Article

A Learning Progression for Understanding Interdependent Relationships in Ecosystems

Sara J. Dozier [1,*], Anna MacPherson [2], Linda Morell [3], Perman Gochyyev [3] and Mark Wilson [3]

1   Department of Science Education, California State University, Long Beach, CA 90840-9504, USA
2   American Museum of Natural History, New York, NY 10024-5102, USA
3   Graduate School of Education, University of California, Berkeley, CA 94720-1670, USA
*   Correspondence: sara.dozier@csulb.edu; Tel.: +1-(562)-985-5921

Abstract: This paper describes a hypothesized learning progression for how secondary students understand interdependent relationships in ecosystems, a key concept in the field of ecology and for public understanding of science. In this study, a hypothetical learning progression was developed and empirically investigated using Rasch modeling of data from 1366 students in a large, diverse, urban school district. We found that the empirical evidence supported the general structure of the hypothesized learning progression for relationships in ecosystems. There were notable exceptions, and we describe the ways in which we altered the items and the learning progression to address empirical inconsistencies with our a priori conceptions. The assessment items developed through this study are immediately available online for formative assessment purposes, and the learning progression can support teachers' thinking about students' understanding of ecosystems. In particular, the upper reach of the learning progression offers a more complete description of the ways in which students might develop their understanding of complex interactions in ecosystems, beyond what is currently offered in the literature and standards documents about students' understanding.

Keywords: science education; learning progression; ecology; ecosystems; Rasch model; item response theory; validity

1. Introduction

An understanding of ecosystems, or biological communities of interacting organisms and their physical environment, has emerged through observations by humans over time and through empirical studies by ecologists. Knowledge of ecosystems has been beneficial to human endeavors, including agriculture and management of natural resources, since before recorded history [1,2]. However, an understanding of ecosystems by the general population is urgently needed today [3], as humans face critical challenges such as global pandemics and climate change. While knowledge of ecosystems is not sufficient to reach sustainability, it is a prerequisite. As Robert Egerton wrote in his history of ecology, the field "may be the most important science for managing the earth as an abode for humanity" [4] (p. 93). Life science education, for the most part, has supported the study of ecosystems in the curriculum. Ninety-eight percent of high schools offer a biology course and biology is available to the largest number of students compared to other science courses [5]. Therefore, nearly all students will learn about ecology during their formal schooling, making research in this area of science education potentially impact most students.

1.1. The Value of Learning Progressions

In recent years, learning progressions have been identified as useful tools for both instructional planning and formative assessment practice [6]. Learning progressions have been described as "conjectural or hypothetical model pathways of learning over periods of time that have been empirically validated" [7] and as "the successively more sophisticated

ways of thinking about a topic that can follow one another as children learn about and investigate a topic over a broad span of time" [8]. Others have argued, however, that narrow views of how learners develop conceptual understanding may exclude valid and useful ways of reasoning (e.g., Ref. [9]). In this study, we offer a map of the ways that an understanding of interactions in ecosystems develops in middle-grade students. Our proposed model is not a rigid, stepwise progression that describes misconceptions that need to be eradicated from students' repertoire as they progress toward a singular "best" understanding of ecosystems [10]. Rather, this progression takes up Sikorski's [9] "upper reach" by describing student understandings of ecosystems with differing degrees of cognitive complexity [11] that are employed in context-dependent ways. We have created this progression to intentionally avoid the use of student "misconceptions" at the less complex levels of the progression, such that each level is useful for reasoning about ecosystems [12]. Rather, we view each level of this progression as concepts that can be used by students in varying contexts or situations. It is this perspective that has led us to adopt a probabilistic approach using item response theory (explicitly the Rasch model [13]) to mathematically express the cognitive structure that we are building on (i.e., the construct map style of learning progression).

After more than a decade of increased attention to learning progressions in the science education research literature [14], many researchers agree that although learning progressions may be complex and challenging to validate empirically, they can be generative tools for curriculum development [15] and teachers' formative assessment practice [16,17]. In addition, learning progressions have gained prominence in the past decade for supporting the creation of instructional and assessment tools that support teaching and assessment of students by identifying increasing levels of complexity in any content area [6,18].

Research into learning progressions for ecology is critical because of the potential impact on learning about this topic for virtually all K-12 students. The concept of the food web, the central model used to predict and analyze interactions in ecosystems, is ubiquitous in middle and early high school biology courses. Thus, research on how students develop their understanding of ecosystems will enable the development of better instructional materials and assessments that reach a large number of students. Previous work has uncovered the ways in which students across a range of ages, from elementary students through graduate students, come to understand different concepts in ecology [19–23]. We build on this prior work, probing how students reason with the concept of interdependent relationships in ecosystems. Rather than focusing on recall, this learning progression describes "understanding," which we define as being able to reason with the concept and could be evidenced by making a prediction or identifying a correct explanation.

The Framework for K-12 Science Education [24] and the Next Generation Science Standards (NGSS) [25] provide guidance about how science is taught in grades K-12 in the United States. These documents identify the concept of "interdependent relationships in ecosystems" as a separate strand in the set of disciplinary core ideas (DCI) addressing life science. Four DCIs reflect unifying principles in life sciences: the DCI Life Science 2 (LS2) is called "Ecosystems: Interactions, Energy and Dynamics" and describes organisms' interactions with each other and their physical environment. LS2 is divided into sub-ideas A–D. LS2.A, the focus of this work, describes, specifically, "interdependent relationships in ecosystems." The NGSS suggest a progression for the way in which students might learn about ecosystem interactions. In elementary school, students learn that there are living and nonliving components of an ecosystem. In middle school, they learn that organisms in an ecosystem are interdependent, and a change to one population can affect other populations. In high school, students are introduced to the idea of carrying capacity or the idea that an ecosystem has a limit to the number of organisms and populations it can support.

Unlike some previous work, our learning progression focuses on a single DCI—interdependent relationships in ecosystems—rather than a set of DCIs or a combination of DCI and scientific practice. There are certainly benefits to examining learning progressions of practices, and practices combined with core ideas. In fact, the larger project in which

this work is situated has also examined learning progressions in physical science and scientific argumentation [26,27]. However, the analysis presented in this paper focuses exclusively on one DCI in an effort to focus on students' developing understanding of interdependent relationships in ecosystems, without making claims about their competence in a scientific practice or other DCIs. Other learning progressions research about ecology has looked at different concepts in ecology, in particular, biodiversity [19,20,23] and systems thinking [28,29], rather than focusing on interdependent relationships. We suggest that our learning progression, taken together with the various hypothesized and empirical learning progressions in the literature, offers a more complete picture of students' development of understanding in this discipline. As Duncan and Gotwals point out [30], research in learning progressions benefits from competing models, which can be compared and evaluated.

Our approach focuses on the complex thinking necessary for a student to understand ecosystems. Identifying the degree of cognitive complexity in science assessment tasks requires, in part, "the demands of the domain in which these cognitive activities are manifested" [11] (p. 37). To examine the cognitive demands of the ecological domain, we employ a framework describing how children reason about open dynamic systems, rather than the closed, decontextualized systems described in Piagetian literature [31]. In this framework, reasoning about dynamic systems requires, among other skills, systemic synthesis, the understanding that changing one part of a dynamic system would eventually affect all components of the system. The upper reach of this progression requires that students use systemic synthesis to explain the interactions between more than two populations in an ecosystem, which is the highest degree of cognitive complexity described in the literature about student learning in ecosystems (described in Section 1.2) and typically taught in a general high school biology course.

1.2. Review of Prior Learning Progression Research in Ecosystems

The learning progression presented in this paper draws on previous ecology-related learning progressions. In 2009, Songer, Kelcey, and Gotwals [23] proposed a learning progression that consisted of two dimensions: a content dimension about biodiversity and an "inquiry/reasoning" dimension (which now according to the Next Generation Science Standards, NGSS, we would refer to as a scientific practice) that described how students constructed evidence-based explanations. In a subsequent paper, Gotwals and Songer [20] described a progression for how sixth-grade students reasoned about food chains and food webs. They proposed a matrix for how content knowledge could be combined with the practice of constructing evidence-based explanations. They focused in particular on students' intermediary knowledge, that is, not the upper and lower anchors of the progression. They described a "messy middle" that learners seemed to progress through as they developed more sophisticated reasoning about ecosystems.

Continuing to build on this work, Gotwals and Songer [19] described a progression that consisted of constructing evidence-based explanations and three foundational ideas in ecology: classification, ecology, and biodiversity. However, the "ecology" strand of the Gotwals & Songer [19] progression combined several different ideas: the lower bound specified students having an understanding that "every organism needs energy to live" and the upper bound that "a change in one species can affect different members of the food web". Though both of these ideas are addressed in the field of ecology, they represent different core ideas of the discipline, that is, the idea that organisms need energy versus the idea that organisms have interdependent relationships. In fact, these are two separate strands of disciplinary core ideas in the Next Generation Science Standards (Life Science 2A and Life Science 2B).

Hokayem and Gotwals [21] extended prior work on developing a learning progression for ecosystems to younger grades. They developed a progression of reasoning about ecosystems that spanned from anthropomorphic reasoning (students reasoning with personal feelings only, rather than relying on scientific concepts) through complex, causal reasoning.

They found that students often used multiple levels of reasoning in a single scenario, making it difficult to place students along a linear progression of development. However, they found that early elementary students were capable of causal reasoning about ecosystems, which is not a typical expectation for elementary science.

A significant body of research investigates how students develop reasoning about feedback loops and, in particular, predator-prey systems [22,32–34]. The main finding from these studies is that linear, one-way causal thinking dominates reasoning patterns across age cohorts and grades. For example, Hokayem, Ma, and Jin [34] asked elementary students to explain how populations change in two contexts: a self-sustaining ecosystem and an ecosystem that is missing predators. Very few students recognized the cyclical relationships among populations in a sustainable ecosystem, leading the authors to develop a learning progression for how students might develop feedback loop reasoning in early grades.

Eilam [32] studied much older students but found similar results. The aim of the study was to examine 9th graders' understanding of the complex, multilevel, systemic construct of feeding relations, nested within a larger system of a live model. Fifty ninth graders interacted with the model and manipulated a variable within it. Even at the end of the learning module, many students did not recognize the cyclic nature of relationships and defaulted to unidirectional, linear descriptions of predator-prey relationships.

