

RESEARCH ARTICLE

Open Access



Exploring the influence of plant and animal item contexts on student response patterns to natural selection multiple choice items

Sara Catherine Heredia^{1*}, Erin Marie Furtak² and Deb Morrison³

Abstract

Background: Research has shown that students have a variety of ideas about natural selection that may be context dependent. Prior analyses of student responses to open-ended evolution items have demonstrated that students apply more core ideas about natural selection when asked about animals, but respond with the same number of naive ideas for plant and animal items. Other research has shown that changing an item to ask about trait loss or gain shifted the types of naive ideas applied by students in their responses. In this paper, we take up both of these findings to determine if differences exist in the types of ideas students apply to similar items with either a plant or an animal in the item stem.

Results: In order to understand if students applied different ideas to plants or animals in distractor-driven multiple-choice questions, we analyzed high school biology students' responses to matched-item pairs. Dichotomous scoring revealed that students chose the correct response more often for the animal items as compared to the plant items. Chi squared analyses revealed significant differences in the distribution of student responses to matched items. For example, more students chose responses that defined animal fitness as related to their strength and plants' fitness related to its longevity.

Conclusions: These results suggest that varied context of plants or animals in item stems on diagnostic assessments can provide teachers with a more complete picture of their students' ideas about natural selection prior to instruction. This is particularly important in assessments used prior to instruction; as teachers will gain greater insight into the variety of ways students think about natural selection across different types of plants and animals.

Keywords: Natural selection, Diagnostic assessment, Distractor-driven multiple-choice items

Introduction

Evolution through the process of natural selection is the unifying concept in the field of biology (Dobzhansky 1973) and is among the disciplinary core ideas in life sciences in the *Next Generation Science Standards* (NGSS Lead States 2013). Unfortunately, studies conducted over the past 25 years have indicated that students have a number of difficulties understanding this concept

(Bishop and Anderson 1990; Gregory 2009; Nehm et al. 2009; Shtulman 2006). For example, students may assign agency to organisms, suggesting that they can intentionally change themselves in response to changes in their environment, or they may confuse the concept of biological fitness with everyday meanings of the term (Bishop and Anderson 1990).

These naive conceptions about natural selection are extremely robust (Ferrari and Chi 1998) and diagnosis of these ideas is particularly important for teachers to support conceptual change in their students (Carey 2000). A number of approaches to this type of 'preassessment'

*Correspondence: scheredi@uncg.edu

¹ University of North Carolina at Greensboro, Greensboro, NC, USA
Full list of author information is available at the end of the article

exist; for example, teachers may use activities and questioning strategies to elicit and respond to student ideas prior to and during instruction (c.f. Furtak 2006). Another approach involves diagnostic assessments, which are generally given to students prior to a unit of instruction and are based on cognitive maps of student understanding of a particular content domain (Treagust 1995). These assessments are designed to provide teachers with information to inform instruction based on an accounting of what their students know and understand.

Although diagnostic assessments may provide teachers with quick and easily interpretable information about student thinking, they also have the potential to generate misleading information about students' ideas. Students' intuitive ideas are generally found to be context-dependent, thus students may respond differently to similar questions when framed in different ways (Hammer et al. 2005). Therefore, the ways in which diagnostic assessment items are written might be more or less likely to draw upon students' intuitive ideas, leading to response patterns that are inconsistent across contexts. Indeed, prior research in biology has indicated that students' varied responses to similar items are dependent upon the context in which the question is framed (e.g. Duncan et al. 2009).

Assessment is, by nature, an inferential process (Bennett 2011), and as such, teachers must make inferences about what students know based upon evidence of their understandings (Pellegrino et al. 2001). This process rests upon the assumption that assessment items elicit valid information about student thinking; however, if student responses vary according to the context in which assessment items are framed, the inferences made about student thinking on the basis of those items are no longer valid. If this is the case, it follows that assessments should span a variety of contexts to support teachers in having ample evidence to make inferences about what students know. In this paper, we explore high school students' response patterns to matched pairs of multiple-choice items about natural selection, varying item context through the use of plants or animals in the item stems. We conclude with implications for diagnostic assessment and teaching of natural selection.

