Cover page

Is the environmental literacy of university students measurable?

Kerry Shephard[1], John Harraway[2], Brent Lovelock[3], Sheila Skeaff[4], Liz Slooten[5], Mick Strack[6], Mary Furnari[1] and Tim Jowett[1,2]

[1]Higher Education Development Centre, HEDC; [2]Department of Mathematics &Statistics; [3]Department of Tourism; [4]Department of Zoology; [5]Department of Human Nutrition; [6]School of Surveying; all at the University of Otago, Dunedin, New Zealand.

Abstract

We report the development and piloting of an evaluative instrument and process for monitoring the environmental literacy (EL) of undergraduate students in one large research-led university in New Zealand. The instrument addresses knowledge, affect and competencies in the general area of environmental literacy in line with this institution's adoption of EL as a graduate attribute (or in a USA context, a general-education learning outcome, and something to be fostered throughout a student's education). The instrument and associated processes were designed to fit within conventional institutional mechanisms that manage student feedback on the quality of teaching. The instrument was tested with more than 600 students from more than 8 programmes over the course of a year and its use stressed that students were anonymous within the survey. We conclude that evaluating (or in a USA context, assessing) the extent to which students acquire environmental literacy is an achievable objective and is a reasonable expectation for any higher education institution that claims to foster this attribute.

Article for peer review

Is the environmental literacy of university students measurable?

Abstract

We report the development and piloting of an evaluative instrument and process for monitoring the environmental literacy (EL) of undergraduate students in one large research-led university in New Zealand. The instrument addresses knowledge, affect and competencies in the general area of environmental literacy in line with this institution's adoption of EL as a graduate attribute (or in a USA context, a general-education learning outcome, and something to be fostered throughout a student's education. The instrument and associated processes were designed to fit within conventional institutional mechanisms that manage student feedback on the quality of teaching. The instrument was tested with more than 600 students from more than 8 programmes over the course of a year and its use stressed that students were anonymous within the survey. We conclude that evaluating (or in a USA context, assessing) the extent to which students acquire environmental literacy is an achievable objective and is a reasonable expectation for any higher education institution that claims to foster this attribute.

Introduction

Universities around the world are addressing the broad range of environmental, cultural and economic issues associated with sustainable development and global citizenship. Many universities have signed accords such as the Talloires Declaration (Association of University Leaders for a Sustainable Future, 1994) agreeing to be actively involved. Campus sustainability has become a common institutional theme alongside the research into sustainability-related subjects that our institutions are well known for. But our higher education institutions may be falling behind societal hopes and expectations in relation to the impact that we may have on our students through curriculum engagement with sustainability issues. As Sterling suggests " …*contemporary and future socio-economic and environmental challenges set an overarching context for higher education as a whole, not least, indicating the kinds of aptitudes, understanding and competencies that may be needed by our graduates, both now and for an uncertain future*" (Sterling, 2012, page 9). There is increasing evidence that at least some students (Bone & Agombar, 2011) and some employers (Business in the Community, 2010) expect higher education to rise to this challenge.

There is evidence of innovative curriculum engagement in this domain for those students who have already shown some commitment towards learning about sustainability. The discipline of environmental science, for example, is spearheading the move towards educating 'sustainability leaders' (Shriberg, 2012). For the broader student population, however, the rhetoric, and action, remains focused on 'greening the curriculum', a movement initiated in the 80s (Toyne, 1993), but still very much in 'planned-for mode' today (Sterling, 2012). There have been some specific developments, often describing the changes in student perspectives that are anticipated or hoped for. The University of Otago, for example, has recently adopted Environmental Literacy as a graduate attribute [something to be fostered by university teachers in our institution and possessed by our graduates (University of Otago, 2011) and similar in concept to a general education requirement or a core curriculum component in the USA]. We argue, however, that whatever efforts higher education is making towards education for sustainability, sustained progress is unlikely if we fail to evaluate our impact on student learning. In particular higher education needs to better understand its current impact, before investing heavily in change. Throughout higher education there are few, if any, systematic studies of how the sustainability attributes of students change as they experience higher education, as a precursor to educational and institutional change, or as a consequence (although more information on this topic is available at school level. See for example Boyes & Stanisstreet, 2012).

There are sound reasons for this deficit. Higher education is clearly not united behind the call for 'education for sustainable development', 'education for sustainability' or even 'environmental education'. In particular, university teachers hold a range of different viewpoints on their and their institution's role in this area (Shephard & Furnari, 2012) and understand concepts such as 'education for sustainability' in different ways (Reid & Petocz, 2006). Certainly not all university teachers are prepared to 'advocate for sustainability' (Jickling & Wals, 2008). The nature of the 'sustainability attributes' sought by those who 'educate for sustainability' is also highly contested. The 'aptitudes, understanding and competencies' of students,indicated by Sterling (2012) remain poorly defined and therefore difficult to measure, monitor, assess or evaluate. Some institutions address sustainability, others focus on the environment. Some include global perspectives and even cultural sensitivity in a general but often poorly defined set of aspirational objectives. In addition, and inevitable in such a poorly structured and understood domain, the tools, instruments and processes that may be employed to support this measurement are themselves subjects of intense disagreement (see for example, Buissink-Smith, Mann, & Shephard, 2011). Our own research has used and developed the Revised New Ecological Paradigm scale (NEP) to monitor changes in the environmental attitudes, or concerns, of our students (Shephard, Mann, Smith, & Deaker, 2009; Harraway, Broughton-Ansin, Deaker, Jowett, & Shephard, 2012). At the time that the original NEP was developed by Dunlap and Van Liere, in 1978, already more than 300 research studies on environmental concern had been established (Dunlap & Van Liere, 1978). More than 30 years on, academia still has profound differences of opinion about what environmental concern is, what its relationship might be to learning, and teaching, whether it is worth measuring, and if it is, how best to do it. The project is not quite deadlocked but is certainly slow moving. We should also recognise profoundly different usage of key educational terms in different parts of the world that are hampering progressive moves to common goals. The graduate attributes of Australasia, notions of graduateness in the UK and general-education learning-outcomes in the USA have much in common. Processes to determine the extent to which they are achieved by institutions, often based on cohort analysis, are generally described as 'evaluation' in Australia, NZ and the UK, but 'assessment' in the USA. For readers in the USA, much of this article is about 'general-education learning-outcome assessment'.

