Development of an Integral-Motivated Lectical® Ecological Stewardship Assessment

Howard Drossman, [1,2] Shanti Om Gaia[3] and Theo Dawson[3]
Colorado College,[1] Catamount Center[2] and Lectica®, Inc.[3]

Abstract

The Catamount Center and Colorado College, in collaboration with Lectica, present the results of a pilot study conducted as the first step in developing the Lectical Ecological Stewardship Assessment (LESA), a new set of assessments of ecological stewardship which rely on responses to a series of open-ended question relating to an environmental dilemma. Fifth-grade student performances on the LESA were scored for Lectical level and coded for content. Comparing students' Lectical scores on the established Reflective Judgment Assessment (RFJ) and students' Lectical scores and qualitative performances on the pilot LESA indicated small gains in Lectical scores over time, small to no statistically significant effect of gender, statistically significant differences in the frequency of responses from higher and lower level scorers on some themes, and a correlation between the frequency of responses to different themes based on students' RFJ and LESA scores. These initial results suggest that we can use the LESA to learn more about the development of students' reasoning related to ecological stewardship by supporting the optimal development of students' understanding of ecological stewardship and, ultimately, to support the development of citizens who are effective environmental stewards.

Introduction

Environmental education researchers largely agree that our ability to make effective environmental decisions about complex problems requires environmental literacy (e.g. Coyle 2005), which includes good reasoning skills (Berkowitz, Ford, & Brewer 2005; Hollweg *et al*. 2011). The environmental education community has reached a general consensus that ecological and socio-political knowledge, skills for identifying and acting upon environmental issues, apt dispositions (e.g. environmental awareness and attitudes), and pro-environmental actions are all important factors contributing to environmental literacy (Simmons 1995; Simmons 2005, McBride *et al*. 2013). However, the tools we have available to enhance these literacies and inspire ecological stewardship could be greatly improved by learning how ecological conceptions[1] develop over time (Esbjörn-Hargens & Zimmerman 2009, 215-242), and how children's and adults' levels of understanding affect their ability to make environmental decisions.

Catamount Center, whose mission is to inspire ecological stewardship, in collaboration with the Colorado College Education Department and Environmental Program, has begun to address the challenge of assessing reasoning skills relating to environmental literacy through its innovative Teaching and Research in Environmental Education (TREE) Semester. Here, undergraduate students teach environmental education to local K-12 students and research how their students learn in outdoor laboratory school and indoor classroom settings (Catamount Center 2015). The Catamount Center's integral approach to Environmental and Sustainability Education[2] (ESE) is summarized in its "Seven S" curriculum model (Fig. 1). Starting with Sustainability as the ultimate goal of ESE, the model emphasizes the essential the role of human actions towards the environment (Service) for achieving sustainability. We define sustainability as the planetary conditions that allow today's human and non-human life to thrive without compromising the ability of future generations to thrive. The model indicates the central role of developing ecological stewardship from an all-quadrant (Science, Systems, Society and Spirit) perspective. We define ecological stewardship as the skill of working towards the highest good regarding the impact one's decisions and actions have upon all life and the planet across time, while considering the relative importance of significant factors and impacts.

1. We use an integral definition of "ecological conception" as provided by Esbjörn-Hargens and Zimmerman (2009), who describe more than 200 different types of ecology. This more holistic conception is similar to McBride's (2013) use of "ecoliteracy."
2. Though environmental education and education for sustainability are often separated due to different goals and approaches (e.g. Eilam & Trop 2010. Similarly to Garrison *et al*. (2014), we refer to the combined approaches of Education for Sustainability (EFS) and Environmental Education (EE) as environmental and sustainability education (ESE).

TREE Semester™ faculty members, students, and fellows from Catamount Center and Colorado College have begun working with Lectica® to develop the Lectical Ecological Stewardship Assessments (LESA), a new set of assessments focused on how children and adults learn to reason about ecological stewardship. Each form of the LESA will include an ecological dilemma and a set of open-ended questions requiring written responses. Over the next few years, each form of the LESA will be administered to a diverse range of children and adults, whose responses will be analyzed to determine how reasoning about ecological stewardship develops over the lifespan. The end result will be a fully developed set of Lectical Assessments—richly diagnostic learning tools that are designed to support learning. They do this by engaging teachers and their students in virtuous cycles of learning (VCoLs)—repeated cycles of goal setting, information gathering, application, and evaluation that optimize learning while enhancing the learning experience (Stein *et al*. 2014). Below, we describe some of the first steps in developing the LESA and report on some initial findings regarding students' understanding of ecological stewardship and the relation between the development of reasoning about ecological stewardship and reflective judgment skills.

Our Approach is Integral

We have chosen to collaborate with Lectica to develop the LESA because their approach to educational testing is fundamentally integral, coordinating concerns about the experience of the individual test taker, the objectivity of tests and their results, the broader culture of test taking, and the functional fit of tests to collective needs (Figure 2). Our research questions relate to lines, levels and types:

1. Lines: How closely related are ecological stewardship and reflective judgment skills?
2. Levels: How does students' reasoning about ecological stewardship change as a function of development?