Hovardas [22] worked with pre-service science teachers and found that during three teaching units about predator-prey dynamics, a majority of participants defaulted to unidirectional, linear reasoning about a predator-prey system (that is, as wolves increase, deer always decrease), rather than an oscillating model, even when presented with data that showed the non-linear nature of the relationship. An implication of this work is that thinking may not proceed in an orderly, linear way through a learning progression about relationships in ecosystems; rather, understanding may proceed through various stages that might not be easily predicted or facilitated by learning experiences.

Jin et al. [28,29] contributed to clarifying the way in which students' thinking may develop by proposing a learning progression for systems thinking in ecosystems. Their progression describes how secondary students use discipline-specific systems thinking concepts [35] to explain interdependent relationships in ecosystems. Using an assessment consisting of constructed response items in which students explained real-world ecological phenomena (e.g., wolves in Yellowstone, hares, and lynx in Canada), they investigated a four-level learning progression. Jin et al. created a multidimensional construct that included all three dimensions of the NGSS: the disciplinary core ideas of interdependent relationships in ecosystems and human impacts on earth systems; the crosscutting concept of systems and systems models; and the science and engineering practice of constructing explanations. The resultant progression describes students' ability to reason using all three dimensions of the NGSS to explain natural phenomena. Because of the multidimensionality of the construct and the constructed response item format, it is difficult to tell which component of the construct, or the language demands, determines students' performance.

The learning progression described in this paper builds on these previous studies to address the practical demands of current science reforms to create a tool that can be immediately useful for teachers' use in instructional interventions, teacher education, and professional development [29]. Previous work has provided useful information about how students' thinking about ecology develops. Our study builds on these findings by (1) using context-rich tasks that center on natural phenomena; (2) using carefully constructed selected response items that reduce the language demands of describing reasoning at the highest levels of the progression [26]; and (3) examining students' performances using items targeted toward their conceptual understanding of ecosystems within tasks (scenarios) that target two or more dimensions as described in the NGSS. The scope of this study is restricted to the items targeting students' conceptual understanding of ecosystems, while the items targeting other dimensions will be described in future studies.

This paper is structured to first describe the iterative process used to develop the learning progression and companion assessment, and this is followed by the empirical findings, a discussion of how groups might use the material, and suggestions for future research.

1.3. Research Question

We aimed to describe, test, and revise a learning progression for a key idea in science, interdependent relationships in ecosystems, that (a) could be immediately useful to middle school science teachers, and (b) provide a model of how we and other researchers could go about investigating and validating a learning progression for other ideas, practices, or concepts. The research question that guided our study was: How do middle and high school students develop an understanding of the interdependent relationships in ecosystems?

2. Materials and Methods

We used the Standards for Educational and Psychological Testing [36] to frame our validity argument, and as recommended in Developing Assessments for the Next Generation Science Standards [37], we used a construct modeling approach called the Berkeley Evaluation and Assessment Research Assessment System (BAS) [38] to develop and empirically investigate a learning progression for student understanding of interdependent relationships in ecosystems. The BAS provides the assessment framework to develop and refine the learning progression itself and the assessment materials used to validate the progression. BAS also provides support for using the progression and assessment materials in the classroom. As with Mislevy's Evidence-Centered Design (ECD) [39,40], the BAS systematizes the assessment development process and provides a model for understanding the connections between different assessment elements [37]. The BAS includes four building blocks used for constructing quality assessments: the construct map, items design, outcome space, and measurement model. These building blocks map to the National Research Council Assessment Triangle developed by the NRC Committee on the Foundations of Assessment [41]. Each building block is described in the following text and shown in Figure 1. These building blocks represent steps in a cycle of development, which may repeat several times in order to refine the construct maps, the items, and the scoring guides. Because the final, empirically validated construct map is very similar to the learning progression in both structure and content, we will use the term learning progression throughout this paper (see Section 2.3).
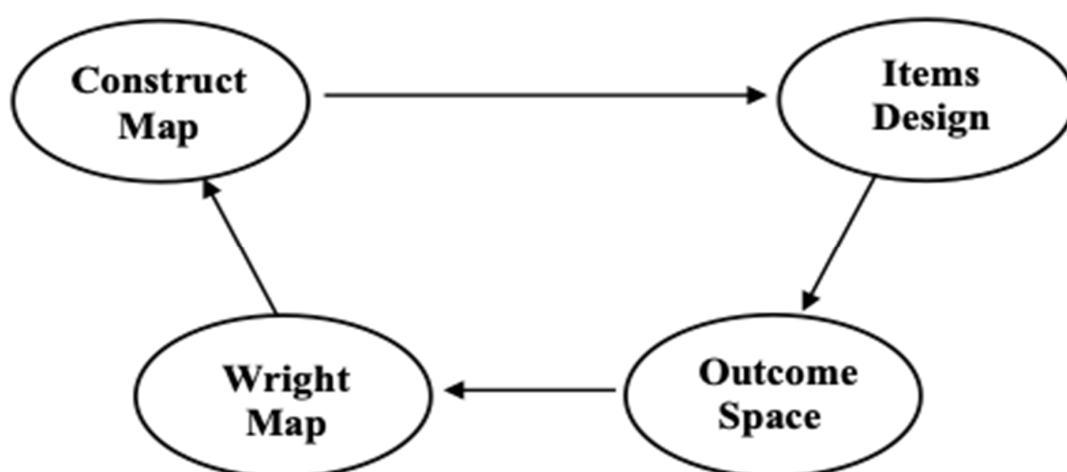

Figure 1. Schematic of the BEAR Assessment System (BAS).

2.1. Sample

A total of 1366 6–10th grade students in a large, diverse, urban school district participated in answering the computer-based assessment. All students provided active assent to participate in the study and active parental consent was obtained for students who partici-

pated in think-aloud interviews, as required under our IRB protocol code 2010-09-2241. The district enrolls over 55,000 students, including students who identify as African American (11%), Chinese (28%), Filipino (6%), Latin (26%), White (11%) and other (18%). In this district, 15% of students are classified as English Language Learners, 13% have an IEP, 33% are identified as GATE students, and 58% qualify for the federal free or reduced-price lunch program.

2.2. Overview of Procedure

Over two years, we engaged in an iterative process of creating a hypothesized learning progression, developing items using data from student think-aloud interviews and teacher focus group interviews; testing these items in computer-based administrations; analyzing the data from the computer-based administrations; modifying and reclassifying some items; and modifying the learning progression. Each step of our research followed the BEAR Assessment System (Figure 1) and is briefly described in Figure 2. The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board of the University of California, Berkeley (protocol code 2010-09-2241; date of approval: 22 January 2020).
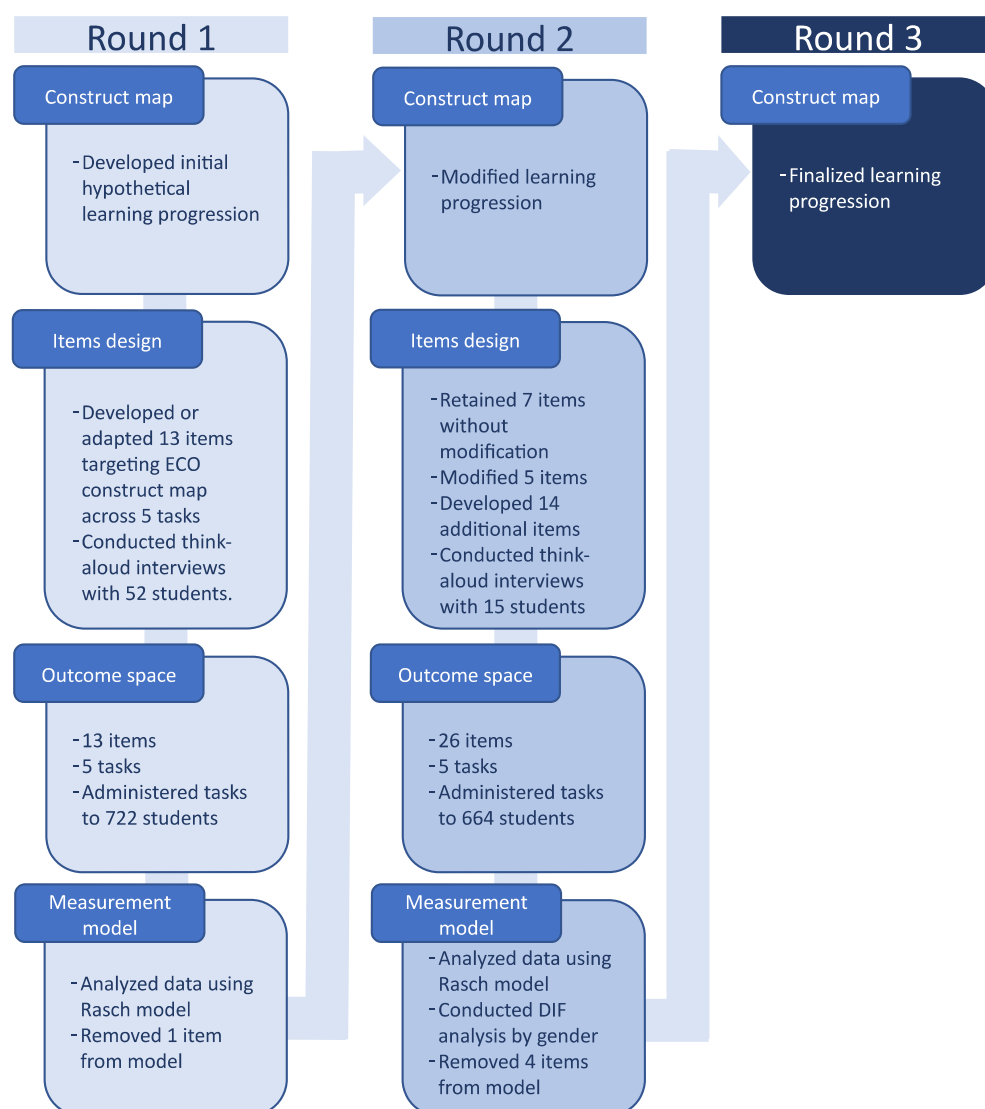

Figure 2. Overview of data collection and analysis procedures.