Background

Study context: student ideas about natural selection

Our work is situated in the conceptual domain of the process of evolution by natural selection, or the process by which species of living organisms descend from other species (Darwin 1859). Following the work of prominent biologists studying assessment of students' ideas about natural selection (e.g. Anderson et al. 2002; Nehm and Reilly 2007), we base our work on Mayr's (1982, p.

479–480) five facts and three inferences about natural selection. Mayr's description builds an explanation for natural selection in a stepwise fashion, first establishing the potential of a mated pair of organisms to reproduce exponentially, then addressing the role of the environment in leading to differential survival and reproduction, and ultimately describing the process of speciation over long periods of time. Mayr's (1982) facts and inferences are similar to accounts of natural selection in the *Next Generation Science Standards* (NGSS Lead States 2013).

A considerable amount of research has been conducted into the various ideas that students have prior to, and as a result of, instruction about natural selection. Perhaps the best-documented intuitive ideas of natural selection and evolution have to do with what Shtulman (2006) called transformationist ideas that attribute change to "a single process operating on the 'species' essence" (p. 173). These ideas have been identified in many young children, elementary and secondary students, college students, and adults (Gregory 2009).

Transformationist ideas include teleological or purpose-driven language, such as an organism 'needs' to change, and reflexive language reflecting goal direction or agency, such as an organism 'changed itself;' these ideas have been shown to appear in students' interpretations of many evolutionary phenomena, including that individual differences are minor (Mayr's Fact 4), a trait's heritability depends upon its adaptive value (Fact 5), and differential survival and reproduction are irrelevant to adaptation (Inference 2) (Shtulman 2006). For example, students may think that an animal will adapt to its changing environment by gaining a trait that supports their survival. This idea may represent an idea students have based on their everyday understanding of adapt, such as the way humans make changes in response to variations in the environment like wearing warmer clothes in the winter.

In addition, students may confuse the meaning of the word 'fitness' in everyday situations with biologists' use of the term 'fitness.' Students may activate an understanding of fitness related to exercise and health, rather than a scientific understanding of the term that organisms are fit if they survive to reproduce. For example, a student may apply an intuitive understanding of fitness that an animal is strong when in fact the more successful trait in a particular environment may be small stature or the ability to store large amounts of water. Furthermore, students commonly pick up on the use of the phrase 'survival of the fittest,' equating the idea of fitness in biology with the idea of fitness in its everyday use or, as Anderson et al. (2002) stated, they equate fitness with "strength, speed, intelligence, or longevity" (p. 964).

Similarly, students' intuitive ideas about the role that origin of traits plays in creating variation within a

population can influence their understanding of natural selection (Shtulman and Schulz 2008). Research indicates that students struggle to understand the role of random processes in multiple areas of science (Ferrari and Chi 1998; Odom and Barrow 1995), and commonly attribute a goal or purpose to inanimate objects. In that sense, students are challenged to think about new traits as arising due to multiple, random genetic processes, such as recombination of genes through sexual reproduction, gene shuffling and random mutations in genetic sequences (Garvin-Doxas and Klymkowsky 2008). Part of the challenge to students' understanding is the fact that new mutations are not necessarily beneficial, but may also be deleterious or have no effect on an organism's phenotype (Muller 1932). Given the way that genetic mutations are often portrayed in science fiction and popular culture, students are more likely to state that new traits arise as a result of an organism's needs or in response to environmental changes rather than occurring through random genetic processes (Bishop and Anderson 1990; Dagher and Boujaoude 2005; Geraedts and Boersma 2006; Shtulman 2006).

Diagnostic assessment of student ideas

Given the wide range of ideas students have, some form of pre-assessment is needed to inform teachers' planned instruction for any given group of students. Diagnostic assessments are designed to support teachers in understanding the raw material that they and their students will be working with in the classroom to construct understandings of scientific ideas (Treagust 1995). These assessments are quite different than traditional assessments that are often used to assess students' ability to recall facts.

