Longitudinal analysis of the environmental attitudes of university students

Authors

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

[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.

Corresponding author

Professor Kerry Shephard, Higher Education Development Centre, University of Otago, Dunedin, New Zealand, Tel 0064 3 479 8439, kerry.shephard@otago.ac.nz

Abstract

This article addresses the important questions that higher education institutions ask concerning their impact on their students' sustainability-related attributes "How do our students' worldviews change as they experience higher education with us?" The process of monitoring such a dynamic entity is fraught with statistical complexity but may not be impossible for an institution willing to ask whether or not its educational efforts in 'education for sustainability', 'education for sustainable development' or 'environmental education', and campus sustainability developments, are paralleled by changes in the attitudes of its students. We describe here a longitudinal survey process based on the revised New Ecological Paradigm scale, with two cohorts of students, in three programmes of study, operating over four years, with multiple survey inputs by each student. We implemented the longitudinal analysis using a linear mixed-effects model and describe here the development and testing of this model. We conclude that higher education institutions can benchmark the sustainability attributes of their students and monitor changes, if they are minded to. We invite higher education practitioners worldwide to join us in further developing suitable research instruments, processes and statistical models; and in further analysing the assumptions that link higher education to sustainability and to global citizenship.

Key words

Monitoring change; sustainability attitudes; learning in the affective domain; longitudinal modelling; linear mixed-effects model

Longitudinal analysis of the environmental attitudes of university students

Abstract

This article addresses the important questions that higher education institutions ask concerning their impact on their students' sustainability-related attributes "How do our students' worldviews change as they experience higher education with us?" The process of monitoring such a dynamic entity is fraught with statistical complexity but may not be impossible for an institution willing to ask whether or not its educational efforts in 'education for sustainability', 'education for sustainable development' or 'environmental education', and campus sustainability developments, are paralleled by changes in the attitudes of its students. We describe here a longitudinal survey process based on the revised New Ecological Paradigm scale, with two cohorts of students, in three programmes of study, operating over four years, with multiple survey inputs by each student. We implemented the longitudinal analysis using a linear mixed-effects model and describe here the development and testing of this model. We conclude that higher education institutions can benchmark the sustainability attributes of their students and monitor changes, if they are minded to. We invite higher education practitioners worldwide to join us in further developing suitable research instruments, processes and statistical models; and in further analysing the assumptions that link higher education to sustainability and to global citizenship.

Introduction

The educational enterprises of 'education for sustainability', 'education for sustainable development' and 'environmental education' have differing backgrounds and rationales, but all seek changes in students that extend beyond simple knowledge about sustainability or the environment. Within the complex educational objectives intrinsic to "what students know, what skills they have to put this knowledge to effect and what they may choose to do with the knowledge and skills at their disposal" differences in approach, and emphasis, focus predominantly on the directness of planned-for influence on learner behaviour. Pedagogic approaches may be described as democratic (Jensen & Schnack, 2006), pluralistic (Kronlid and Österbergh, 2011) or otherwise identify limits to values-based expectations (Shephard, 2008); but it may be reasonable to state that all three enterprises hope to have an impact on the environmental concerns, or attitudes, of learners, in as much as these affective attributes at least enable sustainable behaviour. It is this domain that this article addresses.

Higher education, internationally, may not be uniformly disposed to these complex objectives. Particularly problematic for higher education may be students in general, rather than those who specifically choose to study environmental or sustainability-focused courses. Institutions vary from making relatively little commitment, or accommodation, to these objectives (Dawe, Jucker, and Martin 2005; Shephard 2010; Sterling & Scott, 2008) to whole-scale transformations, in rhetoric or operation or both (Association of University Leaders for a Sustainable Future, 1994).

This article focuses on the pressing educational matter of how higher education might recognise that it is, or is not, succeeding. It seems, to us that educational development is misplaced if it does not constantly, and progressively, monitor its effect. Without this monitoring built into the change process, how would we know that our educational efforts are effective or even point in the right direction? We also note that 'We cannot improve at scale what we cannot measure' is one of six core principles for educational improvement adopted by the Carnegie Foundation http://www.carnegiefoundation.org/improvement-research/approach). But it should be clear that the educational outcomes involved are complex, including knowledge, competencies and affect (values, attitudes, dispositions that may enable learners to make sustainably, or environmentally, directed choices) (Shephard, 2008; Cotton, Warren, Maiboroda, & Bailey, 2007; Shephard & Furnari, 2012). It seems likely that few other higher education objectives are this complex; although it has been argued that medical education may be similarly so (Shephard, 2008). Conventional university assessment processes, focussing on individual assignments and examinations, may not provide sufficient or appropriate means to monitor the attainment of ES, ESD or EE objectives (Krathwohl, Bloom, & Masia, 1988; Buissink-Smith, Mann, & Shephard, 2011) or equivalent complex, or 'wicked' competencies (Knight & Page, 2007). The general area of environmental concern or environmental attitude has received special attention since the 1970s when one particular instrument (the New Environmental Paradigm scale, or NEP) was developed to monitor the existence of an 'environmental paradigm', as an alternative to the 'social paradigm' thought to be dominant in societies at the time (Dunlap, 2008). Several decades on, and following extensive use, validation and development, the NEP (now the revised New Ecological Paradigm scale) is still especially highly regarded. Following an extensive review of its use over a 30-year period, Hawcroft & Milfont (2010, p. 151) concluded that '*until a gold-standard EA* [environmental attitude] *measure has been widely accepted, it is probably advisable for researchers to continue using the NEP scale as a standardised measure of EA*'.