2.3. Construct Map/Learning Progression

A construct map is a foundational feature of the BAS. It provides an ordering of waypoints, qualitatively different levels of student understanding, and performance focusing on one characteristic derived in part from research into the underlying cognitive structure of the domain, and in part from professional judgments about what constitutes higher and lower waypoints of performance or competence. Construct maps are informed by empirical research about how individuals perform in practice. They are the essential link between the instructional design aspect of the learning progression and the assessments used to evaluate student learning. Wilson [38] identifies several possible types of relationships between the construct map and a learning progression. We adopt one of the relationships he identifies here which indicates "the levels [waypoints] of the learning progression are the [waypoints] of a single construct map" (p. 725). Thus, we will refer to modifications of the learning progression rather than the construct map.

The research team first developed a hypothetical learning progression (Student Understanding of the Interdependent Relationships in Ecosystems, or ECO, learning progression) based on our own prior research, which involved interviews with ecologists in various sub-fields [42] and sources in the relevant literature (e.g., [19]). We relied on published assessment items and standards documents—the Atlas of Science Literacy [43], the Next Generation Science Standards [25], and A Framework for K-12 Science Education [24]—for information about the ways in which the progression of these ideas in ecology are operationalized for students and teachers. The hypothetical learning progression was revised in two rounds in response to empirical evidence provided by students. We grounded our decisions about how to revise the learning progression in evidence about whether the learning progression fully captured and adequately differentiated students' conceptual understanding. When nuanced student thinking was not captured by the existing waypoints of the progression, progression was expanded to include these ideas. Waypoints were combined when items had similar difficulty and we deemed that combining the waypoints would not lead to lost information about students' understanding [44]. The final learning progression, shown in Table 1, contains three qualitatively distinct waypoints, which form the hypothesized increasing complexity of student understanding.

Table 1. Learning Progression of Understanding Interdependent Relationships in Ecosystems.

Level	Description
	Complex Relationships
3	Students predict changes in more than two components in an ecosystem based on changes in microscopic or macroscopic populations or the availability of non-living resources [31].
	Indirect Relationships
	Students predict the effects of change in one population on another population with an indirect relationship [21,43].
2	Students predict the effects of the availability of and competition for resources (e.g., food, space, water, shelter, and light) on populations [43].
	Direct Relationships
	Students predict the effect of a change in the size of one population on the size of another population in mutual, commensal, or parasitic relationships [19,20,43,45].
1	Students predict the effect of a change in the size of one population on the size of another population in a predator-prey relationship [19–21,43,45].
	Students predict the effects of changes in plant populations throughout the food web using the knowledge that plants form the base of the food web and are living organisms [19,20,43,45].
	Notions
0	Students express naive knowledge about ecosystems.

Note: The initial map was hypothesized from existing literature about the development of children's ideas about ecosystems and findings from AAAS Project 2061. Level 3 expands on the existing literature (e.g., Ref. [43]) by conjecturing that students can reason about complex relationships in ecosystems and that this might represent a meaningful "upper anchor" of a progression of understanding of interdependent relationships.

2.4. Items Design

The items design is a process where the construct map/learning progression is operationalized by developing tasks that provide evidence of the highest level of cognitive complexity demonstrated by a student on the construct of interest. This is achieved through the systematic design of tasks to elicit the specific types of evidence about student knowledge as described for the waypoints of the learning progression.

Based on the hypothetical progression, we developed a total of 27 selected response items (ECO items) to address the three waypoints of the ECO learning progression in the context of five assessment tasks named Foxes, Invasive, Lions, Succession, and Whales (all tasks and items are shown in Supplemental Materials). The scenario shown in Figure 3 describes the phenomenon of mountain lions interacting with other organisms in an ecosystem, including a cattle ranch. Two items from the Lion task are shown. Each task comprises a scenario, 3 to 8 items targeting the ECO learning progression, and 1 to 5 items targeting a previously published argumentation learning progression [27]. Although students responded to the items targeting scientific argumentation, we consider the argumentation items theoretically distinct from the ECO items. Therefore, the argumentation items were not included in the analysis for the current study.

Lion Task

The organisms in the food web live near large cattle ranches. Over many years, mountain lions occasionally killed a few cattle. One year, a few ranchers hunted and killed many mountain lions to prevent future loss of their cattle.

The arrows point from the organisms being eaten to the organisms that eat them.
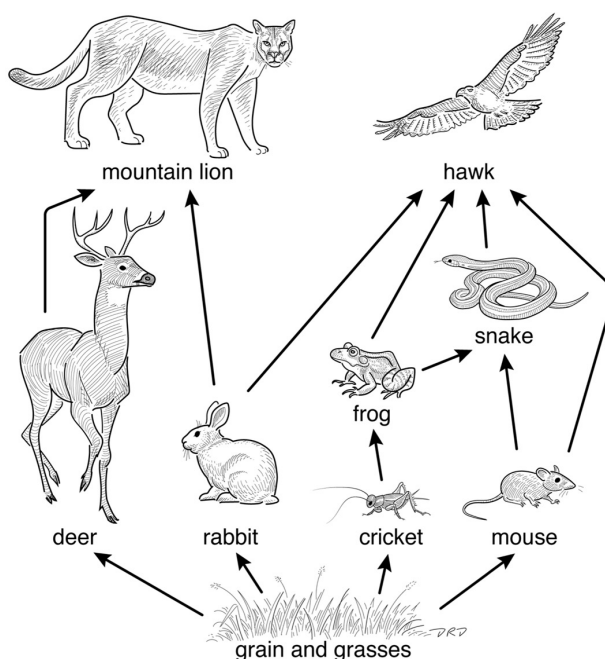
The frog and the cricket have a relationship which is best described as...

    A. <u>One benefits while the other is harmed.</u>
    B. They both benefit from each other.
    C. They compete for resources.
    D. One benefits while the other is unaffected.

Identify two specific populations that most likely increased in number after the cattle ranchers reduced the mountain lion population.

    a. Hawk, frog
    b. <u>Deer, rabbit</u>
    c. Frog, rabbit
    d. Deer only

Figure 3. "Lion" task with selected items. The best answer choices are underlined in this figure but not in student-facing materials.

Items were designed and revised based on information from interviews with both teachers and students. Four teacher focus group sessions were conducted with the research team and 3 to 5 middle school science teachers from our participating district. Focus groups were used because of the affordances of having participants discuss the tasks with one another [46]. The teacher focus group participants had previously participated in the larger research project in which the current study was situated. Participants were given time to review each task and asked whether (a) the items reflected content taught in their science classes, (b) students who understood the content would find an item confusing, and (c) the items contained language that was inaccessible to their students. Notes from these focus group interviews were used by the research team to modify items and task scenarios to reduce construct-irrelevant variance.

Think-aloud interviews [47,48] were conducted with 67 middle and high school students to elicit students' interpretations of the items. Interviews lasted 20 to 30 min and were audio-recorded. The interviewer took notes during the interview and wrote an analytic memo following each data collection session [49,50]. Using the think-aloud protocol (Supplemental Materials), the interviewer instructed students to first read the item aloud, then answer the question aloud, to write or mark their response either on a paper-based version of the task or into the computer-based platform, and finally to move on to the next item in the series. Audio recordings of these interviews were used by members of the research team to modify items and task scenarios.

2.5. Outcome Space

The outcome space defines the qualitatively different levels of item responses arising from a particular prompt or stimulus. Essentially, this is where interpretative value is placed on student work through the mapping of the student responses to the waypoints of the construct. These are expressed as scoring guides for particular items or item-types that show how the features of student responses can be mapped onto the construct waypoints.

2.6. Wright Map

A Wright map describes how inferences about student understandings are drawn from the evaluated (scored) work. This process is where the values derived from the outcome space are translated back to the learning progression, typically through a psychometric model. This is a crucial aspect of the validation stage of the process.

The essential validity question was whether the empirical structure of the data matches the theoretical structure of our learning progression. Each item's capacity to distinguish among people helps to capture and categorize any given person's location on the learning progression more accurately. We used the Rasch model [13] to fit the data. The Rasch model is one of several models within the Item Response Theory (IRT) family of analysis. Using this model allows for an investigation of the measurement properties of the test by generating person ability estimates, item difficulty calibrations, item fit statistics, and person fit statistics. As de Ayala [51] indicates "the Rasch model is the standard by which one can create an instrument for measuring a variable" (p. 19). The Rasch model is a model used to construct the underlying latent variable [38,51,52].

These analyses used the partial credit model, a polytomous version of the Rasch model [52], yielding item difficulty estimates, standard errors of the estimates, the weighted mean-squares (fit indicators), and the corresponding t-statistic for each item. The (weighted fit) mean-square value is a measure of the fit of the items to the measurement model; it represents a ratio of the observed residuals to the expected residuals. According to Wilson [38], mean square values between 0.75 (low MNSQ) and 1.33 (high MNSQ) are generally acceptable. Items that had mean squared values outside the acceptable range were examined to understand what might have caused the poor performance. If found to be faulty on their face, misfitting items were removed from further analysis. Note that removing items does not change the underlying substantive theory. The cognitive model is embodied in the construct map, which we describe in detail above; the construct remains

the same, so the substantive theory is not changed. In our approach to measurement [53], the items are seen as potentially enabling transductions from the construct (the "theory" in this case) into observable events (student responses). We examine these for evidence about the success of the transductions and adapt the item set to achieve a sound result, just as any scientist/engineer would do. We argue that retaining those items would increase the likelihood that inferences about student understanding and the increasing cognitive complexity of the construct levels would be incorrect [38].

A unique product of using the Rasch model [13] is a map that shows the modeling of person ability estimates and item calibrations. This visualization, called a Wright map, provides a summary of both the respondent distribution and the item difficulties for the overall assessment together on a single (logit) scale. We used the Wright map to visualize coverage of student ability and the relationship between item difficulty and the waypoints of the learning progression. We used ConQuest 5.0 [54] and other analysis software for this investigation.

We investigated fairness using differential item functioning (DIF) [55] to ensure the survey items are psychometrically sound when used across different respondent subgroups. We investigated each item for DIF based on the gender variable, as this was the only variable with a sufficient number of students to estimate DIF. An item is considered to have DIF if, for example, two respondents from different groups (i.e., girls and boys) have different probabilities of answering in the same response category even though they have the same amount of the latent trait, in this case, the ability to understand the interdependent relationships in ecosystems. Within the Rasch model, the test for DIF is specified by allowing the overall item difficulty to differ across demographic categories while controlling for the overall construct. Items that showed a significant DIF effect were further investigated and when issues with the items appeared to cause bias, the items were removed from the analysis [55].