3. Types: Are there differences between genders in reasoning about ecological stewardship?

Literature Review

Development

We view development, in accordance with the constructive developmental tradition, as "the evolution of skills over time, where early level skills are reorganized into higher-level skills that allow individuals to manage more complex units of information, perspectives, and tasks" (King 2009, 598). Constructivist research has focused on how knowledge and skills are learned over time, but while constructivism has been prominent in science education for almost three decades, there has been relatively little research on constructivist environmental and sustainability education approaches (Robottom 2004). Constructivism, which emerged from the seminal work of Baldwin (1906) and Jean Piaget (1972), posits that knowledge is constructed by the learner in response to interaction with the environment (including the social environment) and is not *directly* transferable from teacher to learner. Social constructivism, often attributed to Lev Vygostsky, emphasizes the role of social interaction, in particular how social engagement in the form of articulating, reflecting, and negotiating is essential for constructing knowledge. Vygotsky (1978) also identified the "zone of proximal development"—the difference between what a learner can do without support and what he or she can do with support. The upper end of the "zone" is the optimal target of instruction—the learning "sweet spot." One reason for studying how students' understanding of ecological stewardship develops over time is that this knowledge will make it possible to optimize learning by using scores on the LESA to determine what an individual is most likely to benefit from learning next.

The reflective judgment "line"

During the last 100 years, developmental sequences have been described for a number of knowledge areas, including reflective judgment, moral judgment, evaluative reasoning about the good, stages of faith, mathematical reasoning, self-understanding, and science reasoning. In the literature, the term "domains" is often used to differentiate between these knowledge areas, which are analogous to Wilber's (2002) "lines". The reflective judgment construct, which focuses on the way people think about knowledge, evidence, inquiry, truth, and certainty, has been the focus of research since Perry (1970) first interviewed college students about their epistemologies. Building upon Perry's work, and strongly influenced by Piaget, Kitchener and King (1994) enriched our understanding of epistemological development with their Reflective Judgment Interview and scoring system through a series of adolescent and adult studies over two decades. Importantly, they also collaborated with Fischer (Kitchener & Fisher 1990) to examine the relation between skill levels and levels of reflective judgment—one of the first attempts to formally align a domain-specific assessment with a general developmental model. Lectica's Reflective Judgment assessments—the LRJA and RFJ—build upon this initial alignment of skill levels with Kitchener and King's levels.

The ecological reasoning literature

There is a small body of developmental research related to ecological reasoning, including socioscientific reasoning and systems reasoning. In one study of high school students' reasoning about socioscientific issues, Zeidler *et al*. (2009) showed, using a dilemma and questions similar to ours (described in Methods), that providing students with opportunities to engage in structured discussions about socioscientific issues, that often include environmental issues, can promote reflective judgment. Sadler, Klosterman and Topcu (2011) used open-ended questions relating to climate change to assess skills associated with socioscientific reasoning:

complexity, inquiry, perspectives and skepticism. They found a statistically significant correlation between complexity and inquiry skills for high school students enrolled in a socioscientific issues-based class, but no statistically significant correlations of perspectives or skepticism with the other skills.

Several researchers have examined the development of "systems reasoning.[3]" In this literature, the term *systems reasoning* and other terms that include the word *system* are frequently used in connection with students' reasoning, even when students do not have an understanding of systems *per se*. For example, Hmelo-Silver and Pfeffer (2004) use a Structure–Function-Behavior model for testing understanding of complex systems. In their model, *structure* represents system elements, *function* denotes system purpose and *behaviors* are mechanisms allowing the structures to carry out the intended function. For instance, a child may argue that "The animals and plants will die if the planet gets too hot and we won't have any food to eat so we will die. I think we should try to stop global warming.[4]" The system elements in this argument would be the lives of animals and plants, and the lack of food for people, the system purpose would be to describe what will happen if the planet gets too hot, and the mechanism for carrying out the intended function would be to try to stop the planet form getting too hot. Hmelo-Silver and Pfeffer's results suggest that younger students most often address the structure and visible (more concrete) functions of systems. Sheehy *et al*. (2000) suggest that children's failure to go beyond these more concrete conceptions may be explained by developmental limitations (Piaget, 1970).

---

[3] Not to be confused with the way the term "systems" is used in our developmental model.

[4] Here, the effects of global warming, an abstract concept, are described in terms of two sequentially related concrete outcomes, an argument structure that is characteristic of reasoning in Lectica's level 9, which is equivalent to Fischer's (1980) single abstractions level.

Wylie *et al*. (1989) interviewed 8- and 11-year old students to assess their level of systems thinking about air pollution and found that even the younger students could express their answers in terms of *simple cycles*[5], which they define as explanations that include inputs, processes and outputs. As an example from one student interview response from their study: "When you put petrol in the cars and all the smoke comes out of the cars it pollutes the air" (123). The input is petrol, the process is smoke coming out of the car, and the output is pollution. According to Wylie *et al*., moderately complex explanations provided interventions for cycles and the most complex answers suggested outcomes for interventions. They also observed few students who could link two different cycles.

In a study examining how 7th grade students used systems reasoning related to climate issues, Shepardson *et al*. (2014) asked students how the components of the climate system influence climate, how an increase in greenhouse gases influences climate based on the climate system, and how global warming influences climate based on the climate system. The researchers observed that the 7th grade students reported simple structures almost entirely and showed little linking of systems components in their explanations.

