understanding. Second, all classes attach comparatively little importance to the role of personal values (the implication being that these are not scientifically-based views, which are described by the 'interpretations' factor), and to the influence of external bodies. This trend is especially marked among Postgraduates, with low mean scores of 2.7 ($\sigma = 1.0$) for personal values and 2.3 ($\sigma = 1.2$) for external ties. It is debatable whether or not this suggestion that external social influences are unimportant demonstrates higher scientific literacy. While it may be seen to show knowledge of the scientific process in an ideal setting, it does ignore the many cases of scientists being heavily influenced by their funding sources, such as in tobacco and pharmaceuticals [Yach 2001]. It might therefore be considered to demonstrate slightly better scientific literacy in the process dimension, but more limited understanding in the society dimension.

Question 6: Political influence on science

This question asked: *Which factors do you think are important in determining how scientific studies are affected by politics?* The available factors were: *The scientific method's requirements of objectivity and independence; The significant media attention that some scientific studies receive; The way that governments often use scientific results in policy-making; The way that governments often fund and regulate scientific studies.* Because this question focused more explicitly on the society dimension, the alignment of factors, as described above, is less exact than for questions 3-5, but is still apparent.

Factor:	*Scientific method*		*Media attention*		*Policy-making*		*Funding*	
Class:	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ	$\bar{x}$	σ
Fundamentals (n=91)	3.8	1.0	3.1	1.2	3.4	1.0	3.7	1.1
Regular (n=133)	3.6	1.2	3.5	1.2	3.4	1.1	3.7	1.2
Advanced (n=50)	3.5	1.3	3.8	1.0	3.7	1.1	4.1	0.9
Postgraduate (n=28)	3.3	1.4	3.7	1.1	3.3	1.0	4.0	1.2

The responses to this question do not present the same clear trends as above. There are no questions where there is a strong deviation of any one class from the others. Similarly, the responses for all four factors are relatively similar, clustering between mean scores of 3.0 and 4.0 (with $0.9 \leq \sigma \leq 1.4$).

From this data, two conclusions may be drawn. First, the similarities between the responses to each factor indicate that students have less well-formed opinions when asked about explicitly society-focused questions with less of a process component. Second, the similarity between class scores indicates that these opinions also change relatively little over time. The sort of supervised research work that the Postgraduates had previously conducted during their undergraduate degrees evidently contributes more to an understanding of the idealised scientific method (process dimension) than to literacy of the complex interaction between science and political influences (society dimension). There is also little direct consideration of these interactions in the undergraduate physics curriculum; although some students may take related courses in areas such as the History and Philosophy of Science, it is evident that many Physics graduates do not have a greatly improved or different understanding of these issues than they had when they began first year study.

Discussion

Based on the results and analysis presented thus far, we can suggest answers to the research questions. First, we aim to determine what features are required for an instrument that can effectively assess scientific literacy in a university setting. We find that it is important to devote attention both to the survey document itself and to the analysis techniques used to extract meaning from the information gathered. The survey must be sufficiently brief to be practically administered in time-constrained laboratory or lecture settings, making our two-page instrument more appropriate than previous surveys. Based on the dominant three-dimensional framework of Miller, the instrument should assess knowledge of scientific content, understanding of the scientific process, and to a lesser degree attitudes toward the relationship between science and society. This last dimension, society, is the hardest to assess and was treated in the least depth in this investigation, but is potentially worthwhile because it has the least overlap with traditional methods of assessment and receives the least attention in most subject curricula. Finally, the survey should undergo an accepted process of expert validation before being deployed.

Suitable analysis techniques are equally, if not more, important to deriving useful conclusions from the raw data. Previous analyses have typically suffered from being too simplistic, perhaps as a consequence of being designed for processing very large data sets. We have derived significant value from adapting techniques more appropriate for the university setting, where detailed analysis is both possible and necessary, because of the relatively small differences between student classes as compared to the broad variations in the general population. The modified SOLO taxonomy is sufficiently fine-grained to detect these distinctions—for example, the statistically significant difference in process dimension scores between the various classes, none of which were revealed by Miller's scheme. Thematic coding allows us to trace the evolution of the various pieces of conceptual understanding that underpin students' increasing scientific literacy throughout their education, giving insight into the basis of their higher SOLO scores. Additionally, previous studies have often used only one method of analysis; we apply multiple techniques to the same set of data to provide an increased level of confidence in the observed trends.