2.7. Test Administrations and Revisions

The items and learning progression were iteratively revised, using evidence from think-aloud interviews and Rasch modeling of the items. We approached this work with the perspective that the items themselves should be fair and reliable and that the learning progression should be supported by empirical evidence from students. The items were administered to 1366 students, in two rounds with 722 and 664 participants, respectively. During the first round of development, the initial hypothesized learning progression was used to create 13 selected response items across the five tasks. Student responses from think-aloud interviews were used to revise the items. Then, we administered the items to middle and high school students in our participating school district (*N* = 722). We found that of the 13 items targeting the ECO learning progression, 12 met the requirements for inclusion (had a mean-square error within an acceptable range), so the final model from round 1 included data from 12 items.

First, we examined each item itself, evidence from teacher focus groups, and student think-aloud data to determine which level of learning progression each item targeted. Of the 12 items retained from the first round of data collection, 5 items were modified to reflect more clearly the waypoints in the learning progression. Based on the data collected during the first-round administration and student think-aloud interviews, we compared the alignment of the empirical difficulty estimates of the remaining 12 items in relation to the hypothesized learning progression. Our hypothesized progression postulated that students would find reasoning about relationships between individual organisms less difficult than reasoning about changes in the populations of those organisms. However, the data from our first round of analysis did not support this hypothesis. We found that difficulty estimates for items describing the relationship between two individual organisms were indistinguishable from items that asked about interactions between populations of organisms. Further, we found that the type of relationship between the organisms (e.g., predator-prey, competition) and the role of abiotic factors in those relationships were

more strongly related to the difficulty of the items. Taken together, these data indicated that understanding relationships between populations of organisms was not contingent on a prior understanding of the relationships between individual organisms. Thus, we modified the progression by collapsing the waypoints distinguishing relationships between individual organisms and relationships between populations [44].

Because level 2 had few associated items leading to low sampling of this level of the LP, we developed 14 additional items to ensure that each level of the learning progression was targeted by at least five items across at least three tasks. These 26 items across the five tasks were administered to students (*N* = 664). Analysis using the Rasch model revealed that two items correlated poorly with the other items targeting this learning progression, with item-rest correlations of −0.18 and 0.06, respectively [56]. After examining the items more closely, we found ambiguity in the response options that could lead to a student with higher ability choosing an incorrect response and removed these two items from the analysis.

The items were administered to students using the BEAR Assessment System Software (BASS) [38], a web-based platform. Students most often used Chromebooks or iPads to complete the tasks under typical conditions found in the school setting. Extra time or other recommended accommodations were made depending on students' needs. Make-up or alternative administrations were made on a case-by-case basis in collaboration with the students' teachers.

3. Results

3.1. Rasch Modeling

Table 2 shows the results from round 2 (*N* = 664) of learning progression development for the final set of 22 remaining items (see DIF section) from the five tasks: Foxes (F), Invasive (N), Lion (L), Succession (S), and Whales (W). Results shown include the item difficulty estimates based on the partial credit model, the standard errors of the estimates, the weighted mean squares, and the corresponding t-statistic for each item. The items ranged in difficulty from −2.804 to 1.795 logits. Each of the final set of 22 items had a mean-square error within the acceptable range (0.75–1.33).

Table 2. Response model parameter estimates for the Interactions in Ecosystem items.

Task	Item Name	Construct Map Level	Difficulty Estimate (Logits)	Standard Error of the Estimate	Weighted Fit MNSQ	*t*
Lion	L1	1	−2.008	0.115	0.99	−0.1
Lion	L12	1	−1.615	0.104	0.99	−0.2
Lion	L14	1	−2.85	0.153	0.98	−0.1
Lion	L15	1	−1.725	0.107	0.93	−1.1
Lion	L16	1	−2.804	0.15	0.93	−0.6
Lion	L18	3	1.724	0.106	1.02	0.3
Lion	L19	3	0.417	0.085	1.19	5.9
Lion	L20	1	−1.795	0.109	0.94	−0.9
Whales	W1	1	−2.752	0.153	0.92	−0.7
Whales	W5	2	−0.832	0.093	0.97	−0.8
Whales	W6	2	−0.772	0.092	1.04	1.1
Whales	W8	3	0.652	0.091	1.13	3.7
Whales	W9	3	0.159	0.089	1.06	2.0
Foxes	F2	1	−2.027	0.127	0.90	−1.2
Foxes	F3	3	0.272	0.091	1.02	0.6
Foxes	F6	1	−1.493	0.112	0.98	−0.4
Foxes	F11	3	0.62	0.095	1.12	3.4
Succession	S1	1	−2.041	0.135	0.96	−0.4
Succession	S3	2	−0.007	0.096	1.02	0.5
Invasive	N1	1	−1.029	0.115	0.85	−2.8
Invasive	N8	2	−0.79	0.112	0.88	−2.5
Invasive	N9	2	−0.844	0.114	0.96	−0.8

The Rasch model yielded evidence about which items were more and less difficult for students to answer correctly. To investigate the match of the empirical results with

the theoretical levels posed by the learning progression, we generated a Wright Map (Figure 4), which places the person-ability estimates of the middle and high school students in our sample and calibrated item difficulties together on the same scale. Each person icon represents the ability estimate for one student. The items are identified by the item name shown in Table 2 and color-coded based on the level of the learning progression they were designed to measure. Results showed that items targeting each of the higher levels of the learning progression were indeed more difficult than the items targeting the lower levels, and items targeting the same level were generally clustered together (see Figure 4). The items Succession 3 (S3) and Invasive 1 (N1) were more difficult than other items in their respective learning progression levels. Both items described plants not as producers that transform energy from sunlight into chemical energy, but rather as organisms competing with other organisms for resources. Based on responses during think-aloud interviews, we hypothesize that students find it more difficult to reason about plants as competitors, rather than simply as producers. The clear separation of the items targeting the successive levels makes it possible to develop a "banding" of both items and students on the Wright map, as shown in Figure 4. Thus, the Wright map can be used to generate interpretable reports for teachers, focused on the educational integration of the instructional implications of the levels of the final learning progression (Table 1).
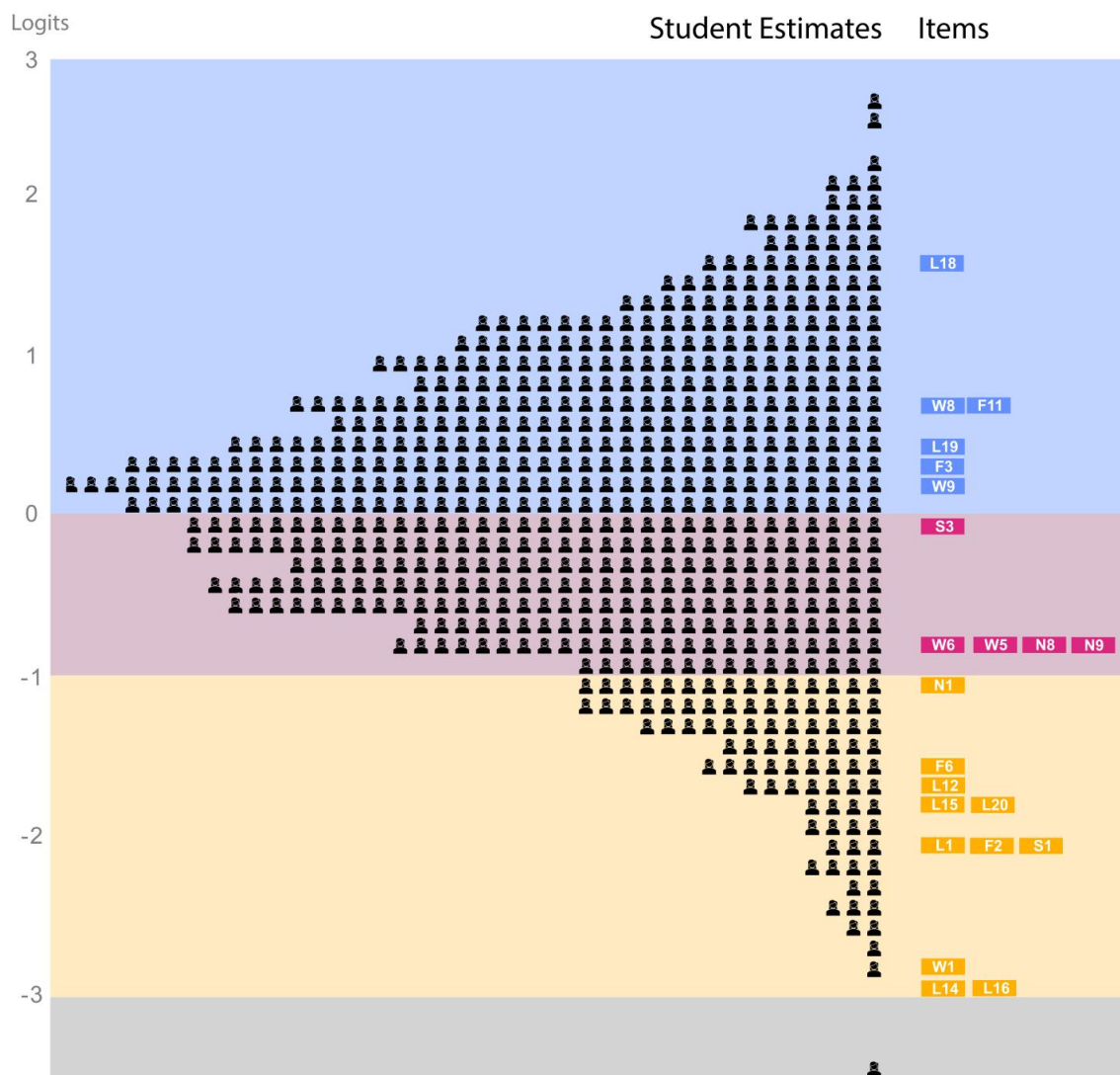
Figure 4. Latent distributions for the understanding of interdependent relationships in ecosystems (*N* = 664). Each 👤 represents one student. Each colored box represents an item with the number

corresponding to the difficulty estimates shown in Table 2. The colored bands show the relationship between student ability estimates, item difficulties, and learning progression levels. The gray band represents level 0. The yellow band represents level 1. The pink band represents level 2. The blue band represents level 3.