One group that has been making systematic progress in this area is the NAAEE (North American Association for Environmental Education). Over many years this group has developed instruments to assess students' environmental literacy and include within this domain knowledge, affect (values and attitudes) and competency (McBeth & Volk, 2009; Hollweg et al., 2011). The research described in our article borrows insights and some elements from this group's research, integrates our own experience using the NEP (Harraway, Broughton-Ansin, Deaker, Jowett, & Shephard, 2012) and took advantage of our own institution's timely adoption of 'environmental literacy' as a graduate attribute. We report here the development and piloting of an Environmental Literacy Evaluation Instrument (ELEI). We stress our use of the term 'evaluation' and define key aspects of its usage here. This instrument is not intended to assess the degree to which individual students achieve environmental literacy. Students within our research are anonymous, other than by cohort. This instrument is designed to enable our academic departments to better understand the degree to which they are fostering this attribute in their undergraduate teaching and the degree to which our undergraduate cohorts possess it.

Methods

This project was managed by an expanding research team meeting monthly. Our team comprised our own Education for Sustainability Research Group, a representative of the University's Environmental Studies Working Group, and additional occasional contributors. We started with in-depth discussions about what, exactly, environmental literacy is, or could be. Following email discussions with members of the NAAEE, we decided to adopt a definition and list of component-parts common to the NAAEE instrument (Hollweg et al., 2011), in particular so that our data could have points of reference elsewhere. Our ELEI was accordingly developed in four parts, designed to record demographic information and to evaluate affect, knowledge and competency. Other design elements of the ELEI were also agreed at this stage. The ELEI was to be capable of being described to and completed by cohorts of students in less than 30 minutes (to maximise eventual acceptability by university teachers). It was to be paper-based (to maximise response rates and to emphasise to respondents their anonymity in the evaluation). Although an early prototype did enable respondents to add hand written comments to their answers, few did so and later versions were entirely optically scan-able (Using software common to our institutional student feedback process, Remark Office OMR, Gravic Inc. PA, USA), to minimise marking and analysis time. The ELEI was also to be designed to be 'environmentally' relevant to students in 2012 living and studying in South Island, New Zealand. During this year, students would have been exposed to media reports about, and perhaps direct experience of, climate change, extreme weather, economic duress, rising graduate unemployment, whaling in the Southern Oceans, dolphin by-catch more locally, over fishing generally, the Rio 20+ Summit, oil drilling by hydraulic fracturing (fracking), land use change, polluted waterways and continuing earthquake aftershocks in nearby Christchurch.

An important aspect of the development of the ELEI was its on-going validation. The ELEI was being developed as an instrument that could be used on more than one occasion with each student by departments (say on an annual basis). It was to exist, therefore, as multiple versions, each drawing knowledge and competency elements from a pool of questions and scenarios. Our research approach assumed that versions of the ELEI would be distributed randomly and equally within a cohort and that differences between versions would not contribute to measured differences between cohorts. Nevertheless we were interested in recording responses to each version so as to better understand the instrument as a whole and to prepare for later compensation between versions in relation to on-going data gathering.

Notions of ‘the correct answer’ are addressed in the discussion, but competency elements were tested by a panel of university teachers and statistical models were fitted to score answers provided later by the students. The prototype ELEI was tested in the School of Surveying (2nd and 3rd year students) and a subsequent research instrument was developed following analysis of student responses and an interview with the staff member concerned. Because of the substantial changes following this first use, these surveying students’ ELEI data are not included in this article although subsequently first year surveying students were involved. An example of a complete ELEI is provided in Appendix 1.

Section A of the ELEI was a set of demographic questions (age, sex, programme of study, ethnicity). Respondents were also routinely invited to record a seven-digit code made up of letters from their name and combinations of numbers from their birth date. The code was designed so that they could recalculate it again in the future (with no recourse to memory) but without researchers being able to identify them through it. The data gathered are therefore available for future use by departments who may wish to continue use of the ELEI and explore individual (but anonymous) changes in EL by these students.

Section B of the ELEI was the 15 Item Revised NEP, variously thought to record ecological worldview (Dunlap, 2008), environmental attitude (Hawcroft & Milfont, 2010) or environmental concern (Dunlap & Jones, 2002), and extensively researched by our own research group (Harraway et al., 2012). For the analysis described here, individuals’ overall mean NEP scores were calculated (averaged over all 15 statements) and reported as Group NEP scores as described by Harraway et al, 2012; with scores between 1 and 5 and high scores indicative of pro-environmental worldview. Comparisons between groups can be visualised by plotting the Group NEP scores and the associated 95% confidence intervals for the scores. The process of obtaining mean NEP scores for each student ensures that the scores are approximately normally distributed and consequently, confidence intervals and hypothesis tests for group NEP scores can be obtained using conventional statistical methods based on the central limit theorem.