The number of factors that arise from the use of the NEP has been reviewed by Dunlap et al. (2000). These authors advise caution in interpreting the NEP results in other than a single dimension but accept that where there is good internal evidence that more than one factor exists and where analysis of each factor provides a useful means of interpretation, separate descriptions of each should be retained. Shephard et al. (2009) identified four factors and used their contributory NEP items to describe four consequential implied "tendencies" (to conserve, to recycle, to be cautious about the future, and to support animal and plant rights). These authors acknowledged the uncertain nature of this four-factor model, of describing identified factors from the NEP as behavioural "tendencies", the advantages of allocating simple descriptive names to them, and suggested that research would continue to address these facets of sustainability. The four-tendency model was developed further by Harraway et al (2012) and the present research continues this process of investigating the usefulness of the four-factor model.

Even where appropriate research or measurement instruments are available to determine the environmental concern or attitudes of students, identifying progressive change in these things is an additional problem. Institutions may grapple with these questions at the individual, cohort or institutional level. They may involve many individuals, multiple test points, extended time periods and multiple overlapping cohorts. In addition, the subjects of this enquiry, students, may be volunteers, unreliable attendees and move between cohorts as they pass, fail and retake courses and change programmes. The process of monitoring such a dynamic entity is fraught with statistical complexity but may not be impossible for an institution willing to ask whether or not its educational efforts in ES, ESD or EE run parallel to the environmental attitudes of its students. Several studies worldwide have investigated a range of approaches already. Paired statistical tests may be used in a 'before and after' format on a single cohort (see for example, Harraway et al., 2012; Packer, 2009; Anderson et al., 2007). Logistic regression modelling (Teisl et al., 2011) or multinomial regression modelling (Jowett et al., 2013) may be used on different, but sequential, cohorts (comparing, for example, first year students with second year students at a particular point in time). Longitudinal modelling approaches are different and make optimum use of data available from dynamic systems. They are widely used in epidemiological studies. Our approach has been briefly described previously for a single cohort of students (Shephard et al., 2012) but not so far for data combining more than one cohort. Multi-cohort analysis is generally necessary to fully describe change processes.

The research described in this article brought together an educational specialist, university teachers and researchers from several departments including statisticians, several years of experimentation and development, several supportive Heads of Department, many willing students, and most recently, institutional commitment to foster all students' attainment of environments literacy (described in more detail by Shephard et al, 2013). We describe here the use of a research instrument, multiple recordings over several years of the environmental attitudes of anonymous student respondents, in two cohorts, studying multiple programmes. We describe here the development and efficacy of a longitudinal statistical model that enables this data to answer the departmental, or institutional, question 'do the environmental attitudes of our students change as they experience higher education with us?'

Materials and methods

Survey processes

Environmental attitudes were measured in the survey using the revised 15-item NEP (New Ecological Paradigm) scale (Dunlap et al., 2000). Participants were asked to rate the level of agreement for each statement on a 5-point Likert-like scale. Individuals' responses were combined into a summated NEP score (by reversing the scores of alternate items) where higher overall scores indicate stronger pro-environmental attitudes (on a scale of 1-5). Previous research (Shephard et al 2009; Harraway et al 2012) used factor analysis to identify 4 tendencies within the NEP data (to conserve, to recycle, to be cautious about the future, and to support animal and plant rights). For the current research we measured the tendency NEP scores by averaging the scores across the statements identified within each of the four tendencies. The participants were also asked to provide, on the survey form, some additional socio-demographic information to aid the analysis (year of study, sex, and self-reported programme affiliations). The back of the survey form contained text describing the research, stressing that participation in the project was entirely voluntary and anonymous in accordance with a University of Otago ethical research approval. To allow for follow up of individuals and at the same time preserving anonymity, participants were asked to write a confidential code word on their survey (described in detail in Harraway et al 2012).

Students entered this research as two cohorts; those starting their university education in 2009 and those starting in 2010. Within each cohort each student completed a NEP survey in a first year statistics course and was followed for the subsequent two years in second and third year courses in one of Human Nutrition, Surveying or Zoology. Students completed the NEP at different stages in each course, depending on when the course teacher though it best fitted their programme. Although most students in each class volunteered to be involved, all of the same students did not necessarily appear in the next class where the NEP survey was used. In addition, some forgot their confidential code or miscalculated it making it impossible to match their second or subsequent entry to their first. To be included in this research, individual students needed to have completed a NEP survey on two or more occasions and to have remembered or correctly recalculated their individual code. The statistical model is sufficiently robust to be able to incorporate irregular sampling times and random missing values but not systematic missing values (so, for example, could not allow new programmes to enter the analysis part-way). The sample size (in terms of the number of students able to be included successfully in this research) for Cohort 1 and 2 was 89 and 125 respectively giving a total sample size of 214, a considerably smaller number from that potentially possible. Table 1 gives the sample numbers of included students per Cohort by Programme of Study and sex. In total, 512 individual NEP responses (214 for Cohort 1 and 288 for Cohort 2) were obtained from 214 included students. For Cohort 1, 52 (58%) of the students responded on two occasions, 25(29%) responded on 3 occasions and 12 (13%) responded on four occasions. For Cohort 2, 87 (70%) of the students responded on two occasions and 38 (30%) responded on three occasions.