3.2. Reliability

We found support for acceptable reliability for the final set of 22 items. The expected (Posteriori/Plausible Value (EAP/PV) reliability is a measure of the overall reliability of the sample based on the student ability estimates from the Rasch analysis. This is the equivalent of Cronbach's alpha in the IRT context and can be interpreted in a similar fashion, with values greater than 0.70 considered to provide supportive evidence of inferences of reliability [57]. The EAP/PV estimate for our analysis is 0.701, which provides evidence in support of adequate reliability.

3.3. Differential Item Functioning

Fairness was examined in the 24 items with high item-rest correlations using DIF based on the gender variable. Two items asking students to identify living and non-living components of an ecosystem (Foxes 1 and Succession 2, shown in the Supplemental Materials) had higher difficulty estimates (0.704 and 0.796 logits) for male-identified students than for female-identified students, based on the student's estimated ability. That is, if a male- and female-identified student had equal ability to reason about interactions in ecosystems, the female-identified student would be much more likely to answer these two items correctly than a male-identified student. Both of these items targeted the ability to identify living and non-living components of ecosystems. Prior research has suggested that boys are less likely than girls to identify plants as living organisms [58–60]. Given this prior research base and the evidence of DIF in our sample, these two items were eliminated from the analysis, which left a total of 22 items for model construction.

3.4. Validated Learning Progression

In the final iteration of learning progression development, we used data from the second round of data collection together with the theoretical literature and standards documents to create our final empirically validated learning progression, shown in Table 1. We found that the least difficult items (in band 1) required reasoning about plants as the basis of the food web and reasoning about the interactions between two populations of organisms directly interacting in predator-prey, commensal, or mutualistic relationships. Because these items were the least difficult, these concepts are likely ones that have been mastered by most middle school students.

The cluster of items of intermediate difficulty (in band 2) required students to reason about competition for abiotic resources or interactions between two populations of organisms that did not interact directly with one another. The items probing this level were more cognitively complex than band 1 because two or more entities are involved (e.g., two species of plants and soil nutrients) and more abstract (students predict what could happen, rather than observing a stated relationship).

The items with the highest difficulty estimates (in band 3) described the relationships between three or more populations of organisms interacting indirectly or competing for resources. Notably, items that required students to reason about interactions with populations of microscopic organisms were very difficult for students. Thus, level 3 of the learning progression is the only level that includes reasoning with microscopic populations. Level 3, the upper anchor, might be considered a context-dependent "upper reach" [9] in that it represents the most cognitively complex reasoning, yet it is not a dogmatic scientific idea that students need to hold to be considered "sophisticated thinkers" in this discipline. This "upper reach" specifies that one of the most difficult ideas to reason about in ecology is that there are endlessly complex interactions in ecosystems, yet one can begin to predict downstream effects using relationships in a food web and observations about nature. Students at

the middle school level are just beginning to move beyond the concrete and be capable of abstraction [61]; thus, this concept is a true "upper reach" for most students.

In our final learning progression, the levels roughly correspond to the grade bands described in the NGSS, with level 1 similar to grades 3–5, level 2 similar to grades 6–8, and level 3 similar to grades 9–12. However, the learning progression described here goes beyond stating the facts that students should know, instead describing the kind of reasoning that students with different levels of understanding can perform. In addition, we describe how reasoning about some kinds of interactions is more difficult than others. For example, students find it more difficult to reason about relationships between organisms that are in indirect competition for resources than for organisms in a chain of predation. We also show that it is more difficult for students to reason about plants and microbes than animals. Thus, this learning progression provides additional detail that is useful for teachers about how these levels differ based on the difficulty for student understanding.

4. Discussion

Knowledge of ecosystems is necessary to understand critical challenges facing humanity such as climate change, pandemics, and the threat of catastrophic biodiversity collapse. Ecology is represented by a set of disciplinary core ideas within the life sciences in the Next Generation Science Standards. Given the prominence afforded the life sciences in school science curricula, and ecology's central place in the biology curriculum, nearly every student encounters ecology during their K-12 education. Today, teachers are increasingly encouraged to engage in formative assessment practices to create learning experiences that are responsive to students' current understandings and help them move forward [62]. Learning progressions have been identified as useful supports for teacher formative assessment practice [18]. Yet assessment materials that target a range of levels of understanding are not widely available for classroom use. The learning progression for ecosystems described here and the items used in our research can support teachers' classroom formative assessment practices.

The final, empirically validated progression developed in this work consists of three broad levels, shown in Table 1. Level 1 consists of students' reasoning about direct relationships in ecosystems. This includes the idea that plants are living organisms and form the basis of the food web and that organisms can have relationships with one another such as predator-prey, mutualism, commensalism, and parasitism. Level 2 of the learning progression describes students' reasoning about indirect relationships in ecosystems. Students predict the effects of a change in one population on another population that is not adjacent on the food web. This level also describes students' reasoning about the effects of the availability of and competition for non-living resources (e.g., space, water, and light). Level 3 describes students' reasoning about changes in more than two components in an ecosystem based on changes in microscopic or macroscopic populations or the availability of non-living resources. It is encouraging to see that the progressions from the NGSS [25] are in general supported by the empirical validation of our learning progression.

4.1. Implications for Practice

Prior research has found that learning progressions can be difficult for teachers to use because student understanding does not necessarily follow a linear pattern so neatly described by the learning progression [63]. Students at the highest levels of understanding are thought to display a more coherent, theory-like understanding of complex scientific concepts, while students whose understanding of the topic is emerging demonstrate knowledge that appears fractured. Indeed, it is simplistic to view a learning progression as a map for all students' learning of a concept over time. Furthermore, critics of learning progressions [16,64] argue that learning progressions, as models of learning, fail to account for the social and situated nature of learning with some exceptions [65,66]. Despite these limitations, scholars have suggested that progressions have a place in teaching practice.

Alonzo and Elby [16] suggest that in addition to empirical validation, learning progressions should be viewed with a lens of fruitfulness; is the learning progression useful to teachers?

First, compared to other attempts to describe a learning progression of ecosystem knowledge, the progression described here has a close association with a particular DCI within the NGSS (LS2.A). This situates the progression as a fruitful tool for helping teachers of a middle school NGSS-aligned ecosystems unit understand their students' current understanding of concepts relevant to standards. While the levels of the learning progression are similar to the grade band designations (3–5, 6–8, and 9–12) described in the NGSS for the DCI LS2.A, we found that middle school students in our participating school district had abilities to reason about interactions in ecosystems that spanned all three grade bands. Thus, this learning progression can help teachers understand their students' trajectory through middle school ecology. The focus here on ecosystems content understanding keeps this learning progression manageable in its scope [67] and in our view more useful to middle school teachers.

Second, a learning progression for a DCI facilitates the development of three-dimensional assessments. Since the NGSS were released in 2013, assessment developers have been charged with developing "three-dimensional" assessments that simultaneously measure scientific ideas, practices, and crosscutting concepts. Three-dimensional assessments are useful to operationalize the NGSS. However, researchers and practitioners observe that content knowledge and practice develop together, and it is hard to disentangle the two [68,69]. Indeed, there is a difference between students' ability to engage in the three dimensions simultaneously during instruction and the ability of a multidimensional task to measure the dimensions separately. Yet, within our larger project, we are investigating a method for measuring multiple dimensions of the NGSS by examining the relationship between students' developing understanding of the scientific practice of argumentation within multiple disciplinary contexts. Our approach focuses on developing learning progressions for a single DCI, practice, or crosscutting concept. We develop multi-item tasks around a common phenomenon, with each item specifically targeting a learning progression for a disciplinary idea, scientific practice, or crosscutting concept. This structure allows us to create a multi-dimensional task using a single scenario to separately estimate students' ability on multiple progressions. Thus, student reasoning about interactions in ecosystems and their ability to argue from evidence can each be estimated separately, while only requiring the student to consider one common scenario. So, even as teachers use the three dimensions together during instruction, this model of assessment provides teachers with information about how students understand each individual dimension.

Third and finally, this learning progression is connected to a set of empirically validated and well-studied assessment items that middle school teachers can use (see Supplemental Materials) as formative assessment tools to understand the conceptions of individual students and their class as a whole along the learning progression [10]. Currently, teachers lack assessment resources that effectively measure students' progress toward the NGSS [70,71]. The assessment resources described here are especially practical for classroom use because they are selected response items and have been shown to assess higher -order thinking effectively [26]. Because these items can be computer-scored, teachers quickly receive information about students' understanding of these concepts. This feature reduces the time between assessment and feedback to students or changes in curriculum, which remains a persistent challenge for classroom assessment [72]. Teachers can access the items within the BAS Software (BASS) platform [38], which integrates the BAS with task delivery, scoring, and reporting, as shown in Figure 5. When students complete the online tasks, the BAS Software platform generates eight types of reports that teachers can choose from describing student proficiency with regard to the construct. We will describe two of these reports in detail here. Examples of the other six report types are available in the Supplemental Materials.
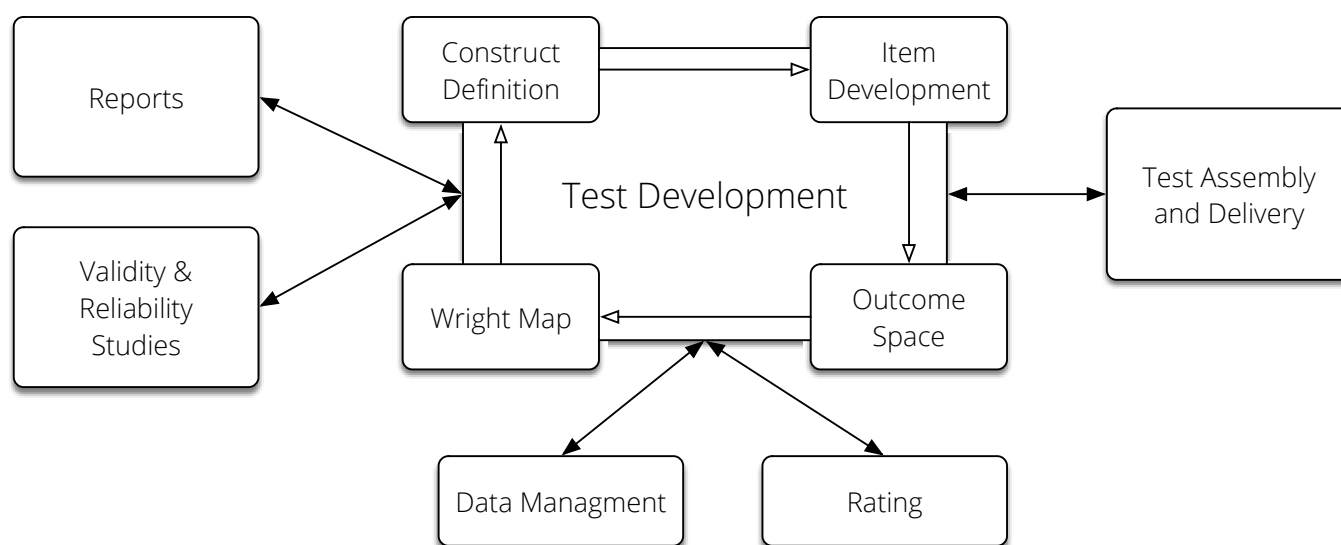
Figure 5. Overview of BEAR Assessment System Software [38].