Section C of the ELEI was a multiple-choice knowledge test. Questions were developed by the authors but also obtained from published research and addressed a wide range of facets of knowledge identified by the NAAEE (Hollweg et al., 2011). We incorporated a core of 5 questions common to all ELEIs used within a cohort, and 5 additional questions drawn from a database. Four different versions of the ELEI were used in each cohort, each version having a different set of additional knowledge question. Marks were allocated using a method known as ‘certainty based marking’ (CBM), (Gardner-

Medwin and Gahan, 2003) where along with their selected answer, students also give their confidence level that the chosen answer is correct on a three level scale: low, medium and high. Correct answers gain 1, 2 or 3 marks (for low, medium and high confidence). Incorrect answers lose 0, 2 and 6 marks (for low, medium and high confidence). Using this marking scheme respondents do not gain systematically by either overestimating or underestimating their confidence and students do best when they have an accurate judgement about their knowledge levels (Dawid, 1986). CBM reduces the likelihood of a student with poor knowledge doing well by guessing and it allows researchers to identify misplaced confidence (Gardner-Medwin, 1995). In this analysis students with strongly held but incorrect beliefs about environmental issues are recorded as having lower environmental literacy than students who have more accurately understood their limited knowledge. Although data from this section tended to be slightly negatively skewed, with sample sizes greater than 30, parametric statistical methods may be used to calculate confidence intervals and for hypothesis testing.

Section D of the ELEI was designed to address the environmental competency of respondents and proved to be the most challenging aspect of our research. We adopted the seven competencies developed by the NAAEE and developed our own scenarios and multiple-choice statements. Our prototype ELEI (piloted in our School of Surveying) incorporated four separate short scenarios, each with multiple-choice outcomes that students were asked to select. Discussion within the research team, however, identified some discomfort about allocating a correct answer, or answers, to these initial competency measures. Different members of the team had different views on which answers were the best. In addition, the team became concerned about the time allocation needed for students to work through four scenarios and multiple-choice statements. The subsequent version of the ELEI incorporated a single scenario (on fracking – hydraulic fracturing - in New Zealand) and expected individual students to tackle multiple-choice elements related to just two competencies. The final ELEI versions (piloted in our Mathematics and Statistics Department and subsequently used for all other groups reported here) incorporated a single competency element common to all students (the capacity to identify environmental issues) and distributed four other competency elements evenly, but randomly, throughout the group (the capacity to: Ask relevant questions, Analyse environmental issues, Investigate environmental issues and Create and evaluate plans to resolve environmental issues). Two additional competencies identified by the NAAAEE (the capacity to: Evaluate and make personal judgments about environmental issues; Use evidence and knowledge to defend positions and resolve issues) proved too challenging for our group to identify and agree possible questions and answers in the time available to us and were not evaluated in this study).

The marks for each multi-choice option of each competency question in Section D were determined in such a way that the maximum score for each question was 10 and the marks given for each option chosen were directly proportional to the distribution of response from an expert panel. The expert panel consisted of 9 experienced researchers with interests in this field who independently answered the questions for all competencies. For some competencies, panel members agreed, but for others there was some variation. For example, in the "Identify Issues" question (The first Section D Question for all four tests) where each student is allowed to select one option, 100% of the panel selected Option C as the correct answer. Because of this, all of the 10 marks were given to students that chose Option C and zero marks for students that chose any of the other options. The same approach was taken for the questions where the students were asked to select two options. For instance in Question D2 of test 1 (create and evaluate plans) 19% of the panel response selected each of options A and B while 33% and 24% of the response selected options C and D. The marks given for each option were determined using the following equation: Marks for any one selected option = 10 marks × the % of the panel that chose the selected option ÷ the sum of the two highest panel response %. With this method, the mark awarded for each of option A and B is 10 × 19 ÷ 57 = 3.33. Similarly, the mark awarded for Option C is 10 × 33 ÷ 57 = 5.79 and option D: 10 × 24 ÷ 57 = 4.21. Therefore, a student choosing option A and B would receive 6.66 marks, options A and C would give 9.12 marks while a student choosing C and D would achieve 10 marks. With this scheme, as required the marks allocated to each option are proportional to the panel response and the maximum marks are 10. In contrast with the previous sections where the data are normally distributed, the distribution of the competency scores of the students was highly skewed. Therefore, we used "boot-strapping" to analyse the data (Efron and Tibshirani, 1994). Using this technique it is possible to test for differences in group-medians and also construct confidence intervals for both the group-medians and differences between group-medians.

Results

These results summarise student responses from 689 students. Most identified that they were studying one of 8 major programmes (Surveying, Tourism, Human Nutrition, Statistics, Physical Education, Health, Psychology and Zoology) although some were studying other subjects but entered this research because they had chosen a course in which the ELEI was distributed. As not all students responded to all questions the numbers in the results below do not all add to 689. Some students, for example, chose not to record their sex. Of those that did, there were 206 males and 470 females; the preponderance of females reflecting their abundance in large programmes such as Human Nutrition. The distribution of students' stated programme and year is provided in Table 1.

Programme	Year 1	Year 2	Year 3	Year 4+
Health	16.10%	6.10%	9.95%	4.44%
Surveying	10.24%	1.83%	0.00%	1.11%
Human Nutrition	22.93%	36.59%	24.61%	28.89%
Zoology	13.66%	40.24%	28.27%	15.56%
Physical Education	1.46%	3.05%	25.13%	21.11%
Psychology	19.5%	1.83%	1.05%	4.44%
Other including Statistics and Tourism	16.1%	10.36%	11.00%	24.45%
Total:	100%	100%	100%	100%

Table 1: The percentage distribution of students in the most common programmes of study in each of the four "Year of Study" groups

On affect (NEP responses)

Figure 1 shows the mean NEP of students in the major programmes of study represented in this research. By fitting a normal linear model to the data we found that the mean NEP value of Zoology students is higher than those in other programmes ($p<0.001$). Students in all programmes have significantly higher means compared with Statistics ($p<0.05$) except for Psychology ($p=0.052$) and Tourism ($p=0.160$).