Many widely used diagnostic assessments have made use of item formats specifically designed to provide information about what students know. Some diagnostic assessments make use of open-ended questions (Opfer et al. 2012), and others use two-tier items, which are multiple-choice assessments that have two sets of responses (Treagust 1995). The first tier includes responses to the question and in the second tier students chose a response that reflects their reasoning for their choice in tier one. Another approach is to use multiple-choice items in which the incorrect responses, or distractors, are linked to students' prior ideas or intuitions. An advantage of this type of multiple choice question, also known as distractor-driven multiple choice (DDMC) (Sadler 1998), is that the teacher gets to know which students have chosen which response, thereby quickly classifying students' understanding into a range of categories that can lead to specific instructional choices rather than classifying students' responses as simply being right or wrong. DDMC

distractors are written on the basis of prior research into student thinking, such as that described above.

The relevance of item context to student ideas

The ways in which student ideas are activated as resources for instruction has been shown to depend on how problems are framed (Hammer et al. 2005). That is, when a student is asked a question within a particular context, they might draw upon one set of prior ideas, but when that context is changed, they may draw on another set of ideas. Duncan et al. (2009) argued that the contexts of assessment tasks could prove problematic for teachers and researchers attempting to identify the level of students' understanding, and that the context of a task or item may influence the way students respond, with students exhibiting higher performance in one context but lower performance in another.

We define item context as the surface features used to frame a question and response choices. Take, for example, the following item about variation from the *Conceptual Inventory of Natural Selection* (Anderson et al. 2002), a commonly used diagnostic assessment for college students' understanding of natural selection: *Populations of lizards are made up of hundreds of individual lizards. Which statement describes how similar they are likely to be to each other?* Anderson et al. (2002) noted that this item is intended to elicit student ideas about variation. While the question is framed in the context of the Canary Island Lizards, it could easily be modified to the context of another organism, such as exchanging lizard for another animal such as a brown bear, or using organisms from another kingdom, such as a plants or bacteria.

Recent studies of student responses to natural selection items found that students gave different responses depending on the context of the item (Nehm and Ha 2011; Nehm et al. 2012; Nettle 2010; Opfer et al. 2012; White and Yamamoto 2011). For example, undergraduate biology students said that organisms lost traits over time because they did not use them or that the organism shifted energy allocation to other more important processes, but did not use those ideas to respond to items about trait gain (Nehm and Ha 2011). Similarly, students provided different responses for items that differed in familiarity of species and if the item referred to a plant or animal; students applied more key concepts about evolutionary change to items that contained familiar and animal items, yet there was no significant difference in the frequency of naive ideas represented with the different contexts (Opfer et al. 2012). Table 1 summarizes research on types of item contrasts.

The studies described in Table 1 demonstrate that both the frequency and type of ideas students applied to similar items shifted when contexts were varied. One

Table 1 Summary of literature on item context and student response to open-ended natural selection items

Authors	Item contrast	Dimension of natural selection	Results
Nehm and Ha (2011)	Scale of change: within or between species changes	Trait gain or loss	Trait gain, as well as within species explanations, had more core ideas than items that asked about trait loss and between species items
Nettle (2010)	Human vs. non-human animal	Process of evolution of a single trait within a population	Human examples elicited more correct ideas about variation. Fewer misunderstandings with human example
Opfer et al. (2012)	Animal vs. plant and familiarity of organism	Evolution of a trait	Respondents used more key concepts to explain evolution for familiar and animal contexts than for unfamiliar and plant items. There was no difference in number of cognitive biases demonstrated with either contrast
White and Yamamoto (2011)	Taxonomic distance	Common ancestry	Negative correlation between taxonomic distance and naive ideas

context that requires further investigation is the contrast of animal or plant in the item stem. While the ACORN assessment (Nehm and Ha 2011; Opfer et al. 2012) contrasted items with plants and animals, analyses of student responses to those items quantified the frequency of intuitive responses rather than identifying differences in types of intuitive responses for plant versus animal items. Nehm and Ha (2011) noted that the plant item elicited more intuitive ideas than the animal item when both asked about trait gain at a larger scale of change, but did not report which ideas each context elicited. Similarly, Opfer et al. (2012) noted that there were differences in student application of ideas to plant and animal items, but did not discuss what those different ideas are for each context.