Next, having confidence in the applicability of the instrument developed, we apply it to answering the second research question. We aim to characterise how students with differing levels of secondary and tertiary physics perform across the three main dimensions of Miller-type scientific literacy: content, process and society. It is important to note that although many studies have examined scientific literacy in adults and school-age populations, ours is the first to directly measure university students at different levels of science education. This higher resolution allows us to more accurately trace the development of scientific literacy, across these three dimensions, as students progress.

Analysis of the open-ended question on radiation demonstrates that a higher level of physics education is associated with improved scientific content literacy, as measured with the SOLO taxonomy. Postgraduate students are more likely to be able to provide an integrated, coherent definition that is placed in the higher SOLO bands. As seen in Figure 1, nearly 60% of Postgraduates are categorised into the relational or extended categories of the SOLO taxonomy, indicating they are able to incorporate multiple ideas about radiation into a structured response. In contrast, less than 15% of Fundamentals students reach those levels, with almost 40% only managing either incoherent (prestructural) or very simplistic (unistructural) definitions. Regular students perform similarly to Fundamentals students. Advanced students show intermediate performance: while most of them can identify a few aspects of radiation (the multistructural band), only 25% can give a reasonably detailed definition. Postgraduates are also less likely to fixate on ionising radiation from nuclear activity, instead identifying radiation as a broader physical concept based on the emission of energy. The usage of nuclear-focused definitions generally decreases as students receive more physics education, as seen in Figure 2.

These differences between students in the various first-year cohorts, none of whom had received a significant amount of university physics tuition at the time of the study, demonstrate the impact of senior secondary school science. This relationship is also illustrated directly in Figure 9(a), which shows that students who studied science in their final year of secondary school have a reduced likelihood of giving an incoherent definition, and an increased chance of being able to identify more than one feature of radiation. It is worth noting in this context that it is not only through studying secondary physics that students achieve better definitions of radiation. The performance of the Fundamentals and Regular classes was not statistically different, despite the Regular students all having previously studied physics and the Fundamentals not having done so. A probable explanation is that a large proportion of Fundamentals students had studied some other science (such as chemistry or biology), where radiation may also be involved in the curriculum.

There are thus two general conclusions that we make about the content dimension. First, university physics education does contribute to scientific literacy in the context of content knowledge, since Postgraduates show a statistically significant increase in the sophistication of their responses, and use more physically accurate themes in their responses. Second, although secondary science education can have a significant positive impact in comparison to studying no science at all, it is necessary for students to perform well at physics (sufficiently well to join the Advanced cohort) before they are likely to be scientifically literate enough to give an adequate definition of radiation. This point has important implications for judging the usefulness of scientific literacy as a concept in the public understanding of science. If a large fraction of students who have chosen to study science at university level are unable to provide a detailed definition of radiation, is it genuinely reasonable to expect members of the wider population to be able to do so? The concept of relative scientific literacy has merit here. It may be sufficient for most sectors of society to meet a lower threshold (perhaps recognising that not all radiation is ionising radiation), with more stringent requirements placed on those in science and related fields, and on relevant decision makers.