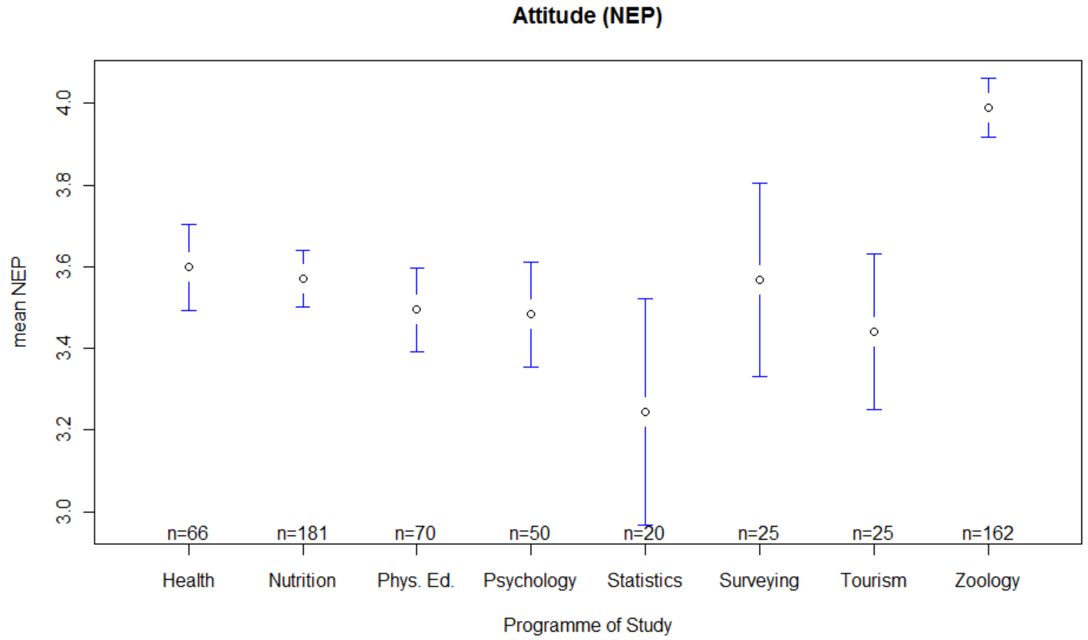

Figure 1: Confidence intervals for mean NEP for the different programmes of study groups

Males and females in this study did not have significantly different mean NEP values (The 95% confidence interval for the difference between means is (-0.021, 0.156) (p=0.13, 206 males, 473 females). Years of study did have a significant influence on the mean NEP values obtained (Anova F test, F(3, 680) =2.68, p=0.046), with a particular increase between Year 1 and Year 2 apparent (Figure 2), although as the distribution of high-scoring Zoology students was not constant within year groups [Zoology students make up a much higher proportion of the Year 2 students (40.24%) than Year 1 students (13.66%)] this result should be treated with caution.

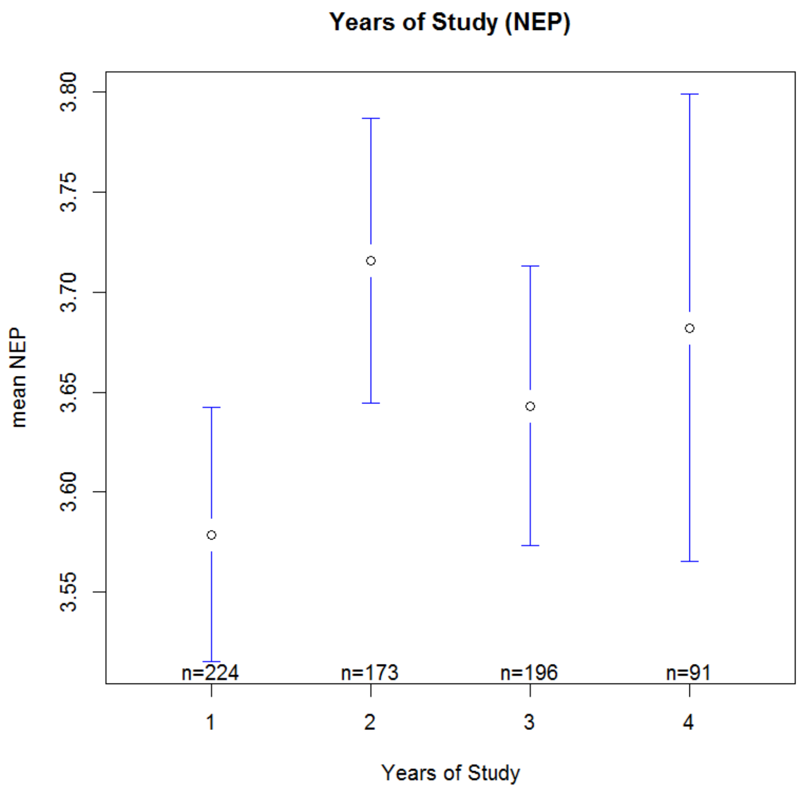

Figure 2: Confidence intervals for mean NEP for the different year of study groups.

On knowledge

The distribution of knowledge scores was slightly skewed with students in general gaining more marks than losing them with this certainty-based marking scheme. However, given the large sample size, in accordance with the central limit theorem, confidence intervals constructed using parametric statistical methods are valid. Uncorrected knowledge scores and certainty-corrected knowledge scores were significantly correlated with one another ($r=0.79$, $p<0.001$) making it reasonable in this analysis to focus primarily on certainty-corrected knowledge scores.

Figure 3 shows the group statistics and confidence intervals related to differences in certainty-corrected knowledge scores for different programmes of study. Zoology had the highest mean so was used as the reference category. Our analysis, using a normal linear model, found no significant difference between Zoology compared with Nutrition, Psychology and Surveying. Health ($p=0.006$), Physical education ($p=0.020$), Statistics ($p=0.037$) and Tourism ($p<0.001$) all had significantly lower certainty-corrected knowledge scores than Zoology. Males and females in this study did not have significantly different mean certainty-corrected knowledge scores (t-test, $p=0.44$, 206 males, 470 females) and the

95% confidence interval for the difference between means is -0.84, 1.92, p=0.44, 206 males, 470 females). Years of study did not have a significant influence on the mean certainty-corrected knowledge scores obtained (Anova F test, $F(3, 680) = 0.41$, $p = 0.747$).