Table 1: Sample characteristics for the longitudinal study.

		Female	Male
Cohort 1	Surveying	4(10%)	36(90%)
	Human Nutrition	25(93%)	2(7%)
	Zoology	19(86%)	3(14%)
	Totals	48(54%)	41(46%)
Cohort 2	Surveying	10(24%)	31(76%)

Human Nutrition	40(93%)	3(7%)
Zoology	23(56%)	18(44%)
Totals	73(59%)	52(41%)

Results

Basic data analysis

Prior to model development, mean NEP scores were obtained by combining all of the observations in the longitudinal study. Females and Males had means of 3.76(SD=0.55) and 3.42 (SD=0.60) respectively. Note that there is a large discrepancy in the balance of sexes between the programmes, and in our data sets, with Surveying and Human Nutrition having a high and low percentage of males respectively. The means for the Programmes of Study were: 3.31(SD=0.52) for Surveying, 3.60(SD=0.56) for Human Nutrition and 4.06(SD=0.44) for Zoology.

To get a crude measure of differences over time, means of first and third year students were compared. First year students were found to have a mean of 3.54(SD=0.58) and third year 3.68 (SD=0.57). It should be noted that standard confidence intervals or tests for differences of means are not valid using this data because there are multiple NEP scores from individuals and therefore the observations are not independent of each other. The longitudinal analysis given below is able to produce confidence intervals and valid tests because the lack of independence between observations is allowed for within the statistical models.

Longitudinal Analysis

We implemented the longitudinal analysis using a `linear mixed-effects model' (LME). A distinctive feature of LME models is that each subject (in our case, students) in the study has a unique response profile over time and each response profile is modelled by both fixed and random effects. The fixed effects are modelled by parameters that are common to all of the participants in the study, either as cross-sectional or longitudinal fixed effects. In this analysis, the cross-sectional fixed effects are the *Programme of Study* and *Cohort* and the longitudinal fixed effect is *Time.* In a "random intercept" LME model it is assumed that each student has a unique `random effect' where the i^{th} student has a deviation from the population mean (for a given combination of fixed effects) equal to b_{1i} and this deviation is assumed to remain constant throughout the full extent of the study period. The main objective of LME modelling is to find estimates for fixed effects. Estimates of the random effects themselves are not of major interest but random effects do need to be allowed for within the model because they provide a mechanism through which valid inference of the fixed effect parameters can be obtained. Our main interest is to assess the extent to which students' NEP responses change with time. Because multiple observations are taken from individual students, the observations can be thought of as being student-specific "clusters" of correlated NEP scores. This correlation acts to reduce the effective sample size. If this correlation were not allowed for then the confidence intervals for the cross sectional effects would be unrealistically narrow which would result in invalid inference about the significance of the cross sectional effects.

By including "*Cohort*" (C) as an explanatory variable in the model we were able to investigate whether there are any differences in the NEP scores of the students relates to the different cohorts. We defined *Cohort* as a binary variable where a value of 1 was allocated to each student in the second cohort and 0 for a student in the first cohort. The *Programme* effect was modelled using a factor with three levels represented by the three programmes of study: Zoology (the reference category), Human Nutrition and Surveying. The variable "*Time*" was modelled as a continuous variable measured in units of weeks. Because of the strong association between sex and Programme of study only one of these variables can easily be accommodated in the model and sex was not included.

Model development

Model development was in stages, initially using overall NEP scores and subsequently with the NEP scores separated into four tendencies. Initial exploration used a main-effects, random-intercept model. This was adapted by incorporating first slope, and subsequently interaction, into the model to test for improved fit.

Overall NEP score, main-effects, random-intercept model

We initially fitted a main-effects, random-intercept model to the data as given below

$$Y_{ij} = \beta_0 + \beta_1 Hunt_i + \beta_2 Surv_i + \beta_3 Cohort_i + \beta_4 Time_{ij} + b_{1i} + e_{ij}$$

Where Y_{ij} denotes the response variable (i.e. mean overall NEP or mean tendency NEP) for the i^{th} individual (where there are N individuals) at the j^{th} occasion ($j = 1,\ldots,n_i$) where n_i is the number of occasions that the NEP survey was taken by the i^{th} individual. The subject-specific random intercept effects b_{1i} are assumed to be random with a mean of 0 and a variance of $\mathrm{var}(b_1) = \sigma^2_{b_1}$.