Next, we consider the process dimension, through the open-ended question on what it means to study something scientifically. Again, we find statistically significant improvements in the SOLO band distribution for the Advanced and Postgraduate classes, as seen in Figure 4. It is also apparent that few first-year students are placed in the highest SOLO band for their definition. This suggests that the direct research experience afforded to Postgraduates, including that carried out during their undergraduate degree, is helpful in building a coherent and detailed understanding of the scientific process. Such a conclusion is reinforced by examination of Figure 5, indicating which key distinguishing features of science were identified by each class. Several terms related to the investigative process, such as “theory”, “hypothesis” and “falsifiability”, were used more frequently by Postgraduates. All student classes performed relatively well at identifying at least one feature, of which “experiment” was the most commonly given, occurring in roughly 60% of undergraduate responses and 80% of postgraduate ones. Again, these results have implications for how we think about public scientific literacy. Figure 8 shows that only 10% of UK adults were similarly able to cite “experiment” as a key part in the scientific process. It would seem unreasonable to expect the whole population to be able to do this when significant minorities of tertiary physics students cannot do so. However, there is still a large gap between the public’s 10% and the undergraduates’ 60%, which demands to be closed rather than defined away if the public is to effectively engage with the process of science and thus be able to trust its outcomes and prescriptions.

One venue for such improvement may be secondary science. Unlike the content dimension, there is not yet a statistically significant association between having studied science in the final year of secondary school and performing better on the process dimension, as seen in Figure 9(b). Senior science curricula do contain experimental and investigative tasks, which provide implicit exposure to scientific processes for students, even if they are not actively considering what makes these processes important. However, because they are often relatively simple and the experimental or research component is frequently not devised by students, there is limited scope for development of the higher aspects of the scientific process, such as the evolving relationship between theory and hypothesis. This may partially explain the lower proportion of first-year students using these terms in their responses, and the correspondingly low fraction of responses placed in the upper SOLO bands. A stronger process focus may enable secondary science to contribute to this dimension of scientific literacy.

Finally, the society dimension was also examined. The responses to the Likert questions showed that there is little change over time in students' views about the way external influences have an impact on the scientific process. Students in all four classes repeatedly suggested that scientific ideals, such as objectivity and reliance on evidence, were the key priorities for practising scientists, while influences such as external funding, personal opinions, and motivations like fame or job security were comparatively unimportant. Postgraduates also prioritised some more advanced aspects of the scientific process, peer review and theoretical elegance, which undergraduates did not rate highly. On one level, these results may be viewed as a demonstration that science students have been correctly taught to prioritise scientific impartiality. At the same time, influences such as funding frequently are important in the actual conduct of science, and are especially prominent in the way the wider public perceives the work of scientists. Students graduating into practising scientists must be able to recognise and engage with these issues rather than suggesting they are not important. Unfortunately, these results indicate that students do not gain as much from physics education—at either the secondary or tertiary levels—in the society dimension as they do in the content and process dimensions. If promoting all three dimensions of scientific literacy is to be a goal of these educational settings, there is scope for more direct consideration of these society-related issues within curricula for physics and other science subjects.

Future Work

A longer study would be able to achieve two key goals. First, it would be able to add methodological improvements that were not possible under the time constraints of an Honours project. These include a more formalised expert validation process for the instrument, and a full inter-coder review of the data to ensure that multiple independent assessors categorise responses in the same way, improving reliability. If possible, the application of more time-consuming data collection methods such as a limited number of interviews would also help clarify the extent to which students' short responses reflect the full depth of their understanding, which is a perennial challenge in qualitative research.

The second main direction of further investigations would be to expand the study beyond physics students. While a pilot study of pre-service science teachers was conducted as part of this work, not enough responses were received to draw statistically valid conclusions. Testing students outside physics is necessary to gain a more accurate sense of the ability of broad tertiary science instruction to improve scientific literacy. In order for the instrument to be used effectively with non-physics students, it is likely that the content dimension section would need to be revised or extended, as radiation is a physics-centric topic that would unfairly disadvantage students in other disciplines. Because previous work [Rennie and Stocklmayer 2003] has shown that even practising scientists perform little better than the lay public on questions about scientific content outside their discipline, a range of content should be tested. For example, questions for "molecule" and "DNA" might be added to test knowledge of core terms in chemistry and biology. Also, the addition of some multiple-choice or true-false questions might enable direct comparison between the disciplines, adding another layer of assessment to supplement the richer but more ambiguous open-ended questions.