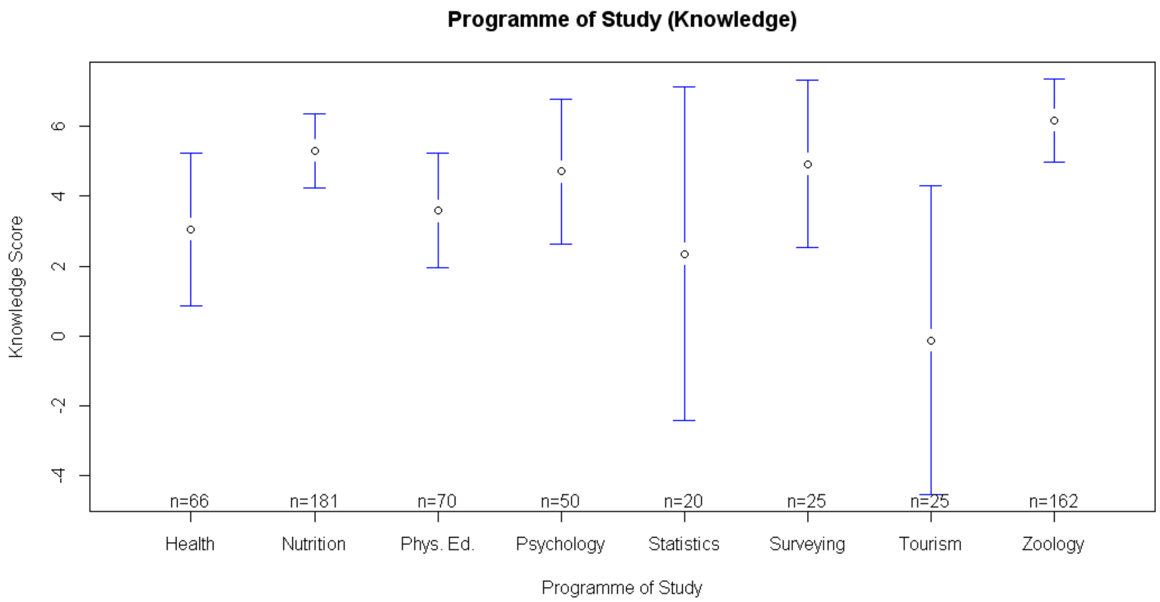

Figure 3: Confidence intervals for certainty-corrected knowledge scores for the different programmes of study groups

Our certainty-based knowledge assessment provided interesting insights into the relative confidence of males and females in answering knowledge questions. To keep the scope of the analysis to a reasonable level we restricted our analysis to the 5 questions that were included in all of the tests: Questions 4,6,7,8 and 9. For each question we counted the number of students responding within each level of certainty classified as “Low”, “Mid” and “High”, for both sexes. With the exception of Question 6 there were marked differences between males and females in the Low and High levels of certainty. In general the % of females with a low level of certainty is much higher than males and males are more likely to have a high level of certainty. Overall these results suggest that females tend to be more conservative in their assessment of their level of knowledge compared with males who tend to be more confident.

To assess whether this effect is significant we carried out a Chi-squared test of association. Significant p values (one sex associated with a high or low percentage compared with the other) were found with questions 4 (0.007), 7(<0.001), 8 (<0.001) and 9 (<0.001), but not 6 (0.091) confirming an association between sex and certainty level. While of great interest to the research, these sex differences in confidence did not translate into significant sex differences in knowledge scores. As with certainty-corrected knowledge scores, uncorrected knowledge scores of males and females was not significantly different (t-test, p=0.13, 206 males, 473 females).

On competence

The distributions of marks in Section D for both males and females were highly negatively-skewed and bimodal. As all students were given question 1, and as this question in effect had just one correct answer (as all members of the expert panel agreed on just one choice), students' answers to this question had a great effect on their competency score. Students who answered question 1 incorrectly contributed to the lower hump in the bimodal distribution. Those who answered it correctly contributed to the higher hump, and there were more of these, making the distribution skewed. Because the distributions are highly skewed, the central limit theorem can no longer be used to construct confidence intervals and perform hypothesis tests. Consequently the analysis in this section was carried out using non-parametric methods. The method of "bootstrapping" (Efron and Tibshirani, 1994) was used to construct 95% confidence intervals for the individual group medians shown in Figure 4 and confidence intervals for the differences between these medians were calculated. The process does not generate p values but differences between medians are considered significant at the 95% level if the 95% confidence interval for a difference excludes zero. Zoology had the highest median so Zoology was used as the reference category with other groups being compared with Zoology. Our analysis found no significant difference between Zoology compared with all programmes other than Tourism which had a significantly lower median competency score (CI for the difference of medians, 4.02, 7.47). The 206 males and 473 females in this study did not have significantly different median competency scores. Years of study also did not have a significant influence on the competency scores obtained.