The fixed effect parameter β_0 is the overall intercept while β_4 represents the overall longitudinal effect of time common to students from all programmes while β_1 and β_2 are the cross-sectional main effects of the *Programme of study* with Zoology as the reference category.

Finally, e_{ij} represents the "within student measurement error" where $\mathrm{var}(e_{ij}) = \sigma^2$.

The above model is referred to as "main effects only" because it does not include parameters for the interaction between *Programme* and *Time* (refer below for a discussion of the interaction model).

The above model was fitted to the data using STATA, a statistical software program (Statacorp LP, Texas, USA, http://www.stata.com/company/). Table 2 below presents the estimated regression coefficients.

Table 2 Regression coefficients and other outputs from the main effects random incercept longitudinal model

Fixed Effect	Estimated regression coefficient	Standard Error	p-value	95% Confidence Interval	
Intercept	4.04	0.066	<0.001	3.91	4.16
Programme (Zoology is reference)	-	-	-	-	-
Human Nutrition	-0.46	0.079	< 0.001	-0.61	-0.30
Surveying	-0.75	0.077	< 0.001	-0.90	-0.60
Cohort	0.014	0.064	0.83	-0.11	0.13
Time	0.00046	0.00037	0.21	-0.00026	0.0012

A Wald test produced a test statistic of 96.76 and a p-value less than 0.001 showing there is evidence that at least one of the main-effects parameters is statistically significant. This is shown in Table 2 where we see that the *Programme* effect is significant. Also, a likelihood-ratio test (Neyman, 1933) gave a p-value < 0.001 showing that the mixed-effects model provided significantly improved fit compared with conventional linear regression.

By reference to Table 2 we see that the *Time* effect is not significant. Therefore there is no evidence of change over time in the average NEP score of students enrolled in the three *Programmes of study* investigated in this project. *Cohort* was also found to have a non-significant effect and consequently there is no evidence to suggest a difference in mean NEP scores between the Cohorts.

Overall NEP Scores, main-effects, random-intercept and random-slope model

Having fitted the main-effects random-intercept model we then elected to see whether there was any justification for using a more sophisticated model that allows the random effects to vary in terms of both intercept and slope. The “main-effects, random-intercept and slope” model given below

$$Y_{ij} = \beta_0 + \beta_1 Hunt_i + \beta_2 Surv_i + \beta_3 Cohort_i + \beta_4 Time_{ij} + b_{1i} + b_{2i} Time_{ij} + e_{ij}$$

is the same as the random intercept model with the addition of the b_{2i} random-slope effect that allows the student-specific response profiles to have different slopes as well as intercepts.

A likelihood-ratio test was carried out to determine if the random-intercept and random-slope model provided a better fit to the data than the random-intercept model. The test statistic of 0.86 results in a p-value of 0.35 from a chi-squared distribution with 1 degree of freedom. Therefore there is no evidence that the random intercept model is improved by adding the random slope component.

Overall NEP Scores, main-effects and interaction, random-intercept model

The main-effects model discussed above assumes that the *Programme* and *Time* effects are the same for both cohorts. This could be an unrealistic assumption as it is possible that, for example, students from the second cohort could have a different rate of change in NEP scores compared with students from the first cohort. These cohort-specific effects can be modelled by including the "interaction" between the main effects to give the following model:

$$Y_{ij} = \beta_0 + \beta_1 Hunt_i + \beta_2 Surv_i + \beta_3 Cohort_i + \beta_4 Time_{ij} + \beta_5 Hunt_i \times Time_{ij} + \beta_6 Surv_i \times Time_{ij} + \ldots$$
$$+ \beta_7 Hunt_i \times Cohort_i + \beta_8 Surv_i \times Cohort_i + \beta_9 Cohort_i \times Time_{ij} + b_{1i} + e_{ij}$$

Table 3 below lists the fixed effect parameter estimates and confidence intervals from the interaction model:

Table 3 Outputs from the main-effects and interaction, random-intercept model

Fixed Effect	**Estimate**	**Standard Error**	**p-value**	**95% Confidence Interval**	
Main Effects					
Intercept	3.97	0.081	<0.001	3.81	4.12
Programme (Zoology is reference)	-	-	-	-	-
Human Nutrition	-0.32	0.11	0.005	-0.54	-0.097
Surveying	-0.65	0.11	<0.001	-0.86	-0.43
Cohort	0.068	0.13	0.61	-0.20	0.33
Time	0.0011	0.00066	0.11	-0.00023	0.0024
Interaction Effects					
Human Nutrition*Time	-0.0015	0.00094	0.11	-0.0034	0.00032
Surveying*Time	-0.00083	0.00090	0.36	-0.0026	0.00095
Human Nutrition*Cohort	-0.11	0.17	0.51	-0.43	0.21
Surveying*Cohort	-0.12	0.16	0.46	-0.43	0.20

Cohort*Time	0.00041	0.00079	0.61	-0.0011	0.0020

As for the previous models we carried out a likelihood ratio test to compare the fit of the `random-intercept' and `random-intercept plus random-slope' versions of the interaction model. The test produced a p-value of 0.77 and as such there is no evidence that the `random-intercept plus random-slope' model fits the data better than the `random-intercept' model.