Children have been found to apply biological ideas differently when considering animals and plants (Anggoro et al. 2008; Carey 1985; Hatano et al. 1993; Leddon et al. 2008; Opfer and Seigler 2004). For example, children tend to not include plants with animals when asked to group things as alive or not alive (Anggoro et al. 2008; Carey 1985; Leddon et al. 2008). Furthermore, children do not associate goal-directed behavior with plants in the same way they do with animals (Opfer and Seigler 2004). This research suggests that students have associated the idea of being alive with the ability to move around, which may have consequences for other concepts in biology that apply to both animals and plants. For example, questions related to origin of traits or fitness may activate the animacy bias because students' intuitive ideas suggest agency and intentional development on the part of the organism to survive; however, they may not apply the same set of intuitive ideas if they do not associate animacy with plants.

Influence of multiple-choice item context on student response patterns

The studies reviewed above illustrate the importance of considering the context of an assessment item when interpreting student responses. While these studies have

identified differences in student response patterns with open-ended items, we do not yet know how item context might influence student responses to ordered multiple-choice items. Furthermore, these studies have identified a number of intuitive responses associated with these contrasts, suggesting that research into the types of ideas students apply to these items is warranted. This is especially true in the case of diagnostic assessments that teachers are using to identify ideas students have prior to instruction. Therefore, to explore the influence of a plant or animal in the item stem on student responses, we designed a study of matched pairs of distractor-driven multiple-choice items for several aspects of student understanding of natural selection. Specifically, we posed the following research questions:

1. How does the use of a plant or animal in the item stem influence the frequency with which students choose responses linked to scientifically accepted ideas?
2. How does the use of a plant or animal in the item stem influence student choice of responses linked to common to intuitive ideas on matched pairs of items?

Methods

Our work in developing a diagnostic assessment for student ideas about natural selection is situated in a long-term study of teachers' everyday assessment practices (see Furtak 2009). As a component of this work, we have sought to create an assessment that could provide teachers with diagnostic information about their students' thinking in an efficient and timely manner. We describe the process of item development, the population of students who responded to these items, and our analytic approach in the sections below.

Item design

We developed the items analyzed in this paper during construction and validation of the Daphne

Assessment of Natural Selection (DANS), which was developed following the Stanford Educational Assessment Laboratory (SEAL) assessment design process (Ayala et al. 2002; Ruiz-Primo et al. 2001). We began with a hypothesized representation of natural selection that sequenced students' increasingly sophisticated ideas about natural selection from naïve to scientifically accepted (Furtak et al. 2014). We then worked to match existing multiple-choice items (Anderson et al. 2002) and a set of open-ended formative assessments authored by our research team (Furtak 2009) to this representation. Then we collected student responses to these assessment items, and conducted think-aloud interviews to better understand student thinking about natural selection and the ways in which the assessment items were eliciting that thinking.

Over time, we modified the representation of natural selection to include both Mayr's (1982, p. 479–480) facts and inferences about natural selection, as well as mechanisms for the origin of variation. We also refined the categories on the representation to better reflect student thinking as measured with assessment items. Across a span of three academic years, this process helped us to iteratively evaluate the way we were representing the construct of natural selection, develop new assessment items, collect more observations, and interpret those observations again. The end product was a set of distractor driven multiple-choice items on the DANS, each of which had one scientifically accepted response and a set of distractors linked to different aspects of student intuitive ideas of natural selection (Furtak et al. 2014).

We worked with a subset of the DANS items in the present analysis, focusing on three aspects of natural selection that observations from the assessment development process suggested would be likely areas for item context to influence student response patterns: fitness, origin of traits, and variation (Table 2). We further disaggregated the origin of traits category into two components: understanding of the genetic mechanisms underlying the origin of traits, and understanding that transformationist explanations of natural selection are not an accurate explanation of change in populations. For each of these four categories we created contrasts for animals versus plants, hypothesizing that students might be more likely to select distractors related to transformationist ideas or to assign agency to organisms when the item was framed in the context of an animal versus a plant. We then wrote generic items for each dimension and then developed two versions of each item with one framed in the context of a plant and the other an animal, working with organisms indigenous to the area in which we conducted the

study so they would be familiar to students. The complete set of matched items can be found in Table 3.