Using a similar approach, Myers, Saunders and Garrett (2004) interviewed different-aged students about their knowledge of animals to determine if knowledge about animals translated to ecosystem and conservation awareness. They found that younger students were more likely to express concern for an individual animal, while older students had some understanding of animals' dependence on aspects of ecological systems. In the existing developmental research on ecological thinking, student age or grade are used as proxies for development, so it is unclear how findings relate to the developmental levels of any developmental framework.

---

[5] The student in this describes how sequential concrete events relate to an abstract outcome, an argument structure that is characteristic of reasoning in Lectica's level 9, which is equivalent to Fischer's (1980) single abstractions level.

Lectical Assessments

Lectical Assessments are calibrated to a single developmental scale with its origins in the work of Baldwin (1906) and Piaget (1972). The Lectical Scale is a refinement of Fischer's (1980) skill scale, which itself is an elaboration of Piaget's stages of development. The skill scale includes 13 levels (0 – 12) of increasing hierarchical complexity, covering the period from birth through adulthood. Over several years Dawson (Dawson 2002; Dawson-Tunik 2004; Dawson, Xie & Wilson 2003; Dawson & Stein 2011) has refined a general developmental scoring system—the Lectical Assessment System—that makes it possible to reliably detect four phases in each of Fischer's skill levels (from levels 6 through 12). During this same period she developed a method—*developmental maieutics*—for using the Lectical Assessment System and a process called *rational reconstruction* to develop descriptions of detailed learning sequences (e.g. Dawson-Tunik 2004; Dawson 2006). Lectica uses these methods to build diagnostic, formative, standardized assessments in a wide range of knowledge domains including ethical reasoning, self-understanding, reflective judgment, science reasoning, and decision-making (Lectica 2015), all of which are designed to support optimal learning.

In the analysis that follows, we examine the relation between reflective judgment and ecological stewardship, and take a first look at how conceptions related to ecological stewardship might change with development.

Methods

Sample

A total of 54 fifth-grade students (26 girls and 28 boys) from a US rural public school participated in this study as part of their regular classwork. Fewer than 10% of the students were from minorities. We did not have access to SES data for the individual students in our sample, but in the school as a whole, 29% of students receive free and reduced lunch. Due to student absences and responses that were unscoreable, there were fewer than 54 total scored assessments during the three assessment periods (Table 1).

Instruments

The Reflective Judgment Assessment (RFJ)

Like many Lectical Assessments, the reflective judgment assessment (RFJ) features an ill-structured dilemma—a real-life dilemma with no correct answer. The RFJs employed in this study (see Appendix 1 for example) provided dilemmas relating to the effects of watching violent television (RFJ001/RFJ002) and the consequences of bullying (RFJ003). Through their responses to a series of questions, students reveal how they think about inquiry, evidence, learning and the mind, truth and certainty, conflict resolution, persuasion, and deliberation.

The Lectical Ecological Stewardship Assessment (LESA)

The 5th grade pilot LESA, which relates to global climate change, is similar to the RFJ in approach, but was designed to assess students' ecological reasoning. We used the following dilemma and associated questions:

> Many people think that air pollution is making our planet warmer. They call this global warming. They say that global warming will cause many problems, and they think people should stop polluting the air. Other people say that warming the planet may not be bad for everyone, so we should not do anything. They also say that it would cost too much money to stop polluting the air.

1. Do you think that air pollution is making the planet warmer? Why or why not?
2. Do you think it will be bad for people if the planet gets warmer? Why or why not?
3. Do you think it will be bad for the planet if it gets warmer? Why or why not?
4. What would be the best way to learn more about global warming? Why?
5. It will cost a lot of money to stop polluting the air, and some people will have to give up things that are important to them. For example, they may have to live in smaller houses, use less electricity, and drive their cars less. Would it be worth it to give up these things in order to stop global warming? Why or why not?

Procedures

The 5th grade pilot LESA was taken online in February 2015 during students' regular one-hour writing class. There were no prior school-based lessons specifically relating to climate change; however, all students had been enrolled in the Catamount Center TREE Semester from August – December 2014, where they participated weekly in experiential outdoor ecology lessons. All students were assigned the same LESA dilemma and questions. The same group of students had twice previously completed the RFJ during their writing class, once in September 2014 (RFJ-T1) and a second time in January 2015 (RFJ-T2). Students were assigned randomly to one of three different RFJ dilemmas (RFJ001, RFJ002, RFJ003) during each assessment time, never receiving the same dilemma twice.

Qualitative Analysis

In preparation for a comparison of students' use of concepts at different developmental levels, LESA student response data were coded in N-Vivo for Mac software (v. 10.2.0). The first stage of coding followed Saldaña's (2009, 66-70) recommendations for "structural coding," in which we used a word or short phrase to represent the content of a segment of student response data. Each segment chosen from a student's response was assigned to one or more codes (termed nodes in NVivo), depending on the number of distinct ideas contained in the segment. In the second stage of coding, nodes with related content were combined into common themes, which are words or phrases that identify a subject (Saldaña 2009, 139). All sub-codes were aggregated at the theme level and recorded at each node.