The Group Proficiency Report shows the ability estimate for each student and the estimated error mapped onto the learning progression. The Group Proficiency Report allows teachers to simultaneously visualize the area where learning is currently occurring for each student in their class. For the example shown in Figure 6 (a fictional teacher and class of students modeled on actual data), the teacher would find that most of their students are actively learning concepts in level 1, such as predator-prey and commensal relationships between two populations, and level 2, interactions between organisms with indirect relationships or competing for abiotic resources. Only 3–5 students are actively learning at level 3, such as making predictions about interactions between 3 or more components in an ecosystem. Similarly, 3–5 students are yet to move into the learning progression, holding naive conceptions about interactions in ecosystems. This hypothetical teacher might use this information formatively to decide whether to continue teaching level 1 material or to move on to thinking about competition and indirect relationships between organisms. Alternatively, the Group Proficiency Report might be used by the teacher to create homogeneous groups and differentiate lesson materials for students based on the level of the progression where they are actively learning. The Group Proficiency Report takes the complicated IRT models used to generate student estimates and translates them into useful tools for teacher use.

Another report type, the Answers Report, displays the frequency with which the different response options were selected for each one of the items in this activity. These reports can be used to identify which incorrect answers are the most attractive to students, which can shed light on how they understand the construct. The example Answers Report shown in Figure 7 shows students' answers based on the same hypothetical teacher and students shown in the Group Proficiency Report (Figure 6). The teacher might look at this report and conclude that all but 7 of their students are able to read and interpret a food web diagram correctly and understand that organisms that are directly connected will be affected by a change in the other species' population. They might check to see if this is true for all direct relationships in the food web, or if it is specific to predator-prey interactions by questioning students or examining other level 1 items in the Answers Report.
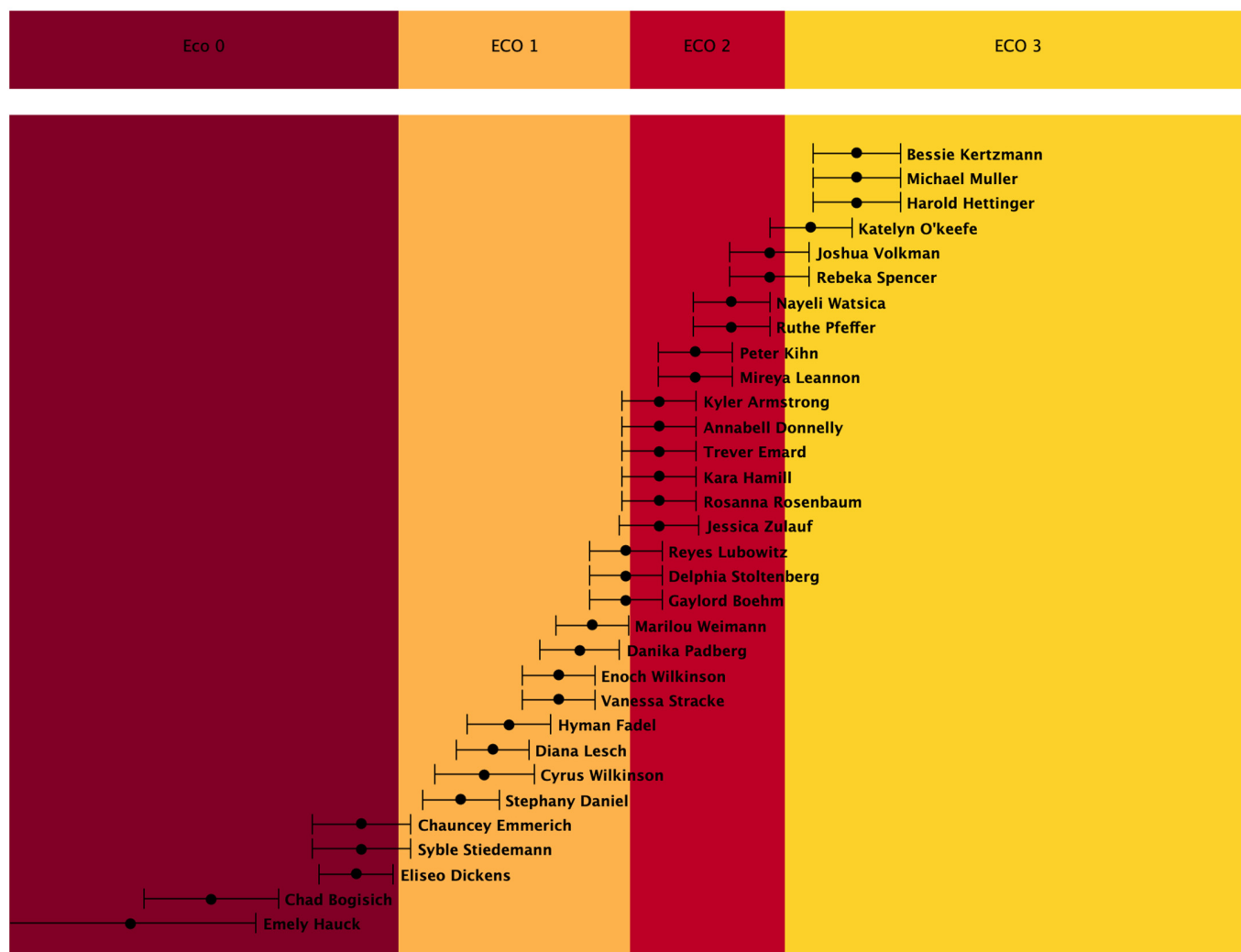
Figure 6. Group Proficiency Report describing students' developing understanding of Interdependence in ecosystems for fictional students. The colors correspond to the levels of the learning progression.

The reporting functions in the BAS System platform play a critical role in translating the empirically validated learning progression into a form that is usable and informative for teachers in their practice. To make use of the information about students' reasoning about interactions in ecosystems, teachers need timely information about how their students understand the construct. The reports available through the BAS Software web interface allow teachers to use this learning progression as a part of their instructional planning and formative assessment practices to support students in making progress toward their understanding of interactions in ecosystems. In short, it offers an important tool to enact the vision of assessment where the feedback is provided rapidly and informed by sound evidence.
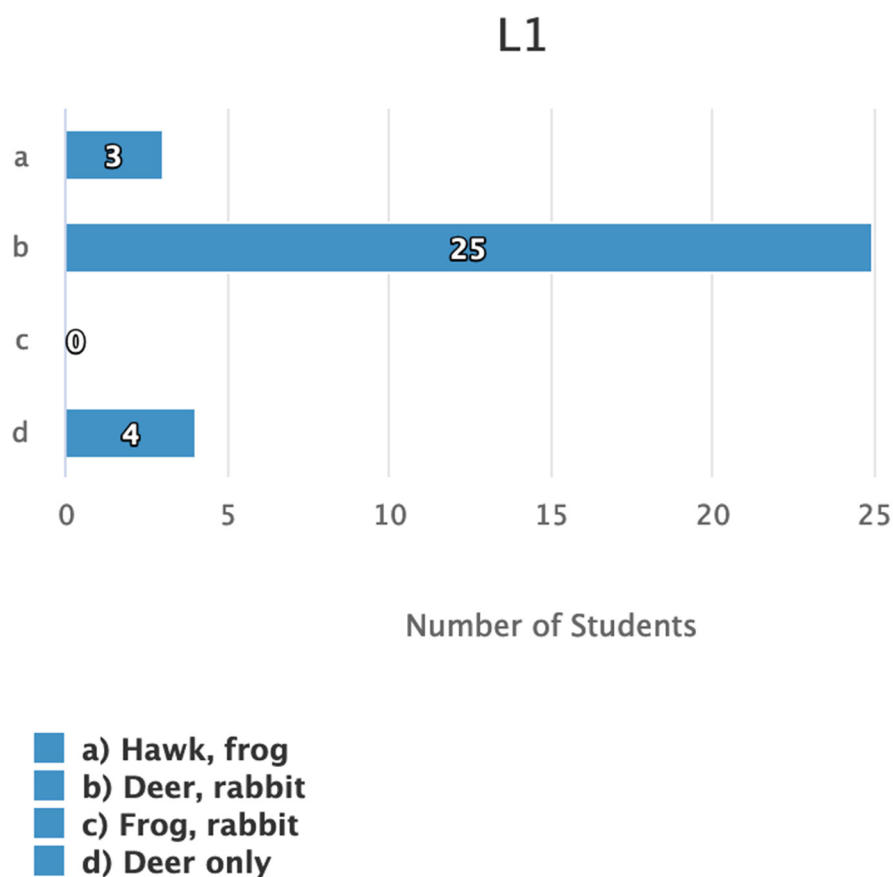
Figure 7. Answers Report for a single item targeting level 1 of the construct map for a fictional class of students.