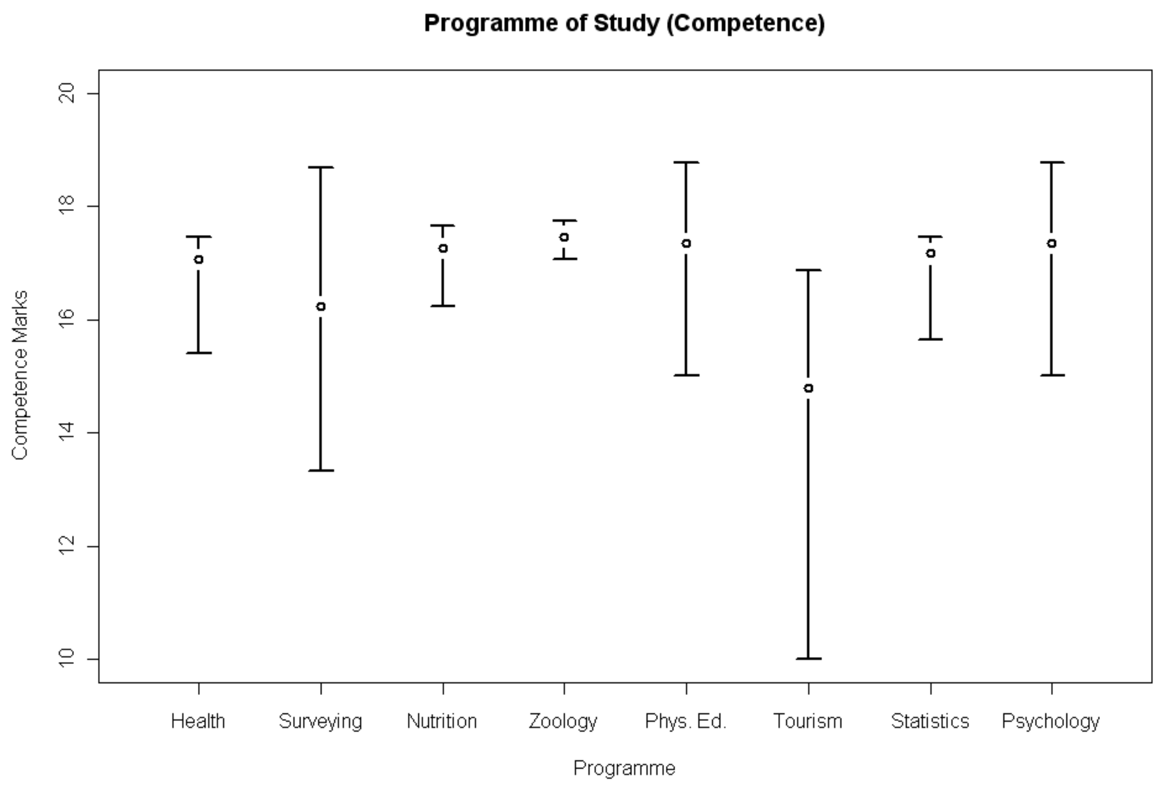

Figure 4: Confidence intervals for competency scores of the different programmes of study groups.

Combining affect, knowledge and competency

Each section of the ELEI results in a numerical score for each student for each of affect, knowledge and competency. To test the extent to which one variable correlates with another, correlation statistics were calculated on standardised data. The Pearson correlation coefficients, p values and coefficient of determination (the proportion of variability of one variable described by the other) are provided in Table 2 for data from 687 students. In all correlations, the relationship between the components is very weak with small correlation coefficients and the proportion of variability of one variable described by the other of less than 3%. Because the p-values are all less than 0.05 there is evidence that the variables are to some extent correlated; but because the correlation coefficients are very low the variables can be considered to be effectively independent of one another.

Variables to correlate	Correlation coefficient	P value	Coefficient of determination
Knowledge/affect	0.12	<0.001	1.15%
Knowledge/ competency	0.15	<0.001	2.4%
Affect/competency	0.11	0.02	1.1%

Table 2: Correlation coefficients, p values and coefficients of determination for the three correlation analyses for knowledge, attitude and competency

Discussion

Our discussion needs to focus on three important questions. Is it possible to measure environmental literacy or is this attribute inaccessible to any form of educational quantification? Do the measurements recorded here make sense in the broader context of higher education in the 21st-century or are they just interesting trivia? And how will these results translate into changed institutional practice in the future?

Our results suggest that it is possible to measure the environmental literacy of undergraduate students. Clearly there are issues related to the definition of environmental literacy and the nature of measurement. There are also issues related to measurement at the individual level or the cohort level and the nature of evaluation rather than assessment. Using a set of understandings about the nature of environmental literacy developed over many years, in particular by the NAAEE (Hollweg et al., 2011), and emphasising the anonymity of students within the process, our research and data described here suggests that the environmental literacy of undergraduate students in higher education is accessible to some form of quantification. In particular we have chosen to adopt a 'strong' interpretation of environmental literacy (Stables & Bishop, 2001); one that includes what students know about the environment (knowledge), what skills they have to put this knowledge to effect (competence) and what they might choose to do with this knowledge and skill (affect). We accept notions of the attitude/behaviour gap in particular as applied to pro-environmental behaviour (Hargreaves, 2012) and emphasise that our interpretation of environmental literacy says little about how students will subsequently behave. In many respects this distinction is common in higher education and we also accept the difference between functioning knowledge and declarative knowledge (Biggs & Tang, 2007). Engineers, for example, may learn how to make safe structures at university, but ensuring that they do so may be beyond the efforts of higher education. We also accept, and within our research group endorse, different viewpoints on the nature of the instruments used in this research. Academic departments will wish to challenge the level and detail of knowledge addressed here, and hopefully supplement our questions with their own. It is, of course, difficult to design questions about 'the environment' that are equally applicable to, and understandable, to all disciplines and it is likely that some of our questions would be seen as over simplistic by more knowledgeable students. Many find fault with the NEP, in particular challenging the nature of the factors that it addresses (see for example Noblet et al 2012) although substantial research supports its use as a useful measure of environmental attitude (Hawcroft and Milfont, 2010). There are many alternative instruments to the NEP that could be used within the ELEI to address affect. Different possible interpretations of environmental competence are addressed below, but here it is necessary to be unequivocal in emphasising the complexity of the concept of environmental competence.