Quantitative Analysis

The 5th grade LESA and RFJ responses were all scored for Lectical Level using Lectica's Computerized Lectical™ Assessment System (CLAS). Statistical tests exploring potential patterns of development (based on the Lectical scores) relative to assessment time, gender, and instrument version were conducted in SPSS for Mac (v. 19). Cohen's *d* (1998) was used to test for effect sizes. All parametric data were tested for normality by visual inspection of P-P plots, and homogeneity of variance with Levene's test. Based on these criteria, normalization was not required for any data.

NVivo attributes (Bazely & Jackson 2013) were created for each student to represent gender, RFJ-T2 score, and LESA score. Since our sample was small, it was statistically convenient to sort observed scores into only two groups to gain sufficient statistical power. The lower-level score range included all student performances scored within 9.31-9.69 and the higher-level range included all scores greater than 9.69. An NVivo matrix query (Bazely & Jackson 2013) was used to determine the frequency of responses at every coded node as a function of scores on both the RFJ-T2 and the LESA. The frequency data were exported from NVivo into Excel for comparative analysis of the average student response rate by node. The average student response rate per node was calculated as the ratio of the frequency of responses (frequency ratio) at a given node to the total number of students in each Lectical range. Though 81 codes and six themes were generated from the students' LESA responses, we statistically analyzed only those codes for which the average student response rate for a node was at least 0.5 and the total number of responses at each code for both CLAS ranges was greater than 10 for both assessments. All aggregated themes and sub-themes were included for statistical analysis (Table 2).

Results

RFJ Performance Over Time

Descriptive statistics for the RFJ-T1, RFJ-T2 and LESA are provided in Table 1. Though the mean value for the RFJ performance increased from September 2014 to January 2015, there was no statistically significant difference in students' Lectical scores ($t(38) = 1.09$, $p = .14$, d =

0.21)) using a paired t-test to compare each over this four-month interval. However, scores for 82% of students either increased (42%) or remained the same (40%) over the four-month period. If changes in observed average developmental growth followed a linear trend over time, the increase would be comparable to a typically observed annual gain of ~0.10 - 0.13 of a Lectical level (Dawson & Seneviratna 2015).

Gender Effects

Independent-samples t-tests were used to assess differences in Lectical level by gender for both the RFJ and LESA (Table 1). We found no statistically significant difference between genders for both RFJ-T1, $t(47) = .628$, $p = .53$, $d = .18$, and LESA, $t(35) = .23$, $p = .82$, $d = .077$. However, for the RFJ-T2 we found a statistically significant difference by gender, $t(39) = 2.13$, $p = .039$, $d = .68$. Given these mixed results, we used the results of a matrix analysis to compare frequency ratios by gender for the most common LESA codes and themes. A paired t-test indicated no statistically significant difference in frequency ratios ($t(17) = .327$, $p = 0.75$, $d = .015$) between genders, and a strong correlation of the frequency ratios between genders ($r(16) = 0.99$, $p < 0.001$). Overall, our results are consistent with prior observations of low effect sizes when comparing gender effects relative to environmental concern (Koger & Winter 2010, 118-119), and gender effects relative to educational interventions (Hattie 2012, 253).

LESA vs. RFJ Instrument Comparison

Using paired t-tests for Lectical level, we found the RFJ-T1 scores were statistically significantly lower ($t(34) = 3.65$, $p = .001$, $d = .83$) than the LESA scores, and that the RFJ-T1 and LESA scores were not statistically significantly correlated, $r(33) = .095$, $p = .59$. The RFJ-T2 Lectical scores were also statistically significantly lower ($t(28) = 2.21$, $p = .036$, $d = .42$) than the paired LESA scores. However, the RFJ-T2 scores were statistically significantly correlated with the LESA, $r(27) = .485$, $p = 0.008$. Overall, it was not possible to separate effects such as student acclimation to Lectical Assessments and student development over time from observed differences between the RFJ and LESA scores. Given the closer timing of LESA (February 2015) with the RFJ-T2 (January 2015) than with the RFJ-T1 (September 2014), and that students' scores increased from RFJ-T2 to RFJ-T1, RFJ-T2 scores were considered to provide a more appropriate comparison with LESA. At this time, we are unable to explain the differences between the LESA and RFJ-T2 scores.

The statistically significant correlation between the RFJ-T2 and LESA motivated further inquiry into the distribution of LESA codes in performances that received higher and lower Lectical scores on the LESA and RFJ-T2. To provide insight, we used the results from our matrix query to determine the average frequency of responses for each threshold code and each of the RFJ-T2 and LESA Lectical score groupings. Table 2 shows which topics were addressed more or less frequently at the different Lectical levels.

Thematic results

Frequency ratios for the RFJ-T2 and LESA were similar for three of the six themes: *Sources of information*, *Effect of changing climate* (including sub-themes of negative and positive effects), and *Consequences of climate change,* though more students who received lower-level scores on this last theme suggested not taking action in their responses to the LESA while more students who received higher-level scores suggested taking action. The frequency ratio of LESA respondents who provided information related to the theme, *Causes of climate change* was about three times as large as the frequency ratio of the RFJ-T2 respondents, even though the frequency ratios for the most frequently coded cause of climate change (hot gases) was similar. There was also a difference in frequency ratio between the themes *Pollution* and *Actions*, but unlike *Causes of climate change*, the frequency of responses was greater for the lower-level Lectical scores than for the higher-level scores on both assessments.