The results shown in Table 3 are very similar to those obtained for the main-effects only model with *Programme* being the only significant main effect. All of the interaction effects were non-significant with p-values greater than 0.05. This suggests that all of the programmes of study tend to behave similarly over time and within the different cohorts.

We compared the fit achieved by the interaction model with that of the main effects model using a likelihood ratio test. The test statistic of 3.58 results in a p-value of 0.61 from a chi-squared distribution with 5 degrees of freedom. Therefore there is no evidence that the main effects random-intercept model is improved by adding the two-way interactions. Given this, the main-effects random-intercept model is preferred.

Overall NEP Scores, results summary

Using the preferred main effects, random-intercept model, there is evidence of a difference in average NEP scores between the different Programmes of Study (Table 2). With Zoology as the reference category the estimated mean score for Zoology students is equal to the intercept term, 4.04, and the average NEP score of Human Nutrition students was 0.46 lower than the average NEP score of Zoology students. The 95% confidence interval for the difference in mean scores between Zoology and Human Nutrition was (-0.61,-0.30). The mean NEP of surveying students was even lower with an estimated difference of means of -0.75 with a 95% confidence interval of (-0.90,-0.60). In both cases the differences are significant with p values less than 0.001.

There is no evidence that mean NEP scores change with time. Note from Table 2 that the 95% confidence interval for the time parameter includes zero.

Tendencies NEP Scores

In this section we analyse the NEP data in terms of the four tendencies identified by the factor analysis carried out by Shephard et.al (2009). The tendencies and associated NEP items are described in the Table 4 below.

Table 4 Distribution of NEP items to behavioural tendencies

Factor	Items	Tendency Description
1	1,6,11	Recycle
2	3,5,9,10,13,15	Conserve

3	2,7,12	Animal and Plant Rights
4	4,8,14	Cautious about the future

As for the analysis of overall NEP scores, we fitted both the main effects and interaction models and found that the main effects, random intercept model to be the preferred model for all tendencies. Scores for each tendency were taken to be the average of the items associated with each tendency as listed in Table 4.

The estimated model coefficients and associated confidence intervals are given in the Tables 5 to 8 below.

Table 5 Estimated model coefficients for the main effects, random intercept linear mixed effects model fitted to the Recycle tendency scores.

Fixed Effect: Recycle	Estimate	Standard Error	p-value	95% Confidence Interval	
Intercept	3.67	0.091	0.000	3.50	3.85
Programme (Zoology is reference):	-	-	-	-	-
Human Nutrition	-0.74	0.10	0.000	-0.94	-0.53
Surveying	-0.74	0.10	0.000	-0.93	-0.54
Cohort	0.02	0.082	0.81	-0.14	0.18
Time	0.0019	0.00068	0.006	0.00055	0.0032

Table 6 Estimated model coefficients for the main effects, random intercept linear mixed effects model fitted to the Conserve tendency scores.

Fixed Effect: Conserve	Estimate	Standard Error	p-value	95% Confidence Interval	
Intercept	4.27	0.074	0.000	4.12	4.41
Programme (Zoology is reference)	-	-	-	-	-
Human Nutrition	-0.43	0.083	0.000	-0.59	-0.27
Surveying	-0.77	0.081	0.000	-0.93	-0.61
Cohort	-0.0063	0.065	0.92	-0.13	0.12
Time	0.00040	0.00056	0.48	-0.00070	0.0015

Table 7 Estimated model coefficients for the main effects, random intercept linear mixed effects model fitted to the Animal and Plant rights tendency scores.

Fixed Effect: Animal and Plant Rights	Estimate	Standard Error	p-value	95% Confidence Interval	
Intercept	4.26	0.099	0.000	4.07	4.46

Programme (Zoology is reference)	-	-	-	-	-
Human Nutrition	-0.44	0.12	0.000	-0.66	-0.21
Surveying	-0.88	0.11	0.000	-1.11	-0.66
Cohort	-0.053	0.092	0.569	-0.23	0.13
Time	-0.000025	0.00064	0.969	-0.0013	0.0012

Table 8 Estimated model coefficients for the main effects, random intercept linear mixed effects model fitted to the Cautious tendency scores.

Fixed Effect: Cautious	Estimate	Standard Error	p-value	95% Confidence Interval	
Intercept	3.69	0.084	0.000	3.52	3.85
Programme (Zoology is reference)	-	-	-	-	-
Human Nutrition	-0.22	0.094	0.017	-0.41	-0.040
Surveying	-0.59	0.092	0.000	-0.77	-0.41
Cohort	0.062	0.075	0.406	-0.084	0.21
Time	0.00085	0.00065	0.19	-0.00042	0.0021

The result of interest from the models fitted to the four tendencies is that the time effect is highly significant (p=0.006) but only for the Recycle tendency. Although this is a significant change, the change is very small; on average across the population the estimated annual change is +0.099 (52 weeks by 0.0019) NEP units. Put another way, for the benefit of those who might identify normative sustainability objectives and seek to use these approaches to establish appropriate educational strategies; it might take in the order of seven to eight years for our students' existing experiences to change the recycling tendencies of our students by the NEP difference regularly recorded between our student surveyors and our student zoologists.