4.2. Limitations and Implications for Future Research and Theory

Our approach allows students to express their understanding of scientific concepts and provides researchers the opportunity to disentangle students' content understanding about interactions in ecosystems from their proficiency in scientific practices [69]. This study provides an empirically validated learning progression that can be used for the development of additional tasks and items that measure students' understanding of interactions in ecosystems, as well as their ability to engage in the science and engineering practices and understanding of the crosscutting concepts. Our team has already begun the process of developing and testing items within these ecosystems tasks that target a learning progression we developed during an earlier phase of the larger project for the practice of arguing from evidence [27]. In addition, we are developing and validating learning progressions for the crosscutting concept of patterns and the DCI of natural resources using these ecosystems tasks. Rather than creating a three-dimensional learning progression that assumes student ability will progress similarly across all dimensions of the NGSS, we develop tasks based on a single phenomenon with sets of items that target each of the three dimensions, allowing us to disentangle students' abilities in each of these dimensions.

Yao et al. [69] showed that although content knowledge is correlated with the ability to engage in scientific argumentation, each is a distinct construct. An outstanding question is whether the ability to engage in practices or use crosscutting concepts transfers from one scientific content area to another. For example, if a student has a high degree of proficiency in argumentation about interactions in ecosystems, will they have a similarly high ability in argumentation related to chemistry? The tasks described here provide a unique kind of instrument for answering this question.

One surprising result from this study was the differential item functioning for two items that suggested male-identified students had more difficulty than female-identified

students in identifying microbes, insects, birds, and small mammals as living organisms. Prior studies have documented gender differences in whether students identify plants as living or non-living [58–60]. However, these items did not ask students to distinguish plants as living, but rather whether different types of animals and microorganisms were living or non-living. Further research is needed to understand gender differences in how learners distinguish living from non-living components of ecosystems.

This study presents a hypothesized learning progression supported by empirical results from 1366 students. While broad coverage of the Wright Map suggests that the current item set was sufficient to identify the separation between the hypothesized levels, another area for future work is in the development of level 3 items. Because higher-order items are necessarily complex, it is difficult to avoid introducing construct-irrelevant variance because of language demands such as challenging syntax. In addition, it is difficult to write a distractor that is both plausible and incorrect for items that target extremely complex phenomena [73]. Because interpretations about complex ecological phenomena are often subtle, most of the distractors describe possibilities that are not dramatically weaker than the best answer. Thus, it is difficult to write distractors that are substantially different from the best answer but not obviously inaccurate. That is not to say that the development of items that target the highest levels of the learning progression is impossible, only that it is time and resource-intensive, requiring extensive think-aloud interviews and analysis of constructed responses to similar items. Despite these challenges, the development and validation of additional items targeting level 3 of this learning progression would be a useful addition to this work.

Another limitation of the work is that the development of progression was completed in tandem with the development of items. We assert that this is beneficial for both the research and practitioner communities because it creates a learning progression grounded in the realities of teaching and learning that is well-situated to be useful to teachers with tasks and items aligned to the progression available for immediate use in classrooms. The practical demands of this process require that the usability of the items for making inferences about student understanding be considered at all stages of the research, including during model construction. As a consequence, we chose to remove items that performed poorly during student think-aloud interviews or had misfitting item parameters. In all cases, the items were examined prior to removal to understand what might have caused the poor performance. Removing items that do not perform well has the potential to change the underlying model and present a cleaner model than if the items were retained. However, we argue that retaining those items would increase the likelihood that inferences about student understanding and the increasing cognitive complexity of the construct levels would be incorrect [38].

5. Conclusions

In this study, we developed and investigated the qualities of a learning progression about how middle school and high school students understand interdependent relationships in ecosystems. As previous research has found, students' understanding likely does not proceed along a simple, linear path. For example, it appears to be an oversimplification to hypothesize that as the interactions grow more complex, they are necessarily harder to understand. We found that the complexity arising from thinking about changing populations of organisms was not appreciably more difficult than considering the relationship between individual organisms for our middle and high school students. However, increasing the number of populations and nonliving factors interacting in a scenario did make items significantly more difficult. Furthermore, items about populations with predator-prey relationships were less difficult to reason about than populations competing for resources.

The contribution of this work is a practical and useful learning progression that middle school teachers can use to think about the way their students might develop an understanding of ecosystems. Our work builds on prior research on learning progressions in ecology [20,29]. In prior studies, researchers developed learning progressions that

include DCIs and one or two different dimensions of the NGSS (i.e., science and engineering practice, crosscutting concept). The approach used in the study described here allows us to investigate student understanding of a DCI independent of their ability to engage in a science and engineering practice while creating space within tasks for the inclusion of additional items targeting the science and engineering practice or crosscutting concept. Moreover, the model described here for developing a learning progression and assessment tasks in tandem, through the rigorous examination of empirical evidence, can be used by other researchers to develop tools that support the teaching and learning of science. Each band is associated with items that teachers can use to probe students' understanding and the assessment items, which underwent extensive testing, are available for immediate use in teachers' formative assessment practice (see Supplementary Materials). Thus, we created sound formative assessment materials that can be used immediately in biology classrooms across the country. In this way, the materials presented here have the potential to make an immediate impact on how students learn about ecosystems and how teachers support student learning about this important topic in class. While knowledge of ecosystems is only one step toward a more integrated view of ecological sustainability, it is surely a prerequisite. Without understanding how ecosystems achieve stability, it is impossible to understand how we might create sustainable solutions that might restore and maintain such stability in the future. Our approach provides a model for understanding how students learn about ecosystems coupled with practical tools that can support teachers in the classroom by disentangling students' understanding of interactions in ecosystems from the science and engineering practices. This disentangling provides actionable information that can be used to propel student understanding to higher degrees of cognitive complexity and a deeper knowledge of ecosystems.

References

1. Berkes, F.; Colding, J.; Folke, C. Rediscovery of Traditional Ecological Knowledge as Adaptive Management. *Ecol. Appl.* **2000**, *10*, 1251–1262. [CrossRef]
2. Hoagland, S.J. Integrating Traditional Ecological Knowledge with Western Science for Optimal Natural Resource Management. *IK Other Ways Knowing* **2017**, *3*, 1–15.
3. Merritt, E.G.; Bowers, N. Missed Opportunities for Observation-Based Ecology in the Next Generation Science Standards. *Sci. Educ.* **2020**, *104*, 619–640. [CrossRef]
4. Egerton, F.N. A History of the Ecological Sciences: Early Greek Origins. *Bull. Ecol. Soc. Am.* **2001**, *82*, 93–97.
5. Banilower, E.R.; Smith, P.S.; Malzahn, K.A.; Plumley, C.L.; Gordon, E.M.; Hayes, M.L. *Report of the 2018 NSSME+*; Horizon Research, Inc.: Chapel Hill, NC, USA, 2018.
6. Scott, E.E.; Wenderoth, M.P.; Doherty, J.H. Learning Progressions: An Empirically Grounded, Learner-Centered Framework to Guide Biology Instruction. *CBE Life Sci. Educ.* **2019**, *18*, es5. [CrossRef]
7. Duschl, R.A.; Maeng, S.; Sezen, A. Learning Progressions and Teaching Sequences: A Review and Analysis. *Stud. Sci. Educ.* **2011**, *47*, 123–182. [CrossRef]
8. National Research Council. *Taking Science to School: Learning and Teaching Science in Grades K-8*; National Academies Press: Washington, DC, USA, 2007.
9. Sikorski, T.-R. Context-Dependent "Upper Anchors" for Learning Progressions. *Sci. Educ.* **2019**, *28*, 957–981. [CrossRef]
10. Furtak, E.M.; Morrison, D.; Kroog, H. Investigating the Link between Learning Progressions and Classroom Assessment. *Sci. Educ.* **2014**, *98*, 640–673. [CrossRef]
11. Baxter, G.P.; Glaser, R. Investigating the Cognitive Complexity of Science Assessments. *Educ. Meas. Issues Pract.* **1998**, *17*, 37–45. [CrossRef]
12. Alonzo, A.C.; Wooten, M.M.; Christensen, J. Learning Progressions as a Simplified Model: Examining Teachers' Reported Uses to Inform Classroom Assessment Practices. *Sci. Educ.* **2022**, *106*, 852–889. [CrossRef]
13. Rasch, G. *Probabilistic Models for Some Intelligence and Attainment Tests*; Expanded Ed.; University of Chicago Press: Chicago, IL, USA, 1980.
14. Corcoran, T.B.; Mosher, F.A.; Rogat, A. *Learning Progressions in Science: An Evidence-Based Approach to Reform*; CPRE Research Reports; Columbia University: New York, NY, USA, 2009.
15. Wiser, M.; Smith, C.L.; Doubler, S. Learning Progressions as Tools. In *Learning Progressions in Science: Current Challenges and Future Directions*; Alonzo, A.C., Gotwals, A.W., Eds.; Brill ǀ Sense: Rotterdam, The Netherlands, 2012; pp. 359–403. ISBN 9789460918247.
16. Alonzo, A.C.; Elby, A. Beyond Empirical Adequacy: Learning Progressions as Models and Their Value for Teachers. *Cogn. Instr.* **2019**, *37*, 1–37. [CrossRef]
17. Furtak, E.M. Linking a Learning Progression for Natural Selection to Teachers' Enactment of Formative Assessment. *J. Res. Sci. Teach.* **2012**, *49*, 1181–1210. [CrossRef]
18. Shepard, L.A. Learning Progressions as Tools for Assessment and Learning. *Appl. Meas. Educ.* **2018**, *31*, 165–174. [CrossRef]
19. Gotwals, A.W.; Songer, N.B. Validity Evidence for Learning Progression-Based Assessment Items That Fuse Core Disciplinary Ideas and Science Practices. *J. Res. Sci. Teach.* **2013**, *50*, 597–626. [CrossRef]
20. Gotwals, A.W.; Songer, N.B. Reasoning up and down a Food Chain: Using an Assessment Framework to Investigate Students' Middle Knowledge. *Sci. Educ.* **2010**, *94*, 259–281. [CrossRef]
21. Hokayem, H.; Gotwals, A.W. Early Elementary Students' Understanding of Complex Ecosystems: A Learning Progression Approach. *J. Res. Sci. Teach.* **2016**, *53*, 1524–1545. [CrossRef]
22. Hovardas, T. A Learning Progression Should Address Regression: Insights from Developing Non-Linear Reasoning in Ecology. *J. Res. Sci. Teach.* **2016**, *53*, 1447–1470. [CrossRef]
23. Songer, N.B.; Kelcey, B.; Gotwals, A.W. How and When Does Complex Reasoning Occur? Empirically Driven Development of a Learning Progression Focused on Complex Reasoning about Biodiversity. *J. Res. Sci. Teach.* **2009**, *46*, 610–631. [CrossRef]
24. Board on Science Education; Division of Behavioral and Social Sciences and Education; Committee on Conceptual Framework for the New K-12 Science Education Standards; National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; Quinn, H., Schweingruber, H.A., Keller, T., Eds.; National Academies Press: Washington, DC, USA, 2012; ISBN 9780309217422.
25. NGSS Lead States. *Next Generation Science Standards: For States, By States*; The National Academies Press: Washington, DC, USA, 2013.
26. Morell, L.; Suksiri, W.; Dozier, S.; Osborne, J.; Wilson, M. An Exploration of Selected-Response Items Compared to Constructed-Response Item Types in Science Education. In Proceedings of the National Council on Measurement in Education Conference, Online, 9–11 September 2020.
27. Osborne, J.F.; Henderson, J.B.; MacPherson, A.C.; Szu, E.; Wild, A.; Yao, S.Y. The Development and Validation of a Learning Progression for Argumentation in Science. *J. Res. Sci. Teach.* **2016**, *53*, 821–846. [CrossRef]
28. Jin, H.; Shin, H.J.; Hokayem, H.; Qureshi, F.; Jenkins, T. Secondary Students' Understanding of Ecosystems: A Learning Progression Approach. *Int. J. Sci. Math. Educ.* **2017**, *17*, 217–235. [CrossRef]
29. Jin, H.; Mikeska, J.N.; Hokayem, H.; Mavronikolas, E. Toward Coherence in Curriculum, Instruction, and Assessment: A Review of Learning Progression Literature. *Sci. Educ.* **2019**, *103*, 1206–1234. [CrossRef]