Discussion

Lectical level distribution

Most developmental studies use grade level as a proxy for developmental level. Although this approach yields interesting observations about how students' thinking changes over time, it tends to obfuscate within-age differences in developmental progress. For example, based on average annual student growth rates of 0.1 - 0.15 of a Lectical level per year, levels of performance in our 5th grade sample, which ranged from 9.31 to 9.94 on the LESA, and 9.31 to 10.06 on the RFJ, spanned 5-6 years. This wide range of scores within a single grade is typical (Dawson & Stein 2008; Dawson & Stein 2012; Dawson 2014), and has important implications for teaching and learning as well as research into how conceptions develop over time. For example, it brings into question the wisdom of attempting to teach exactly the same material to students representing such a wide developmental range or describing learning sequences based primarily on differences between students in different grades (Dawson & Stein 2011a).

Ecological reasoning

The code that appeared with overall highest frequency relating to causes of global warming (Q1) was that hot gases are the cause of global warming (Table 2); almost half the students who answered Q1 implied that a cause of global warming is caused by hot gases added to the air by humans. Here is one typical student response (italics added):

> I do think the air pollution is making the world warmer because I have seen nonfiction books saying global warming is here and all about it. I also think that the amount of people who believe in the pollution is so large it must be making the air warmer. *Air pollution is a warm gas that is in the air making it warmer in my opinion*.

This student, like many others, attributes warming to gases, though he/she does not define the gas as a greenhouse gas, as has been observed in 7th grade students (Shepardson *et al*. 2014) who have been prompted with the term *greenhouse gasses* in their interview questions.

While there were a relatively small number of responses for the code relating to *interdependent life* (Table 2), a few responses indicated a beginning understanding that the existence of plants, animals and people are related. For example, one student wrote, "…we wouldn't have plants and without plants animals would die and without food to hunt we wouldn't have food to eat and we would die." Note that this is a simple string of concrete relations between plants, animals, and people—an early step in the process of developing an understanding of ecological systems.

A beginning understanding of concrete relations between ice, landmass, and water was demonstrated in some students' characterizations of *polar ice melting*. For example, one student wrote:

> Yes it will be bad for people if the planet gets warmer because there is about 70% of water and 20% land. The glaciers will melt and there will be100% water and 0% land and that is bad because sharks got to eat and people got to swim.

These statements provide examples of what Wylie *et al*. (1998) identify as simple cycles including input (plants/food; glaciers), process (eating; melting) and outputs (dying; eating). The answers are also consistent with Hmelo-Silver and Pfeffer's (2004) observation that predict an emphasis on structure (plants/animals; glaciers) and visible functions (eating; warming), but limited explanation of behaviors.

Not only do more students performing at higher levels mention animal health effects, but their responses take into account more factors. For example one student whose LESA score was 9.94 wrote that pollution/global warming, "will kill our life source for food like animals, plants, us, it would probably kill us too." In contrast, a student performing at 9.56 stated that pollution/global warming "will be bad for penguins and all the other animals who live in the northern area because they are used to cold and they would possibly die." The difference between these two answers is consistent with Myers' *et al*. (2004) observation that developmentally, students move from concern for particular animals to an understanding of animals in ecological systems, although as in earlier examples, the concepts being coordinated in the 9.54 response are concrete and related to one another in a simple list. There are none of the abstract concepts, complex arguments, and recursions that characterize true systems.

Given the small sample and the preliminary nature of this research, we are, of course, reluctant to draw any strong conclusions from these qualitative patterns in our data.

LESA and "lines"

Our preliminary LESA results, combined with those from the RFJ, allowed us to explore how reflective judgment skills overlap with reasoning about ecological stewardship. We found that these assessments appear to be measuring related lines. This is not surprising for a couple of reasons. First, we used RFJ-type questions as a model for some, but not all of the LESA questions. For example Q4, relating to *Sources of information,* was very similar to a question asked on the RFJ. The main sources of information used by students in the 5th grade were similar, on both assessments. But we also expected at least some shared variance because reflective judgment skills will inevitably be exercised in deliberation about any contested issue. Our understanding of the nature of knowledge, inquiry, and truth is always a factor in this kind of deliberation.

It has not been our goal here to determine whether or not reflective judgment and ecological stewardship constitute separate or overlapping lines of development. But perhaps, by developing the LESA, we will contribute new evidence that will help future scholars better understand the relation between these dimensions. To paraphrase *Integral Ecology*: we need more research before we can discern whether we are discovering new lines of development, or whether we will simply find how existing lines express themselves in relationship to the environment (Esbjörn-Hargens & Zimmerman 2009, 224).

Limitations

Given the preliminary nature of this study, we faced two significant limitations. First, the small data set restricted our ability to compare codes solely from students who only took the RFJ-T2 *and* LESA. Second, we created our coding categories based on content-related themes, using a process that was unlike the process used by Lectica to develop learning sequences, which focuses on how students understand concepts rather than which concepts they mention. As such, we could only address whether children talked more or less *about* topics based on frequency responses rather than *how* they thought about these topics. Results obtained with Lectica's approach will make it possible to document precisely how students' understanding of environmental stewardship develops over time. Future work developing the LESA will target larger and more diverse groups of children and adults, trial additional dilemmas, and focus on the development of learning sequences for targeted constructs.