In designing any revisions, the instrument must remain brief enough to be practically deployed. This could mean reducing the number of Likert scale questions to focus the instrument more explicitly on the content and process dimensions, at the expense of the society dimension. Another option is to split the instrument into multiple parts that can be used in different circumstances, depending on which dimensions the researcher is most interested in testing, without losing sight of scientific literacy as an integrated assessment of all three. Finally, implementing the instrument as a pre- and post-test for units of study, especially for those that are specifically designed to deal with scientific literacy or the role of science in society, would enable evaluation of those units' effectiveness.

Conclusions

As a widely accepted goal of science education (Kanasa 2008), it is surprising that scientific literacy in tertiary students has not been evaluated more closely before this study. There is scope for significant further work, as outlined above, but three important conclusions can be drawn from this investigation.

First, previously developed instruments and techniques for assessing scientific literacy are frequently lacking in one or more important ways that limit their ability to be applied in a university setting. Some survey documents are too lengthy to be practically deployed on sufficiently large scales in time-poor classes, such as the Views of the Nature of Science (VNOS-C) or Views on Science, Technology and Society (VOSTS) questionnaires. Some analysis methods are too simplistic or limited to detect the fine distinctions between various classes of science students—distinctions which are instructive in determining how education is improving scientific literacy across each of the three dimensions. By creating a relatively brief instrument that still retains open-ended questions supplying rich data, and then subjecting those responses to more detailed analysis methods, valuable information can be gained. The utility of adapting the SOLO and thematic coding frameworks to studies of scientific literacy has been demonstrated, allowing an assessment of students' overall sophistication and coherence in responding, as well as the tracking of different concepts that form part of their understanding.

Second, the results verify that science education is indeed generally effective at raising the content and process dimensions of scientific literacy. This is reassuring; concerns about the content-heavy transmissive style of science teaching have sometimes called into question whether science education does genuinely improve student understanding. The four classes surveyed represented differing levels of physics educational background. Statistically significant increases in the SOLO sophistication ranking, for both the content and process dimensions, were observed in the Advanced and Postgraduate classes, who have received greater prior exposure to physics education. When this was broadened to consider whether first-year students had studied any type of science in their final year of secondary school, a statistically significant improvement was found in content scores for those who had studied science over those who had not, but the improvement was not significant for process scores. Given this low baseline, tertiary physics instruction appears to yield an especially strong improvement in understanding of the scientific process. Because Postgraduate responses referred more frequently to concepts surrounding hypothesis construction and testing, this improved understanding may be a result of the more common firsthand exposure to research practices at university level.

Third, little difference is found between students' conceptions of the relationship between science and society, whether they are a first-year student with no physics background or a postgraduate student who has completed a degree in physics. This result is surprising. Across all classes, students tend to suggest that behavioural norms among practising scientists are those that match with scientific ideals, such as objectivity and the centrality of evidence. They rate other factors, such as external funding, government regulation, and personal ambitions, as being of low influence. Where Postgraduate students do differ—they rate elegance and peer review as more important influences than undergraduate students do—they still prioritise factors internal to science rather than any that reflect the two-way relationship between science and society. This minimal change over time in the society dimension of scientific literacy, and the fact that student opinions generally reflect scientific ideals, may indicate that few core science curricula are devoting attention to these issues surrounding the practice of science as a public enterprise. Understanding these is a central part of scientific literacy.

In a time when science touches almost all public issues, often requiring decision makers and the public to evaluate complex competing arguments in an environment where not all participants are acting in good faith on the best available evidence, scientific literacy remains a vital skill. Consequently, it is important that science education providers remain aware of its principles, and determine ways to build students' abilities across each of the three dimensions of core scientific content, understanding of the scientific process, and knowledge of the relationship between science and society. It is also helpful to test student progress in scientific literacy to ensure that teaching is producing the desired improvements. Since no nation is yet close to having a population that is predominantly scientifically literate, the endeavour of improving scientific literacy is a crucial one and deserves ongoing attention.

References

Aikenhead, G., and A. Ryan (1992), Science Education, 76(5), 477-491.

Bell, R., and N. Lederman (2003), Science Education, 87(3), 352-377.

Biggs, J., and K. Collis (1982), Evaluating the quality of learning: the SOLO taxonomy. New York: Academic Press.