We emphasise the advantages of using a cohort-based evaluation instrument rather than an individual-based assessment instrument. Evaluation at the cohort level allows students to be anonymous in the process and is likely to result in more honest reflections in particular of affect than if students felt that their response to the instrument might in some way influence their degree result. We also emphasise the advantage of using an instrument compatible with conventional university approaches using similar instruments. Most universities nowadays ask their students for feedback on the quality of the teaching they have received. This is also through an anonymous cohort process rather than through an identified-individual process. In most institutions the personnel, structures and processes needed to evaluate the extent to which students become environmentally literate are already in place. We urge those with an interest in evaluating the achievements of higher education to extend the evaluation domain to include pressing higher education and societal matters, such as the sustainability, and environmental, literacies of higher education students.

The results described here do make sense in the broader context of higher education. Although it is difficult to make clear comparisons beyond the institution, some comparative data are available, particularly where standard instruments such as the NEP are used. University of Otago students have, for example, NEP scores in the upper ranges of those described internationally by Hawcroft & Milfont, 2010. The data are also relatively consistent with previous research at the University of Otago. As expected, zoology students tend to have higher certainty-corrected knowledge scores in comparison with some other groups of students. It is noteworthy that overall knowledge scores were very high in all cohorts, although this is obscured somewhat in reporting certainty-corrected knowledge scores. Those who develop the ELEI further may wish to reassess the relative advantages of using an absolute knowledge scale in comparison with a certainty-corrected scale; no doubt reassessing the relative virtues of being confident and of being correct. It is also possible that in later versions of this instrument (perhaps particularly in department-specific versions) users will have higher expectations of knowledge and use more difficult questions and this may create greater discrimination between programmes of study. We were surprised that our data did not find a sex difference in affect (NEP) score as previous research frequently does (Shephard et al., 2009) but there are substantial sex imbalances between cohorts of students in the study and therefore a range of interaction effects. Our observation that the environmental knowledge levels of males and females are essentially the same but that confidence in knowledge is greater in males than in females is a new result and does need to be explored further. There is a wealth of information available within this data for any university department with developing interests in the environmental literacy of their students. From a departmental perspective the results also emphasise the need for instruments that measure EL to incorporate separate tools for knowledge, affect and competency. Each does not correlate with the others. For example, individuals, and cohorts, may be knowledgeable about the environment but also be relatively environmentally incompetent and show only modest dispositions towards environmental protection.

The research has highlighted, particularly in the minds of the researchers, some of the challenges involved in measuring environmental literacy. Measuring environmental competence is likely to remain a particularly challenging task. Even with relatively clearly defined aspects of competence, knowledgeable and involved university teachers in this study did not agree on the 'correct answers' within a multiple-choice situation. It is possible to formalise this divergence of opinion as done here and it may be that this will be the best long-term solution to measuring environmental competence where the problems are themselves wicked. But this does add an extra degree of complexity to the process. Added to this, our research group had extended discussion on what environmental competence actually comprised. Using an extreme example to illustrate the situation, we needed to find approaches to quantify environmental competence that enables very different competencies, and their underlying belief systems, to be considered equivalent to one another. In this model, proponents of nuclear energy as a solution to global warming would be considered equally environmentally literate to opponents, if their arguments are similarly well constructed. It seems likely that such contentious issues will re-emerge in the public debate about environmental literacy in the future.

If higher education institutions do seek to routinely evaluate student attainment of environmental literacy there are still important processes to explore. Evaluation of change is a particularly challenging feat. Single measures of complex outcomes such as EL are unlikely to provide useful indicators of institutional performance (Astin and Lee, 2003). Several studies have investigated a range of approaches and the ELEI described here could be used in several ways to explore change. Paired statistical tests could be used in a before and after format on a single cohort (Harraway et al., 2012). Logistic regression modelling (Teisl et al., 2011) or multinomial regression modelling (Jowett et al. in press) could be used on different, but sequential, cohorts (comparing, for example, first year students with second year students at a particular point in time). A longitudinal modelling approach could combine many cohorts in a single statistical exploration of change over extended time periods (Shephard et al., 2012; Shephard et al, in preparation).

Even if this instrument does provide realistic interpretations of students' environmental literacy there are many challenges to formal adoption. Further progress will require considerable institutional commitment to evaluating the environmental literacy of graduates and there is no doubt that a range of structural and philosophical problems limit the integration of environmental literacy into the higher education curriculum. Nevertheless, this research and related work internationally is focused on finding workable approaches, should higher education choose to adopt them.

Acknowledgements

The authors would like to acknowledge the help of over 600 anonymous students, numerous university teachers, several heads of academic departments and colleagues from around the University (including the University's Environmental Studies Working Group and Sustainability Coordinator) for their efforts in helping to develop and to pilot the ELEI. We would also like to acknowledge the developers of the NAAEE Framework for Assessing Environmental Literacy, who provided invaluable advice and feedback during the development of the ELEI and numerous researchers worldwide for permission to use published questions for our knowledge section.

References

Association of University Leaders for a Sustainable Future. (1994). *The Talloires Declaration 10 Point Action Plan, updated version*. Retrieved February 11 2013 from http://www.ulsf.org/pdf/TD.pdf

Astin, A. W., & Lee, J. J. (2003). How risky are one-shot crosssectional assessments of undergraduate students? Research in Higher Education, 44(6), 657‐672.

Biggs, J., & Tang, C. (2007). Teaching for quality learning at university. Maidenhead: SRHE and OUP.

Bone, E., & Agombar, J. (2011). 2011: First-year student attitudes towards, and skills in, sustainable development. Retrieved January 24, 2013, from http://www.heacademy.ac.uk/resources/detail/sustainability/esd_First_Year_Students_report

Boyes, E., & Stanisstreet, M. (2012). Environmental Education for Behaviour Change: Which actions should be targeted? *International Journal of Science Education*, *34*(10), 1591–1614. doi:10.1080/09500693.2011.584079

Buissink-Smith, N., Mann, S., & Shephard, K. (2011). How Do We Measure Affective Learning in Higher Education? *Journal of Education for Sustainable Development*, *5*, 101–114. doi:10.1177/097340821000500113

Business in the Community, (BICT). (2010). Leadership skills for a sustainable economy. Retrieved January 24, 2013, from http://www.bitc.org.uk/go.rm?id=34082.