Discussion

Our discussion must focus on two relatively separate lines of enquiry: the processes used to measure change in the environmental worldview of our students in recent years; and the nature of the changes observed.

On the first; we identify a research instrument, an approach to repeatedly and systematically use this instrument with groups of anonymous students and a statistical model with which change can be identified and interpreted. We recommend further use and development of these combined processes as higher education grapples with the ES/ESD/EE enterprises. We note here that the model can be developed to accommodate additional cohorts of students and additional programmes of study. Any higher education institution interested in asking itself if its students' environmental concern changes as they pass through its programmes could use this instrument, process and model to good effect; subject of course to the limitations and concerns described below.

On the second; this research, conducted over four years, involving multiple NEP surveys of anonymous students in multiple cohorts engaged in several programmes of study, does not make a strong case for an institutional effect on the environmentally-aligned attributes of our students. In no way would we wish this finding to be seen as potentially generalizable to international higher education, but it should be of interest. We note here other researchers have identified correlation between higher education experiences and commitment to environmental sustainability (Cotton and Alcock, 2012). Departments within our institution may be interested in subtle changes in ecological worldview, or its constituent tendencies, but on balance the students followed by this research arrived at our institution with worldviews aligned to their sex and, to an extent, to their chosen discipline, and left some years later relatively unscathed by their experiences with us. Our university has only recently identified 'environmental literacy' as a graduate attribute. This educational aim is probably not yet widely discussed or generally sought within the institution and is likely interpreted variously in different parts of the university. At this stage, perhaps it is not surprising that our research has shown limited change. But it will be difficult for our institution to ignore repeated longitudinal measures in future years. Our own institutional challenge is to generate willingness to continue to monitor the situation.

We, the authors and researchers, anticipate that institutional and departmental responses will mirror those from academic referees and readers of our related papers and our own concerns and interests. Responses (with our comments following each question) may focus on:

- Validity and potential inadequacy of the research instrument. What does this instrument really measure and is it possible to measure something as nebulous as affect? These matters are addressed extensively within the literature (see for example Hawcroft & Milfont, 2010). No research instrument used in this complex area is immune from criticism, but the NEP is strongly endorsed, and validated, as an appropriate measure of environmental attitude (Hawcroft & Milfont, 2010), environmental concern (Dunlap & Jones, 2002) or ecological worldview (Dunlap, 2008). Nevertheless, the processes and statistical models described could be adapted for use with many alternative research instruments. We stress here that this research uses the NEP, but is not about the NEP. Aversion to the NEP should not be a sufficient cause to discount the processes described here. We recommend, for example, the tools and framework developed by the North American Association for Environmental Education to measure the environmental literacy of learners at many educational levels (Hollweg et al., 2011; Shephard et al, in press). The two-dimensional 2-MEV (Bogner & Wiseman, 2006) may also be appropriate. Even so, it is possible that some educators will not wish to draw evidence from any quantitative research instrument, no matter how valid statistical analysis suggests it to be. We note in particular that there is a developing philosophical discourse in the general area of environmental, and sustainability, ethics (Kronlid & Öhman, 2013). No doubt particular instruments are developed and used within particular ethical frameworks, either deliberately or inadvertently. The objectives that higher education sets itself, and the evaluative or assessment processes that it uses to keep track of progress, all need to be interpreted with respect to the assumptions that underpin their conception. This will be a challenging task for higher education; requiring at least multidisciplinary engagement by those responsible for developing policies, strategies and implementation programmes.

- Is it worthwhile monitoring the environmental attitude of students when we know that there is a substantial attitude/behaviour gap? Although this gap exists and can be explained (Hargreaves, 2012), measurements of behaviour change are far more complex to achieve (Monroe et al., 2013) and may be beyond the expertise of researchers in HE. Besides, although individuals who have high levels of pro-environmental attitude may not behave in a sustainability-focussed manner, it seems less likely that those with low levels of pro-environmental attitude will behave in this way. High levels of pro-environmental attitude, ecological worldview or environmental concern may not be a sufficient predictor for pro-environmental behaviour, but may be a pre-requisite. The possibility of promoting pro-environmental behaviour without, necessarily, changing associated attitudes is an active area of research in the social sciences.

- Impossibility of assessing, monitoring or evaluating development in the affective domain. Especially, did these students really believe that they were anonymous, or were they recording what they thought their tutors wanted to hear? Researchers made every attempt to stress that students were anonymous, and the code system used would be challenging for an institution to break. We have confidence that the responses made by students are not affected by institutional expectations, but suggest here that this would be a useful area for further research. How free from institutional, parental, peer and societal pressures are students to decide for themselves how they might respond within an anonymous survey?