30. Duncan, R.G.; Gotwals, A.W. A Tale of Two Progressions: On the Benefits of Careful Comparisons. *Sci. Educ.* **2015**, *99*, 410–416. [CrossRef]
31. Chandler, M.J. The Development of Dynamic System Reasoning. *Hum. Dev.* **1992**, *35*, 121–137. [CrossRef]
32. Eilam, B. System Thinking and Feeding Relations: Learning with a Live Ecosystem Model. *Instr. Sci.* **2012**, *40*, 213–239. [CrossRef]
33. Hogan, K. Assessing Students Systems Reasoning in Ecology. *J. Biol. Educ.* **2000**, *35*, 22–28. [CrossRef]
34. Hokayem, H.; Ma, J.; Jin, H. A Learning Progression for Feedback Loop Reasoning at Lower Elementary Level. *J. Biol. Educ.* **2015**, *49*, 246–260. [CrossRef]
35. Mambrey, S.; Schreiber, N.; Schmiemann, P. Young Students' Reasoning about Ecosystems: The Role of Systems Thinking, Knowledge, Conceptions, and Representation. *Res. Sci. Educ.* **2022**, *52*, 79–98. [CrossRef]
36. American Educational Research Association; American Psychological Association; National Council on Measurement in Education. *Standards for Educational and Psychological Testing*; American Education Research Association: Washington, DC, USA, 2014.
37. National Research Council. *Developing Assessments for the Next Generation Science Standards*; The National Academies Press: Washington, DC, USA, 2014.
38. Wilson, M. *Constructing Measures*, 2nd ed.; Routledge: New York, NY, USA, 2023.
39. Mislevy, R.J.; Almond, R.G.; Lukas, J.F. *A Brief Introduction to Evidence-Centered Design*; Educational Testing Service: Princeton, NJ, USA, 2003.
40. Mislevy, R.J.; Steinberg, L.S.; Almond, R.G. On the Structure of Educational Assessments. *Measurement* **2003**, *1*, 3–62.
41. Pellegrino, W.; Chudowsky, N.; Glaser, R. *Knowing What Students Know: The Science and Design of Educational Assessment*; National Academies Press: Washington, DC, USA, 2001.
42. MacPherson, A.C. A Comparison of Scientists' Arguments and School Argumentation Tasks. *Sci. Educ.* **2016**, *100*, 1062–1091. [CrossRef]
43. American Association for the Advancement of Science. *Atlas of Science Literacy*; National Academies Press: Washington, DC, USA, 2001; Volume 1.
44. Shea, N.A.; Duncan, R.G. From Theory to Data: The Process of Refining Learning Progressions. *J. Learn. Sci.* **2013**, *22*, 7–32. [CrossRef]
45. Hayes, M.L.; Plumley, C.L.; Smith, P.S.; Esch, R.K. *A Review of the Research Literature on Teaching about Interdependent Relationships in Ecosystems to Elementary Students*; Horizon Research: Chapel Hill, NC, USA, 2017.
46. Parker, A.; Tritter, J. Focus Group Method and Methodology: Current Practice and Recent Debate. *Int. J. Res. Method Educ.* **2006**, *29*, 23–37. [CrossRef]
47. Duncker, K. On Problem-Solving. *Psychol. Monogr.* **1945**, *58*, 1–113. [CrossRef]
48. Ericsson, A.K.; Simon, H.A. *Protocol Analysis: Verbal Reports as Data*; Revised edition; MIT Press: Cambridge, MA, USA, 1993.
49. Birks, M.; Chapman, Y.; Francis, K. Memoing in Qualitative Research: Probing Data and Processes. *J. Res. Nurs.* **2008**, *13*, 68–75. [CrossRef]
50. Saldaña, J.; Omasta, M. *Qualitative Research: Analyzing Life*; SAGE Publications: Thousand Oaks, CA, USA, 2017; ISBN 9781506305493.
51. De Ayala, R.J. *The Theory and Practice of Item Response Theory*; Guilford Press: New York, NY, USA, 2009.
52. Wright, B.D.; Masters, G.N. *Rating Scale Analysis*; MESA Press: Chicago, IL, USA, 1982; ISBN 9780941938013.
53. Mari, L.; Wilson, M.; Maul, A. *Measurement Across the Sciences: Developing a Shared Concept System for Measurement*, 2nd ed.; Springer: New York, NY, USA, 2023; ISBN 9783031224508.
54. Adams, R.J.; Wu, M.; Wilson, M. *ConQuest 5.0*; ACER: Hawthorn, Australia, 2021.
55. Paek, I. *Three Statistical Testing Procedures in Logistic Regression: Their Performance in Differential Item Functioning (DIF) Investigation*; ETS: Princeton, NJ, USA, 2009; Volume 2009, pp. 1–29.
56. Lord, F.M.; Novick, M.R. *Statistical Theories of Mental Test Scores*; Addison-Wesley: Reading, MA, USA, 1968.
57. Field, A. *Discovering Statistics Using SPSS*, 3rd ed.; SAGE: London, UK, 2009; ISBN 9781847879066.
58. Amprazis, A.; Papadopoulou, P.; Malandrakis, G. Plant Blindness and Children's Recognition of Plants as Living Things: A Research in the Primary Schools Context. *J. Biol. Educ.* **2021**, *55*, 139–154. [CrossRef]
59. Melis, C.; Wold, P.A.; Billing, A.M.; Bjørgen, K.; Moe, B. Kindergarten Children's Perception about the Ecological Roles of Living Organisms. *Sustainability* **2020**, *12*, 9565. [CrossRef]
60. Wandersee, J.H.; Schussler, E.E. Preventing Plant Blindness. *Am. Biol. Teach.* **1999**, *61*, 84–86. [CrossRef]
61. Lehalle, H. Cognitive Development in Adolescence: Thinking Freed from Concrete Constraints. In *Handbook of Adolescent Development*; Jackson, S., Ed.; Psychology Press: Seattle, WA, USA, 2007; pp. 71–89. ISBN 9781841692005.
62. Shepard, L.A. Classroom Assessment to Support Teaching and Learning. *Ann. Am. Acad. Pol. Soc. Sci.* **2019**, *683*, 183–200. [CrossRef]
63. Shavelson, R.J.; Kurpius, A. *Reflections on Learning Progressions*; Alonzo, A.C., Gotwals, A.W., Eds.; Sense Publishers: Rotterdam, The Netherlands; Boston, MA, USA; Taipei, Taiwan, 2012; pp. 13–26.
64. Hammer, D.; Sikorski, T.-R. Implications of Complexity for Research on Learning Progressions. *Sci. Educ.* **2015**, *99*, 242–431. [CrossRef]
65. Gunckel, K.L.; Covitt, B.A.; Salinas, I.; Anderson, C.W. A Learning Progression for Water in Socio-Ecological Systems. *J. Res. Sci. Teach.* **2012**, *49*, 843–868. [CrossRef]

66. Covitt, B.A.; Gunckel, K.L.; Caplan, B.; Syswerda, S. Teachers' Use of Learning Progression-Based Formative Assessment in Water Instruction. *Appl. Meas. Educ.* **2018**, *31*, 128–142. [CrossRef]
67. Gotwals, A.W. Where Are We Now? Learning Progressions and Formative Assessment. *Appl. Meas. Educ.* **2018**, *31*, 157–164. [CrossRef]
68. Debarger, A.H.; Penuel, W.R.; Harris, C.J. *Designing NGSS Assessments to Evaluate the Efficacy of Curriculum Interventions*; Educational Testing Service: Princeton, NJ, USA, 2013.
69. Yao, S.-Y.; Wilson, M.; Henderson, J.B.; Osborne, J.; Henderson, B. Investigating the Function of Content and Argumentation Items in a Science Test: A Multidimensional Approach. *J. Appl. Meas.* **2013**, *16*, 171–192.
70. Wertheim, J.; Osborne, J.F.; Quinn, H.; Pecheone, R.; Schultz, S.; Holthuis, N.; Martin, P. *An Analysis of Existing Science Assessments and the Implications for Developing Assessment Tasks for the NGSS*; Stanford Center for Assessment, Learning, and Equity: Stanford, CA, USA, 2016.
71. Badrinarayan, A.; Wertheim, J. Reconceptualizing Alignment for NGSS Assessments. In Proceedings of the National Association for Research in Science Teaching (NARST) Annual International Conference, Baltimore, MD, USA, 31 March–3 April 2019.
72. Wiggins, G. Seven Keys to Effective Feedback. *Educ. Leadersh.* **2012**, *70*, 10–16.
73. Shin, J.; Guo, Q.; Gierl, M.J. Multiple-Choice Item Distractor Development Using Topic Modeling Approaches. *Front. Psychol.* **2019**, *10*, 825. [CrossRef] [PubMed]