Dawid A.P. (1986) Probability forecasting, In *Encyclopaedia of Statistical Sciences*, ed. Kotz, S., Johnson, N.L. & Reid, C.B. . Vol. 7, 210-218.

Dunlap, R. E. (2008). The NEP Scale: From marginality to worldwide use. *Journal of Environmental Education*, *40*(1), 3–18.

Dunlap, R. E., & Jones, R. E. (2002). Environmental concern: Conceptual and measurement issues. In R. . Dunlap & W. Michelson (Eds.), *Handbook of Environmental Sociology* (pp. 482–524). Westport, CT: Greenwood Press.

Dunlap, R. E., & Van Liere, K. . (1978) The new environmental paradigm. *Journal of Environmental Education*, *9*, 10–19.

Efron, B. and Tibshirani, R.J. (1994) *An introduction to the bootstrap* New York, Chapman & Hall/CRC Press.

Gardner-Medwin, A.R. (1995), Confidence Assessment in the Teaching of Basic Science, *Association for Learning Technology Journal*, 3 (1) 80-85.

Gardner-Medwin, A.R. and Gahan, M. (2003), Formative and Summative Confidence-Based Assessment, Proc. 7th International Computer-Aided Assessment Conference, Loughborough, UK, pp. 147-155. Retrieved from www.caaconference.com/ 25th January 2013.

Hargreaves, T. (2012). Questioning the virtues of pro-environmental behaviour research: Towards a phronetic approach. *Geoforum*, *43*(2), 315–324. doi:10.1016/j.geoforum.2011.09.006

Harraway, J., Broughton-Ansin, F., Deaker, L., Jowett, T., & Shephard, K. (2012). Exploring the Use of the Revised New Ecological Paradigm Scale (NEP) to Monitor the Development of Students' Ecological Worldviews. *The Journal of Environmental Education*, *43*(3), 177–191. doi:10.1080/00958964.2011.634450

Hawcroft, L., J., & Milfont, T. (2010). The use (and abuse) of the new environmental paradigm scale over the last 30 years: A meta-analysis. *Journal of Environmental Psychology*, *30*, 143–158.

Hollweg, K., Taylor, J., Bybee, R., Marcinkowski, T., McBeth, W., & Zoido, P. (2011). Developing a framework for assessing environmental literacy. Washington, DC: North American Association for Environmental Education. Retrieved February 11 2013 from http://www.naaee.net/framework

Jickling, B., & Wals, A. E. J. (2008). Globalization and environmental education: looking beyond sustainable development. *Journal of Curriculum Studies*, *40*, 1–21. doi:10.1080/00220270701684667.

Jowett T, Harraway J, Lovelock B, Skeaff S, Slooten L, Strack M and Shephard K Multinomial-regression modelling of the environmental attitudes of higher education students. *Journal of Environmental Education* (in press).

McBeth, W., & Volk, T. L. (2009). The National Environmental Literacy Project: A Baseline Study of Middle Grade Students in the United States. *The Journal of Environmental Education*, *41*(1), 55–67. doi:10.1080/00958960903210031

Noblet C L , Anderson M & Teisl M F (2012): An empirical test of anchoring the NEP scale in environmental ethics, Environmental Education Research, 1–12, iFirst Article

Reid, A., & Petocz, P. (2006). University Lecturers' Understanding of Sustainability. *Higher Education*, *51*(1), 105–123. doi:10.1007/s10734-004-6379-4

Shephard, K, Harraway, J., Lovelock, B., Skeaff, S., Slooten, E., Strack, M., & Jowett, T. (2012). Monitoring changes in the sustainability attributes of higher education students in a New Zealand university. In W. Leal (Ed.), *Sustainable Development at Universities: New Horizons* (p. xx). Frankfurt: Peter Lang Scientific Publishers.

Shephard, K, Mann, S., Smith, N., & Deaker, L. (2009). Benchmarking the environmental values and attitudes of students in New Zealand's post-compulsory education. *Environmental Education Research*, *15*, 571–587. doi:10.1080/13504620903050523

Shephard, Kerry, & Furnari, M. (2012). Exploring what university teachers think about education for sustainability. *Studies in Higher Education*, 1–14. doi:10.1080/03075079.2011.644784

Shriberg, M. (2012). Developing Sustainability Leaders: Applying a Transformative Leadership Framework at the University of Michigan, Paper presented at the annual meeting of the AASHE, D. L. Lawrence Convention Center, Pittsburgh, PA, December 2012. Retrieved from http://www.allacademic.com/meta/p521319_index.html

Stables, A., & Bishop, K. (2001). Weak and Strong Conceptions of Environmental Literacy: Implications for environmental education. *Environmental Education Research, 7*, 89–97. doi:10.1080/13504620125643

Sterling, S. (2012). The future fit programme. Higher Education Academy.

Teisl, M., Anderson, M., Noblet, C., Criner, G., Rubin, J., & Dalton, T. (2011). Are Environmental Professors Unbalanced? Evidence From the Field. *The Journal of Environmental Education*, *42*(2), 67–83. doi:10.1080/00958961003705899

Toyne, P. (1993). *Environmental Responsibility: An Agenda for Further and Higher Education.* London: HMSO.

University of Otago. (2011). University of Otago Teaching and Learning Action Plan. Retrieved February 11 2013 from http://www.otago.ac.nz/staff/otago027123.pdf

Appendix 1 An example of one version of the ELEI