- Is this the business of higher education? For example, what gives educational institutions the right to evaluate their students' environmental worldview, or, for example, to ask their students to agree or disagree with statements such as "humans have the right to modify the natural environment to suit their needs"? Even to ask may be seen as impertinent and disrespectful of the rights of students as citizens. At the University of Otago, this research was subject to ethical approval, emphasising the voluntary nature of the survey. The authors of this paper may agree that informed consent is a necessary part of the research and evaluation processes discussed here. Student responses in our institution have been overwhelmingly positive with response rates in classes routinely in the high 90s%. But taking the processes beyond research may involve additional ethical concerns. At least some higher education institutions have perhaps assumed this right, as they commit themselves to educate for sustainability or for sustainable development, and students know of this commitment before they come to university. It seems unlikely that this commitment would not carry with it an expectation of evaluation. At least one country (Sweden) has enshrined the principles of education for sustainable development into its higher education governance. The Swedish Higher Education Act was amended to specify that its higher education activities will promote socially, economically and environmentally sustainable development (SOU 2004:104). Nevertheless, not all in higher education will agree that institutions have the right to ask questions potentially unrelated to the educational programmes that students register for. It seems likely to us that some higher education institutions have made claims about their intentions for their students' values and attitudes without fully exploring the educational, moral and individual-freedom issues that may be involved. The University of Otago has committed itself to fostering the environmental literacy of its undergraduate students; no more, or less, at this stage.

- Problems identifying and separating effects from within higher education itself and from the experiences that students have in their years of HE that are not controlled by HE. Is this research relevant to higher education? The world 'outside' higher education changed considerably for these students during the period of this research, particularly as portrayed by the media. During this period, students would have been exposed to media reports about (and possibly had direct personal experience of) extreme weather, economic duress, rising graduate unemployment, whaling and overfishing in the southern oceans, oil drilling by hydraulic fracturing (fracking), land use change, polluted waterways, natural disasters such as tsunamis and the earthquakes in nearby Christchurch. It remains possible that all the good work undertaken by institutions is undone by life's challenges outside the institution, or vice versa, but this seems an inadequate rationale for not researching change. Many higher education institutions worldwide expect their students to become environmentally-literate, or sustainability-focused, irrespective of the external social environment that their students inhabit.

- Complexity of this process. Do we really need statisticians to address this issue? Without statisticians, and statistics, educators may be doomed to subsist in a qualitative paradigm of 'what might be'. Statisticians add complexity, but also the promise of repeatability, reliability and transparency. Anonymity adds validity to the data. But as with all research data, interpretation is key.

- What of the workload for universities? Higher education has already developed an industrial-scale undertaking to manage anonymous feedback from students to satisfy itself and its stakeholders of the quality of its educational services. A small fraction of this industry, diverted to evaluate HE's effect on the sustainability attributes of its students, may be a small price to pay for confidence that HE's assertions about its effectiveness are well grounded.

- Are the students self-selecting? All participants were volunteers but nearly all students in classes where the research instrument was used did volunteer (with participation rates in the high 90%s). Perhaps more importantly, participation depended on attendance at classes and this was highly variable. This research therefore tended to select those students who attended classes. There is no data on links between ecological worldview and propensity to attend classes in higher education. An online survey would have different selection properties and different problems.

- Are the NEP values of these students so high already that it would be unreasonable to expect positive change? Is the instrument, and statistical model, sufficiently discriminating to identify change? Students arrive at this institution with different degrees of ecological worldview, significantly aligned with their chosen programme of study. Presumably prior experience at home, at school, or in the wider community has influenced these students (no genetic link with sustainability attributes, possibly other than sex, has been identified yet) and the research instrument is sufficiently discriminatory to detect these differences. Minor changes in the already high scores of Zoology students were detected. There is no reason to doubt the efficacy of this research instrument, or statistical model, in detecting change.

We suggest that higher education institutions can, and perhaps should, monitor their impact on the sustainability attributes of their students; particularly if they claim or intend to have an effect in these areas. We describe here processes that provide a basis for this monitoring and invite higher education practitioners worldwide to join us in further developing the research instrument, the processes and the statistical model; and in further analysing the assumptions that link higher education to sustainability, and to global citizenship.

References

Anderson, M., Criner, G., Teisl, M., Smith, H., Hunter, M., Norton, S., Bicknell, E. (2007). Attitude Change of Undergraduate Students in General Education Courses. *Journal of General Education*, 56(2), 149–168.

Association of University Leaders for a Sustainable Future. (1994). The TalloiresDeclaration 10 Point Action Plan, updated version. Retrieved August 6 2013 from http://www.ulsf.org/pdf/TD.pdf

Bogner, F. X. & Wiseman, M. (2006). Adolescents' attitudes towards nature and environment: Quantifying the 2-MEV model. *The Environmentalist*, 26, 247-254.

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

Cotton, D. R. E., Warren, M. F., Maiboroda, O., & Bailey, I. (2007). Sustainable development, higher education and pedagogy: a study of lecturers' beliefs and attitudes. *Environmental Education Research*, 13, 579–597. doi:10.1080/13504620701659061

Cotton, D.R.E and Alcock, I. (2012) Commitment to environmental sustainability in the UK student population. *Studies in Higher Education* (In press and available online at: http://www.tandfonline.com/doi/abs/10.1080/03075079.2011.627423)

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, Riley E., Van Liere, K. D., Mertig, A. G., and Jones, R. E. (2000). New Trends in Measuring Environmental Attitudes: Measuring Endorsement of the New Ecological Paradigm: A Revised NEP Scale. *Journal of Social Issues*, *56*(3), 425-442. doi:10.1111/0022-4537.00176

Harraway J, Broughton-Ansin F, Deaker L, Jowett T and Shephard K (2012) Exploring the use of the Revised New Ecological Paradigm Scale (NEP) to monitor the development of students' ecological worldviews. *Journal of Environmental Education* 43(3), 177–191.