Boulton-Lewis, G., (1994), Higher Education, 28(3), 387-402.

Brossard, D., and J. Shanahan (2006), Science Communication, 28(1), 47-63.

Burns, T., D. O'Connor and S. Stocklmayer (2003), Public Understanding of Science 12(2), 183-202.

Davis, R., (1958), The Public Impact of Science in the Mass Media. Ann Arbor: University of Michigan.

Durant, J., G. Evans and G. Thomas (1989), Nature 340, 11-14.

Forgione, P., (1998, April). What We've Learned From TIMSS About Science Education in the United States. Speech presented to the Conference of the National Science Teachers' Association.

Hobson, A., (2008), The Physics Teacher, 46 (October 2006), 404-406.

Hodges, L., and L. Harvey (2003), Journal of Chemical Education, 80(7), 785-787.

Hurd, P., Educational Leadership 16 (1958): 13-16.

Kanasa, H., and K. Nichols (2008), In P. Jeffery (Ed.), AARE 2008 Conference Papers Collection, Brisbane: Australian Association for Research in Education.

Koballa, T., A. Kemp and R. Evans (1997), The Science Teacher 64(7), 27-31.

Laugksch, R., (1998), Science Education, 84(1), 71-94

Lake, D., (1999), Journal of Biological Education, 33(4), 191-198.

Lederman, N., P. Wade, and R. Bell (1998), "Assessing understanding of the nature of science: A historical perspective," In W. McComas (ed.) The Nature of Science and Science Education: Rationales and Strategies pp. 331–350. Dordrecht: Kluwer.

Lederman, N., R. Schwartz, F. Abd-El-Khalick, and R. Bell (2001), Canadian Journal of Science, Mathematics, and Technology Education, 1, 135-160.

Lederman, N., F. Abd-El-Khalick, R. Bell and R. Schwartz (2002), Journal of Research in Science Teaching, 39, 497-521.

Levy-Leblond, J., (1992), Public Understanding of Science 1, 17–21.

Martin-Dunlop, C., (2004), PhD thesis, Curtin University of Technology.

Miller, J. D., (1983), Daedalus 112(2), 29–48.

Miller, J. D., (1987), "Scientific literacy in the United States," in D. Evered and M. O'Connor (eds) Communicating Science to the Public, pp. 19–40. London: Wiley.

Miller, J. D., (1998), Public Understanding of Science 7, 203-223.

Miller, J. D., and L. Kimmel (2001), Biomedical Communications: Purposes, Audiences, and Strategies. New York: Academic Press.

Miller, J. D., (2004), Public Understanding of Science 13(3); 273-294

Norris, S., L. Phillips and C. Korpan (2003), Public Understanding of Science 12(2); 123-145

Oliver, C., (2008), PhD thesis, University of New South Wales.

Popli, R., (1999), Public Understanding of Science 8(2) 123–137.

Rennie, L., and S. Stocklmayer (2003), International Journal of Science Education, 25(6), 759-773.

Robinson, M., and D. Crowther (2001), The American Biology Teacher, 63(1), 9-14

Rutherford, F.J., and A. Ahlgren (1990), Science for all Americans. Cary: Oxford University Press USA.

Shen, B., (1975), "Scientific literacy and the public understanding of science," in S. Day (ed.) Communication of Scientific Information. Basel: Karger.

Yach, D., and S. Bialous (2001), American Journal of Public Health, 91, 1745.

NVivo qualitative data analysis software; QSR International Pty Ltd. Version 8, 2008.

Appendix A: Survey Instrument

Thank you for filling out this survey. There are no right or wrong answers; we are investigating different views of the way scientists work. This survey does not contribute to your marks and is completely voluntary.

Personal information

Age:
Gender:
Did you study science in your final year of high school? If so, which subjects?

What degree are you studying?

Questions

---

1. In your own words, how would you **define "radiation"?**

2. Some things are studied scientifically; some things are studied in other ways. From your point of view, what does it mean to **study something scientifically?**

3 – 6.