Conclusions

Our research approach is fundamentally integral (Figure 2), as it considers the environmental concerns expressed by learners, employs objective tests of developmental level, adds to the broader culture of environmental test taking by combining quantitative and qualitative methods to tell a more cohesive story, and is intended to provide a functional fit to the collective needs of educators who seek to enhance environmental and sustainability education. As we continue to develop the LESA, we will continue to inquire into the genetic epistemological question of how ecological understanding, values, and skills develop over time.

To date, no studies have focused on understanding how ecological stewardship develops within individuals over time. We believe that the continued development of the LESA will yield knowledge that can support the design of higher quality, developmentally appropriate environmental and sustainability education curricula. By providing detailed descriptions of reasoning along specific themes at different developmental levels, we can help teachers find and support learning in each student's learning sweet spot. We argue that by supporting the optimal development of students' understanding of ecological stewardship in this way, we can increase the likelihood that students will become effective stewards of the environment.

Figure 1. Seven S curriculum model for Environmental and Sustainability Education (ESE)

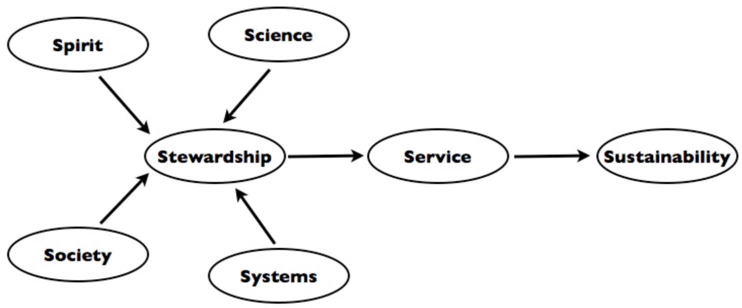

Figure 2. Integral psychometrics, indicating an all-quadrant perspective of this study

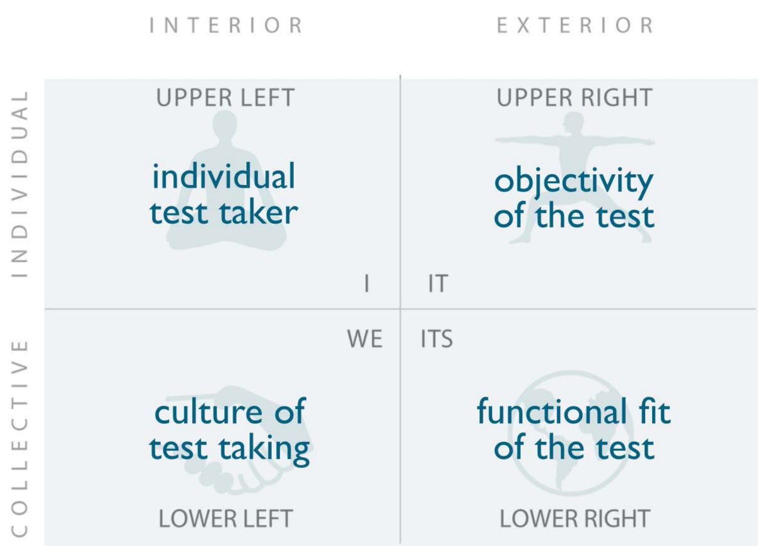

Table 1. Descriptive statistics for RFJ-T1, RFJ-T2, and LESA

	RFJ-T1	RFJ-T2	LESA
All students			
n	49	41	37
Mean	9.61	9.67	9.76
SD	0.17	0.18	0.15
Male			
n	26	24	17
Mean	9.62	9.62	9.75
SD	0.15	0.18	0.19
Female			
n	23	17	20
Mean	9.59	9.73	9.76
SD	0.19	0.15	0.13

RFJ-T1 (September 2014); RFJ-T2 (January 2015) and LESA (February 2015). n = number of students.

Table 2. Average responses per student on RFJ and LESA for threshold codes.

	RFJ-T2						LESA				
	Lower	**Higher**		**Ratio**	**n**		**Lower**	**Higher**		**Ratio**	**n**
Students: number (% of total)	31 (76%)	10 (24%)					13 (35%)	24 (65%)			
Actions	0.32	0.30		0.93	13		0.15	0.38		2.44	11
Causes of warming	0.97	0.70		0.72	37		0.54	1.17		2.17	35
Hot gases	0.35	0.50		1.41	16		0.31	0.54		1.76	17
Consequences of taking action	0.87	1.30		1.49	40		0.77	1.33		1.73	42
Don't take action	0.29	0.50		1.72	14		0.46	0.54		1.17	19
Take action	0.48	0.70		1.45	22		0.31	0.71		2.30	21
Effects of changing climate	2.97	3.80		1.28	130		3.00	4.08		1.36	137
Negative effects	2.55	3.50		1.37	114		2.54	3.63		1.43	120
Animal health effects	0.26	0.60		2.33	14		0.15	0.54		3.52	15
Hotter climate effects	0.29	0.30		1.03	12		0.31	0.33		1.08	12
Human health effects	0.61	0.70		1.14	26		0.77	0.83		1.08	30
Interdependent life*	0.13	0.20		1.55	6		0.54	0.17		0.31	11
Polar ice melting*	0.23	0.30		1.33	10		0.08	0.33		4.33	9
Problems with plants/trees	0.19	0.40		2.07	10		0.23	0.29		1.26	10
Positive effects	0.23	0.20		0.89	9		0.31	0.29		0.95	11
Pollution	0.35	0.10		0.28	12		0.38	0.29		0.76	12
Sources of information	1.23	1.50		1.22	53		1.23	1.58		1.29	54
Internet	0.55	0.50		0.91	22		0.46	0.50		1.08	18
Reading books/articles	0.35	0.70		1.97	18		0.38	0.54		1.41	18