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

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 August 6 2013 from http://www.naaee.net/framework

Jensen, B. B., & Schnack, K. (2006). The action competence approach in environmental education. *Environmental Education Research*, 12, 471–486. doi:10.1080/13504620600943053

Jowett T, Harraway J, Lovelock B, Skeaff S, Slooten L, Strack M and Shephard K (2013) Multinomial-regression modeling of the environmental attitudes of higher education students based on the Revised New Ecological Paradigm Scale *Journal of Environmental Education* 45:1, 1-15.

Knight, P, & Page, A. (2007). The assessment of “wicked” competencies. Open University, UK. Retrieved August 6 2013 from http://www.open.ac.uk/cetl-workspace/cetlcontent/documents/460d21bd645f8.pdf

Krathwohl, D., Bloom, B., & Masia, B. (1988). *Taxonomy of educational objectives, Handbook II: The affective domain*. New York: David McKay Co.

Kronlid, D. O., & Öhman, J. (2013). An environmental ethical conceptual framework for research on sustainability and environmental education. *Environmental Education Research*, 19(1), 21–44. doi:10.1080/13504622.2012.687043

Kronlid, D. and Österbergh, R. (2011) Crossing Boundaries: An Analytical Look at Cemus Educational Model, In M Hald (ed) *Transcending Boundaries* Uppsala, Sweden: Centrum för miljö- och utvecklingsstudier (CEMUS), Uppsala centrum för hållbar utveckling, Uppsala universitet & SLU.

Monroe, M. C., Agrawal, S., Jakes, P. J., Kruger, L. E., Nelson, K. C., & Sturtevant, V. (2013). Identifying Indicators of Behavior Change: Insights From Wildfire Education Programs. *Journal of Environmental Education*, 44(3), 180–194. doi:10.1080/00958964.2012.746277

Neyman, Jerzy, and Egon S. Pearson. "On the problem of the most efficient tests of statistical hypotheses." *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character* 231 (1933): 289-337.

Packer, A. (2009). Service Learning in a Non-majors Biology Course Promotes Changes in Students’ Attitudes and Values About the Environment. *International Journal for the Scholarship of Teaching and Learning*, *3*(1), no page.

Shephard, K. (2008). Higher education for sustainability: seeking affective learning outcomes. *International Journal of Sustainability in Higher Education*, 9(1), 87–98. doi:10.1108/14676370810842201

Shephard, K., Mann, S., Smith, N., and 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, K (2010) Higher education's role in 'education for sustainability*' Australian Universities' Review* 52(1) 13-22.

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

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*. Frankfurt: Peter Lang Scientific Publishers. 117-127.

Shephard K, Harraway J, Lovelock B, Skeaff S, Slooten L, Strack M, Furnari M, Jowett T (2013) Is the environmental literacy of university students measurable? *Environmental Education Research* (in press) available at: http://www.tandfonline.com/doi/full/10.1080/13504622.2013.816268

SOU (2004) *Betänkande av Kommittén för utbildning för hållbar utveckling* Stockholm 2004 SOU 2004:104 http://www.regeringen.se/content/1/c6/03/41/44/0fe2bc94.pdf

Sterling, S., & Scott, W. (2008). Higher education and ESD in England: a critical commentary on recent initiatives. *Environmental Education Research*, *14*(4), 386–398. doi:10.1080/13504620802344001

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

The Authors

Kerry Shephard is Professor of Higher Education Development in the Higher Education Development Centre (PhD). His research interests focus on understanding the nature and roles of higher education.

John Harraway is Senior Lecturer in the Department of Mathematics and Statistics. John is currently President of the International Association for Statistical Education and has research interests in applied statistics and statistics education.

Brent Lovelock is Associate Professor in the Department of Tourism and Co-Director of the University of Otago's Centre for Recreation Research (PhD). Brent researches the application of sustainable tourism within protected natural areas.

Sheila Skeaff is Senior Lecturer in the Department of Human Nutrition (PhD). Sheila researches the iodine status of at-risk groups, childhood eating patterns and issues affecting women's health.

Liz Slooten is Associate Professor in the Department of Zoology (PhD). Liz researches the assessment of risk and uncertainty using resource management models, with particular interests in marine mammals.

Mick Strack is a lecturer in the (NZ) National School of Surveying. (PhD). Mick's surveying interests are in land law, Māori land and aboriginal title, environmental ethics, urban design and planning, and our relationship with land, property and resources.

Tim Jowett and Mary Furnari are research assistants to the group and are currently completing master's theses in statistics and higher education respectively.

…

Corresponding Author email: kerry.shephard@otago.ac.nz

Word count; 7420 including references