Participants

This study was conducted in three high schools in the same large, suburban school district in the Western US. We administered all eight items to students enrolled in 10th grade biology courses at each high school early in the fall semester prior to their evolution unit. A total of 740 students took the assessment. These students were linguistically, racially, and economically diverse. Only 709 students were used in the analyses that follow because we only included students that answered all eight items.

Analytic approach

We performed two separate analyses of student responses to each of the eight items. The first analysis involved scoring the items dichotomously and comparing the mean score on the animal items to the mean score on the plant items with a paired t test. Although these items were not designed for dichotomous scoring, this created a mechanism to calculate a continuous measure of students' responses that could then be used to statistically compare students' success on each set of items. We calculated Cohen's *d* to calculate an effect size of item context on students' scores.

Next, to explore the proportions of students' responses for each pair of matched items, we performed Chi squared tests of homogeneity. Chi squared analyses compare observed to expected values to see if the distributions are homogeneous, and significant value indicates that the distribution of student choices were significantly different for matched items. Since Chi squared analyses assumes independent observations, we used SPSS software to randomly sample half the students and compared the distribution of their responses to the animal item to the distribution of the other half of the students' responses to the plant item. This resulted in a total of 709 observations, $N = 331$ for the plant items and $N = 378$ for the animal items.

To perform tests of significance on responses to matched items, we calculated the standardized residuals as post hoc analyses on matched items (the difference between the observed and expected values converted to a z score). We used the standard threshold for determining significance with z-scores, with a standardized residual less than -1.96 or greater than 1.96 was determined to be significantly different ($p < 0.05$) from the expected value.

Results

Our analyses indicate that student responses differed significantly depending upon whether a plant or animal was used as the context for the item stem. The average score

Table 2 Linkage of common student ideas to Mayr’s facts and inferences assessed

Mayr’s fact or inference	Dimension	Description
Fact 4	Target idea	<i>New traits arise as a result of random genetic processes</i>
	Mutation-based ideas	Changes occur as a result of genetic mutations in direct response to the environment
		Student refers to mutations or random changes leading to new traits but does not describe a mechanism for how that happens
		Description of differences in traits not given at genetic level
	Target idea	<i>Organisms do not purposefully change their traits in response to the environment</i>
	Transformation-based ideas	Organisms change as a direct result of environmental changes
Theoretically ambiguous description of change		
No mention of transformationist ideas		
Target idea	<i>Individuals within a population vary among themselves, even though they may look generally the same</i>	
Variation-based ideas	Includes idea of variation but use is unclear or vague OR mentions variation but not at the level of population	
	No mention of variation	
Inference 2	Target idea	<i>Individuals in a population are fit if they are able to reproduce the most number of offspring who survive to reproductive age</i>
	Fitness-based ideas	Confusing everyday meaning of fitness with fitness in its biological sense No mention of fitness

on the animal items was 1.85 (*SD* = 1.06), whereas the average score on the plant items was 1.30 (*SD* = 0.90). A paired t test indicated a significant difference between students’ scores on animal versus plant items, $t(708) = 11.06, p < 0.01, d = 0.44$. These results suggest that students chose the correct response more frequently for the animal items than for the plant items.

However, this dichotomous scoring of the items only gives us information on which students chose the scientifically accepted response and does not give us information on which distractor students chose for each item. Chi squared analyses revealed that for three of the four matched pairs there was a significant difference in the distribution of student responses to each item. The distribution of student responses to the fitness, genetic origin of traits, and transformationist items indicated significant differences in how students responded to those items. However, there was not a significant difference in the distribution of student ideas for the matched items that assessed students’ ideas about variation. In the next section, we will present results of the Chi squared analyses to understand if and how students’ responses differed to matched item pairs.