Lower = scores 9.31-9.69. Higher = scores >9.69. Ratio = (average higher response frequency per student)/(average lower response frequency per student). n = sum of total responses of higher and lower for each node for each assessment. Items in bold are the broadest level of coded themes, items in italics are sub-themes, and indented items are single (non-aggregated) codes. Codes marked with an * were not used for statistical analysis due to low sample size.

References

Baldwin, J. M. (1906). *Thought and things: A study in the development of meaning and thought or genetic logic* (Vol. 1-4). New York: Macmillan Co.

Bazeley, P., & Jackson, K. (2013). *Qualitative data analysis with NVivo*. Thousand Oaks, CA: SAGE Publications Limited.

Berkowitz, A. R., Ford, M. E., & Brewer, C. A. (2005). A framework for integrating ecological literacy, civics literacy and environmental citizenship in environmental education. In E. A. Johnson & M. J. Mappin (Eds.), *Environmental Education and Advocacy: Changing Perspectives of Ecology and Education*. (pp. 227-266). Cambridge: Cambridge University Press.

Catamount Center (2015). http://www.catamountcenter.org/tree/ Accessed May 12, 2015.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Coyle, K. (2005). Environmental Literacy in America: What Ten Years of NEETF/Roper Research and Related Studies Say About Environmental Literacy in the U.S. Washington, D.C.: The National Environmental Education & Training Foundation.

Dawson, T. L. (2002). A Comparison of Three Developmental Stage Scoring Systems. *Journal of Applied Measurement*, 3(2), 146-189.

Dawson, T. L. (2006). Stage-like patterns in the development of conceptions of energy. In X. L. W. Boone (Ed.), *Applications of Rasch measurement in science education* (pp. 111-136). Maple Grove, MN: JAM Press.

Dawson, T. (2014). A confirmatory Rasch analysis of the RFJ001. International Objective Measurement Workshop. Philadelphia.

Dawson, T., & Seneviratna, G. (2015). *Cultivating the integral mind: fostering the emergence of increasingly complex and integrated neural networks*. Paper presented at the Integral Theory Conference, Santa Rosa, CA.

Dawson, T. L. and Z. Stein (2008). "Cycles of research and application in education: Learning pathways for energy concepts." *Mind, Brain, & Education* 2(2): 90-103.

Dawson, T. L., & Stein, Z. (2011). We are all learning here: Cycles of research and application in adult development. In A. Hoare (Ed.), *Oxford handbook of reciprocal adult learning and development*. London: Oxford University Press.

Dawson, T. L., & Stein, Z. (2011a). *Virtuous cycles of learning: redesigning testing during the digital revolution*. *The International School on Mind, Brain, and Education*. Paper presented at the Ettore Majorana Center for Scientific Culture, Erice (Sicily), Italy.

Dawson, T. L. and Z. Stein (2012). Measuring the growth of reflective judgment with cognitive developmental assessments. Annual Meeting of the Northeastern Educational Research Association. Rocky Hill, CT.

Dawson, T. L., Xie, Y. Y., & Wilson, M. (2003). Domain-general and domain-specific developmental assessments: do they measure the same thing? *Cognitive Development, 18*, 61-78.

Dawson-Tunik, T. L. (2004). "A good education is:" The development of evaluative thought across the life-span. *Genetic, Social, and General Psychology Monographs, 130*(1), 4-112.

Eilam, E., & Trop, T. (2010). ESD Pedagogy: A Guide for the Perplexed. *The Journal of Environmental Education, 42*(1), 43-64.

Esbjörn-Hargens, S., & Zimmerman, M. E. (2009). *Integral ecology: Uniting multiple perspectives on the natural world*. Boston, MA: Shambhala Publications.

Fischer, K. W. (1980). A theory of cognitive development: The control and construction of hierarchies of skills. *Psychological Review, 87*(6), 477-531.

Garrison, J., Östman, L., & Håkansson, M. (2015). The creative use of companion values in environmental education and education for sustainable development: exploring the educative moment. *Environmental Education Research, 21*(2), 183-204.

Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*. New York: Routledge.

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science, 28*(1), 127-138.

Hollweg, K. S., Taylor, J. R., Bybee, R. W., Marcinkowski, T. J., McBeth, W. C., & Zoido, P. (2011). Developing a framework for assessing environmental literacy. Washington, DC: North American Association for Environmental Education.

King, P. M. (2009). Principles of development and developmental change underlying theories of cognitive and moral development. *Journal of College Student Development, 50*(6), 597-620.