Each of the following questions will name a scientific **situation** and several **factors** that may or may not contribute to it. For each factor, please tick the box corresponding to how **important** you think that factor is to scientists working today.

3. Which factors do you think are important in determining **how objectively scientists report on their results?**

			Importance		
	Less				More
Scientific ideals, like reporting results accurately and without bias.					
The knowledge that their results will be checked by other scientists.					
The knowledge that their work affects the wellbeing of society.					
A desire to please governments and corporations that fund them.					

4. Which factors do you think are important in determining **how scientists decide whether to accept a proposed new theory?**

			Importance		
	Less				More
The body of evidence and testing that supports the theory.					
Whether the theory is elegant, explaining the evidence simply.					
Personal views or attachments to the theory or opposing theories.					
Factors like job security, fame and success.					

5. Which factors do you think are important in determining **why scientists disagree on some scientific theories?**

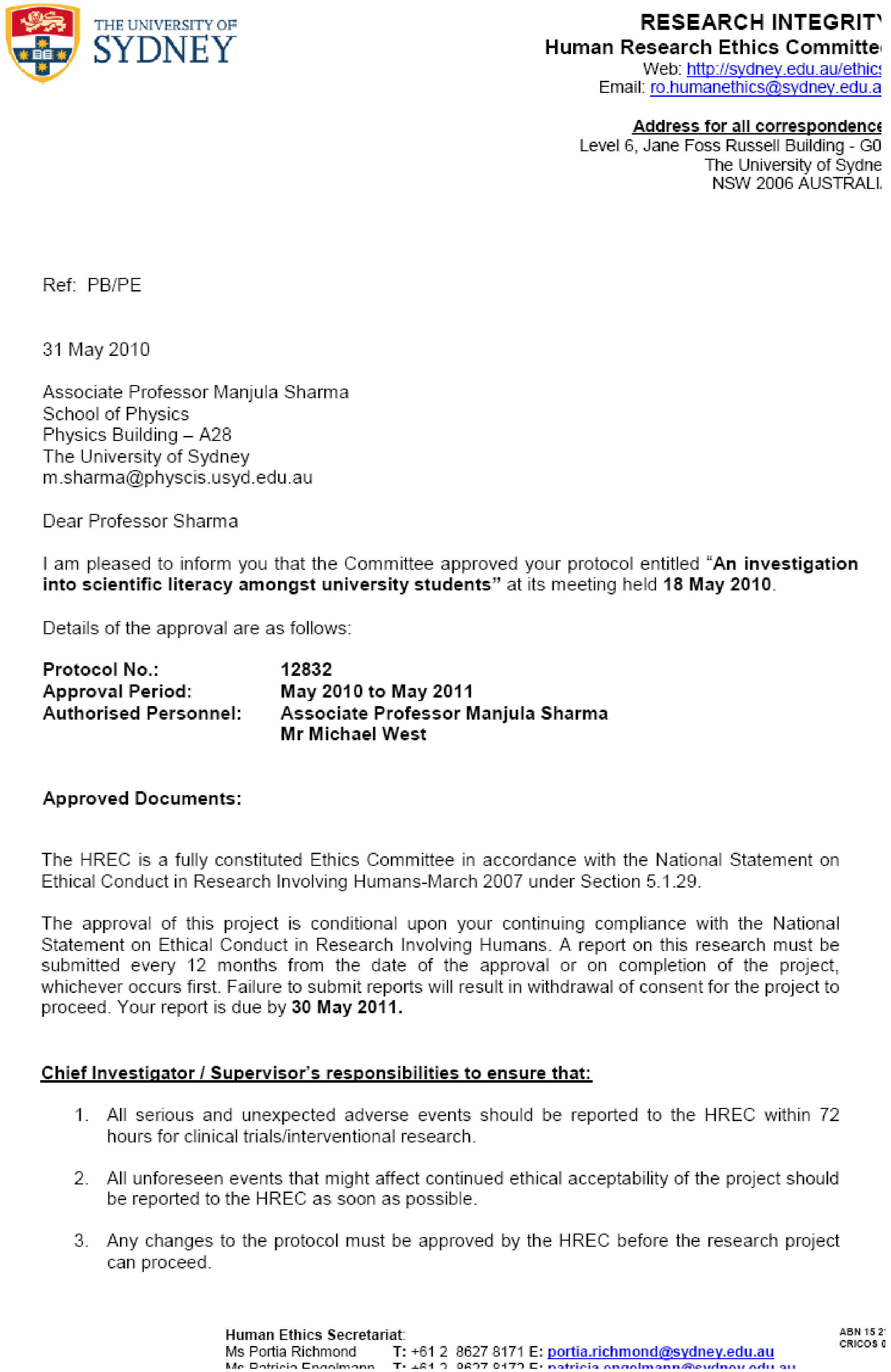
			Importance		
	Less				More
Not all of the relevant facts have been discovered.					
Different scientists interpret the available facts differently.					
Different scientists have different personal opinions or moral values.					
Different scientists have different ties to companies and governments.					