Fitness

Items 1A and 1B (Table 3) asked students to define fitness, and used the context of a plant or an animal. Response choices for these items included statements that reflect the common intuitive ideas that an organism’s fitness is defined by their longevity, strength, or success in obtaining food.

There was a significant difference in the distribution of student responses $\chi^2(3, n = 709) = 41.2, p < 0.001$, and the proportion of students that chose the scientifically accepted answer for these items was similar (34% for the animal and 37% for the plant). Post-hoc analyses indicated a significant difference ($p < 0.05$) in the number of student responses for each of the distractors than would be expected given homogeneous distributions for each of the items. A greater proportion of students chose the answer that provided a definition of fitness as strength for the animal (43%) than for the plant (23%). A greater proportion of students chose the answer that provided a definition of fitness as longevity for the plant (30%) than for the animal (19%). Table 4 summarizes the distribution of student responses, as well as the standardized residuals for each response choice.

Origin of traits: genetic mechanisms

Items 2A and 2B (Table 3) explored students’ understanding of the origin of traits as a result of random genetic processes, and used item contexts of a blue jay and prickly pear cactus. The response choices for these items contained the scientifically accepted response that traits change through random genetic processes and intuitive ideas that relate to the environment causing genetic changes, gene mutations without a specified mechanism for how this change leads to variation in a population, and no genetic basis for change.

There was a significant difference in the distribution of student responses for each item, $\chi^2(3, n = 709) = 26.14, p < 0.001$. Post-hoc analyses showed that three of the four

Table 3 Matched items by surface context contrast

Animal	Plant
<i>Matched pair 1: fitness</i>	
1A. Which characteristic would a biologist find most important in deciding which animals are the “most fit”? The animal	1B. Which characteristic would a biologist find most important in deciding which plants are the “most fit”? The plant
a. that lives the longest	a. with the largest number of seeds that grow into new plants
b. that eats the most food	b. that is the strongest and dominates others in the population
c. with the most number of babies that survive and reproduce	c. that lives the longest
d. that is the strongest and dominates others in the population	d. that has the most leaves
<i>Matched pair 2: origin of trait/genetic mechanism</i>	
2A. Within a population of blue jays, there are differences in beak size. What causes these differences?	2B. Within a population of prickly pear cactus plants, there is variation in the number of spines that keep predators from eating them. What causes these variations?
a. Blue jays have lots of different mutations in their genes	a. Random genetic changes within the population lead to differences in the number of spines
b. Random genetic changes within the population lead to differences in beak size	b. Changes in the predators that eat them lead to random mutations that change the number of spines
c. Changes in the food source lead to random mutations that change beak size	c. Cactus plants have lots of different mutations in their genes
d. Beak size is not controlled by genes	d. Genes do not control the number of spines
<i>Matched pair 3: origin of trait/genetic mechanism</i>	
3A. Although most skunks are black and white, some are grey and others are light-colored. Where did the variations in fur color within a population of skunk most likely come from?	3B. Although most chili peppers are spicy, some are not so spicy and others are very spicy. Where did the variations in spiciness within a population of chili peppers most likely come from?
a. The skunks needed to change in order to survive, so new colors developed	a. The chili peppers needed to change in order to survive, so more spicy chilies developed
b. The skunks changed their color in response to changes in the environment	b. Changes in DNA that happen by chance led to different levels of spiciness
c. Changes in DNA that happen by chance led to new colors	c. The environment caused mutations in the peppers’ DNA that helped them to survive
d. The environment caused mutations in the skunks’ DNA that helped them to survive	d. The chili peppers changed their spiciness in response to changes in the environment
<i>Matched pair 4: variation</i>	
4A. Populations of bighorn sheep are made up of hundreds of individual sheep. These sheep	4B. Populations of yucca plants are made up of hundreds of individual plants. These plants
a. are identical to each other	a. vary in many ways
b. vary in many ways	b. are identical to each other
c. are each unique in every way	c. are each unique in every way

responses contributed significantly ($p < 0.05$) to the difference in the overall distribution of student responses. A greater proportion of the students responded with the scientifically accepted response to the blue jay item (40%) as compared to the prickly pear cactus item (24%). While a greater number of students chose distractors for the prickly pear cactus, the pattern of choices from greatest to least for the two items was similar. The greatest proportion of students chose the distractor that the environment causes genetic changes for both items; however, more students chose this distractor for the cactus item (38%) than for the blue jay item (27%). More students also chose the response that genes do not control traits for the prickly pear cactus (14%) than for the blue jay (8%), although this

was the least popular choice for both items. Table 4 summarizes the distribution of student responses, as well as the standardized residuals for each response choice.