King, P. M., & Kitchener, K. S. (1994). *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults*. San Francisco, CA: Jossey-Bass

Kitchener, K. S., & Fischer, K. W. (1990). A skill approach to the development of reflective thinking. *Contributions to Human Development, 21*, 48-62.

Koger, S. M., & Winter, D. D. (2011). *The psychology of environmental problems: Psychology for sustainability*. New York: Psychology Press.

Lectica (2015). https://www.lectica.org/ Accessed May 12, 2015.

McBride, B., Brewer, C., Berkowitz, A., & Borrie, W. (2013). Environmental literacy, ecological literacy, ecoliteracy: What do we mean and how did we get here? *Ecosphere, 4*(5), art67.

Myers Jr., O. E., Saunders, C. D., & Garrett, E. (2004). What do children think animals need? Developmental trends. *Environmental Education Research, 10*(4), 545-562.

Perry, W. G., Jr. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart, and Winston.

Piaget, J. (1970). Piaget's theory. In P. Mussen (Ed.), *Carmichael's Manual of Child Psychology*. New York: John Wiley.

Piaget, J., & Kegan, P. (W. Mays, Trans.) (1972). *The principles of genetic epistemology*. London: Routledge.

Robottom, I. (2004). Constructivism in environmental education: beyond conceptual change theory. *Australian Journal of Environmental Education, 20*(2), 93-101.

Sadler, T. D., Klosterman, M. L., & Topcu, M. S. (2011). Learning Science Content and Socio-scientific Reasoning Through Classroom Explorations of Global Climate Change. In T. D. Sadler (Ed.), *Socio-scientific Issues in the Classroom* (Vol. 39, pp. 45-77): Springer Netherlands.

Saldaña, J. (2009). *The coding manual for qualitative researchers*. Thousand Oaks, CA: SAGE Publications.

Sheehy, N. P., Wylie, J. W., McGuinness, C., & Orchard, G. (2000). How Children Solve Environmental Problems: Using computer simulations to investigate systems thinking. *Environmental Education Research, 6*(2), 109-126.

Shepardson, D. P., Roychoudhury, A., Hirsch, A., Niyogi, D., & Top, S. M. (2014). When the atmosphere warms it rains and ice melts: seventh grade students' conceptions of a climate system. *Environmental Education Research, 20*(3), 333-353.

Simmons, D. (1995). Working paper# 2: Developing a framework for national environmental education standards *Papers on the development of environmental education standards* (Vol. 53, pp. 58). Troy, Ohio: North American Association for Environmental Education.

Simmons, D. A. (2005). Developing guidelines for environmental education in the United States: The national project for excellence in environmental education. In E. A. Johnson & M. J. Mappin (Eds.), *Environmental education and advocacy* (pp. 161-183). Cambridge, UK: Cambridge University Press.

Stein, Z., & Hiekkinen, K. (2009). Metrics, models, and measurement in developmental psychology. *Integral Review, 5*(1), 4-24.

Vygotsky, L. S. (1978). Mind and society: The development of higher mental processes. Cambridge, MA: Harvard University Press.

Wilber, K. (2000). *Integral psychology: Consciousness, spirit, psychology, therapy*: Shambhala Publications.

Wylie, J., Sheehy, N., McGuinness, C., & Orchard, G. (1998). Children's Thinking about Air Pollution: a systems theory analysis. *Environmental Education Research, 4*(2), 117-137.

Zeidler, D. L., Sadler, T. D., Applebaum, S., & Callahan, B. E. (2009). Advancing reflective judgment through socioscientific issues. *Journal of Research in Science Teaching, 46*(1), 74-101.

Appendix 1. Dilemma and associated questions for RFJ 001

The Dilemma
Some scientists think that all violent TV shows are bad for children, because they can teach children to hurt other people. Other scientists disagree. They think it is okay for children to watch some violent TV shows.

Question 1
Some scientists think that violent TV shows are bad for children. Others think some violent TV shows are okay.

a. Which group of scientists do you think is right? (Check the box next to your answer.)

- ❑ The scientists who think violent TV is bad for children
- ❑ The scientists who think some violent TV is okay
- ❑ Neither group
- ❑ Both groups

b. Explain your choice.

Question 2
How would you decide which group of scientists was right?

a. If you were trying to decide which group of scientists was right, what would you need to know? Why?

b. If you were trying to decide which group of scientists was right, where would you look for more information? Why?

c. If you were trying to decide which group of scientists was right, how would you know if you could trust what you learned?

d. If you were trying to decide which group of scientists was right, what else could you do to help you decide? Why would that help?

Question 3
If you were one of the scientists who thought violent TV was bad for children, what could you do to convince the other group of scientists that you were right?

a. Describe what you would do to convince the other scientists.

b. Why do you think this would get them to agree with you?

Question 4

If you were one of the scientists who thought some violent TV was okay for children, what could you do to convince the other group of scientists that you were right?
a. Describe what you would do to convince the other scientists.

b. Why do you think this would get them to agree with you?

Question 5
How is it possible that the two groups of scientists have such different ideas? Explain your answer.

Question 6
Is it possible to know for sure if violent television is bad for children?

a. Check the box next to your answer.

- ❑ yes
- ❑ no
- ❑ maybe

b. Explain your answer.