6. Which factors do you think are important in determining **whether scientific studies are affected by politics, or are impartial?**

			Importance		
	Less				More
The scientific method's requirements of objectivity and independence.					
The significant media attention that some scientific studies receive.					
The way that governments often use scientific results in policy-making.					
The way that governments often fund and regulate scientific studies.					

Thank you for your help in completing this survey.

Appendix B: Human Research Ethics Approval



THE UNIVERSITY OF
SYDNEY

RESEARCH INTEGRIT
Human Research Ethics Committe
Web: http://sydney.edu.au/ethics
Email: ro.humanethics@sydney.edu.a

Address for all correspondence
Level 6, Jane Foss Russell Building - G0
The University of Sydne
NSW 2006 AUSTRALI

Ref: PB/PE

31 May 2010

Associate Professor Manjula Sharma
School of Physics
Physics Building – A28
The University of Sydney
m.sharma@physcis.usyd.edu.au

Dear Professor Sharma

I am pleased to inform you that the Committee approved your protocol entitled "**An investigation into scientific literacy amongst university students**" at its meeting held **18 May 2010**.

Details of the approval are as follows:

Protocol No.:	**12832**
Approval Period:	**May 2010 to May 2011**
Authorised Personnel:	**Associate Professor Manjula Sharma** **Mr Michael West**

Approved Documents:

The HREC is a fully constituted Ethics Committee in accordance with the National Statement on Ethical Conduct in Research Involving Humans-March 2007 under Section 5.1.29.

The approval of this project is conditional upon your continuing compliance with the National Statement on Ethical Conduct in Research Involving Humans. A report on this research must be submitted every 12 months from the date of the approval or on completion of the project, whichever occurs first. Failure to submit reports will result in withdrawal of consent for the project to proceed. Your report is due by **30 May 2011.**

Chief Investigator / Supervisor's responsibilities to ensure that:

1. All serious and unexpected adverse events should be reported to the HREC within 72 hours for clinical trials/interventional research.
2. All unforeseen events that might affect continued ethical acceptability of the project should be reported to the HREC as soon as possible.
3. Any changes to the protocol must be approved by the HREC before the research project can proceed.

4. All research participants are to be provided with a Participant Information Statement Consent Form, unless otherwise agreed by the Committee. The following statement appear on the bottom of the Participant Information Statement: *Any person with conc or complaints about the conduct of a research study can contact the Deputy Mana Human Ethics Administration, University of Sydney on +61 2 8627 8176 (Telephone); 2 8627 8177 (Facsimile) or* [*ro.humanethics@sydney.edu.au*](mailto:ro.humanethics@sydney.edu.au) *(Email).*

5. You must retain copies of all signed Consent Forms and provide these to the HRE request.

6. It is your responsibility to provide a copy of this letter to any internal/external gra agencies if requested.

7. The HREC approval is valid for four (4) years from the Approval Period stated in this l Investigators are requested to submit a progress report annually.

8. A report and a copy of any published material should be provided at the completion o Project.

Please do not hesitate to contact the Ethics Office should you require further informatio clarification.

Yours sincerely

Associate Professor Philip Beale
Chair
Human Research Ethics Committee