Origin of traits: transformationist ideas

Items 3A and 3B (Table 3) explored students’ understanding of the origin of traits as a result of random genetic processes with distractors tied to transformationist ideas, and used item contexts of a skunk and chili pepper plant. The distractors for those items included that the organism changes itself, the environment changes the organism, and the environment changes the organisms’ genes.

Table 4 Percentage of students' responses compared by plant and animal and standardized residuals of differences

Response choice	Proportion of student responses (%)		Standardized residual responses (%)
	Blue jay	Cactus	
<i>Origin of traits</i>			
Changes in traits arise through random genetic processes	40	24	±2.7*
Changes occur as a result of genetic mutations in direct response to the environment	27	38	±1.8*
Refers to mutations in genes, but not related to changes in traits	24	24	±0.0
No genetic mechanism for trait change	8	14	±1.7*
<i>Variation</i>			
Individuals within a population vary	50	57	±1.0
Individuals within a population are identical to one another	19	17	±0.5
Individuals within a population are all unique	31	26	±0.9
<i>Transformationist Ideas</i>			
Changes in traits arise from random genetic processes	26	44	±2.9*
Organism needed to change in order to survive	11	11	±0.1
Organism changes their traits in response to changes in their environment	40	22	±3.0*
Changes occur as a result of genetic mutations in direct response to the environment	24	23	±0.2
<i>Fitness</i>			
Fitness is an individual's capacity to reproduce	34	37	±0.3
Fitness is defined by how long an organism lives	19	30	±2.1*
Fitness is defined by the strength of the organism	43	23	±3.2*
Fitness is defined by the organism's capacity to obtain food	4	11	±2.4*

**p* < 0.05

There was a significant difference in the distribution in student responses for these items $\chi^2(3, n = 709) = 35.24, p < 0.001$. Two out of the four responses revealed significant (*p* < 0.05) differences in the frequency of student choices for these items in post hoc analyses. A greater proportion of students answered this question with the scientifically accepted response for the skunk (44%) compared to 26% of students choosing that response for the chili pepper. A greater proportion of students chose the response that the organism changed itself in response to a change in the environment for the chili pepper (40%) as compared to the skunk item (22%). Table 4 summarizes the distribution of student responses, as well as the standardized residuals for each response choice.

Discussion

The preceding sections indicate that students responded differently for three of the four matched item pairs, suggesting that the context of the item was influencing their response patterns. Students chose the scientifically accepted response for the animal items as compared to the plant items, and selected different distractors related to the fitness of animals versus plants. In addition, students selected different distractors for items related to plants versus animals with respect to the origin of new traits. In the following section we will discuss how each

of these findings provides diagnostic information about what students know about natural selection prior to instruction, and will then discuss the implications of these findings for diagnostic assessment design and instruction.

Inferences about what students know

The data presented above demonstrate that students responded differently to matched item pairs with varying context for three of the four paired items measured. In this section, we will discuss what types of inferences teachers could make about these data if represented as either dichotomous or disaggregated by student response choice.

As we have illustrated, inferences that are possible when the items are scored dichotomously are limited. Nevertheless, we can conclude that students had a poor understanding of these aspects of natural selection, regardless of whether a plant or animal was used in the item stem. While the difference in their scores was statistically significant, the effect size (*d* = 0.4) reflected only a medium-size effect of item context on students' application of scientific explanations for natural selection. However, students did respond more favorably to items with animals in item stems, so teachers would know that their students understand the concepts better when the items are framed with animals. Teachers could act on this

inference by providing students more examples in class using plants and other organisms, such as bacteria or fungi.

Opfer et al. (2012) found similar findings when they varied test items with plants or animals, with students providing more key evolutionary concepts for animal items than plant items. They attributed this response pattern to students' greater exposure to animals over plants as example organisms in science classes. It is possible that students in our study have also been exposed to more animal examples in class; however, it is also possible that students apply more evolutionary ideas to animals in contrast to plants because they hold biased ideas that only animals are alive. More research is needed on student understanding of natural selection in plants and animals to further explain these findings.

When teachers are provided information on student responses that are disaggregated by student response patterns, there are more inferences that can be made about what students know and can do then from a report of dichotomously scored items. The distribution of student responses allows for a teacher to identify a variety of intuitive ideas students are activating to understand change over time across different organisms. In this way, teachers can identify possible patterns in how students' ideas about plants and animals influence their thinking about evolutionary processes and how they may begin to address those differences in their teaching.

One pattern could be inferred from the similar response pattern of the items that asked about variation among a population of organisms of the same species: half of the students chose the scientifically accepted response that organisms vary within a population for both items, and chose distractors in similar proportions. Therefore, it is possible that students have a coherent set of resources that are activated to respond about variation. Opfer et al. (2012) similarly suggest that the use of similar intuitive ideas reveals a robust cognitive architecture. Ideas about variation within a population have been identified as important indicators of understanding natural selection as a mechanism for evolution (Shtulman and Schulz 2008); as such, it is imperative that teachers address this issue of variation with their students across different types of organisms. Teachers would need to use explicit teaching strategies (Hammer et al. 2005) to work with the ideas that all individuals are unique or identical and demonstrate that, while all individuals carry some similarity, there are slight variations among individuals that are not necessarily unique, but make individuals different from one another.

Students' performance on the fitness items shows an important difference for how students related the concept of fitness to plants versus animals. Students chose

strength to represent fitness for animals and longevity for plants. This reflects a common everyday definition that students have been shown to draw upon when thinking about fitness (Anderson et al. 2002), as well their tendency to attribute human-like deliberate acts to animals (e.g. Ferrari and Chi 1998). Furthermore, this finding is consistent with research that shows children readily categorize plants with animals when asked about whether or not they grow and die (Anggoro et al. 2008; Carey 1985; Leddon et al. 2008), which suggests that an animacy bias is also being activated when students answer questions related to fitness. This difference in response choices also points to the importance that teachers address student ideas about fitness differently with plants and animals.

The student response distributions on the origin of traits items provide an interesting pattern in students' responses to the matched item pairs. For both of these sets of items, students chose the scientifically accepted response more for the animal than the plant item. Taken together, the highest proportion of student responses on both items related plants' change in traits as a direct response to the environment changing. This response pattern makes sense as students likely have seen plants change with the seasons, losing leaves, changing color, flowering and going through cycles of death and regrowth. Teachers can use this intuitive idea expressed by their students to support their understanding of the underlying genetic plasticity that plants have as a trait for successful reproduction in particular environments.

The variability of student responses to matched items provides more information for teachers to use diagnostically to inform their instruction of natural selection. While it is important to note that no assessment can provide teachers with all the information they need to understand what their students know and can do (Pellegrino et al. 2001), assessments can be designed to give teachers more and better information about what their students know to guide their inferences toward instructional action (Bennett 2011).

Implications for diagnostic assessment and instruction

Research on conceptual change in students highlights the importance of teachers understanding the nature of student understanding prior to instruction (e.g. Carey 2000; Hammer et al. 2005; Strike and Posner 1992). The findings of this study indicate the influence of item context on student pretest responses, and as such they indicate the importance of teachers eliciting and building upon the ideas that students bring to learning about natural selection. It follows that teachers should rely upon a variety of instructional practices to learn about the ideas students bring to understanding natural selection, as well as

other scientific concepts, so that they can better attend to student thinking during instruction.

At the same time, our findings highlight the critical role that the item's sample organism may play in influencing students' response choices. Assessments that feature only plant or animal items will result in an incomplete assessment of what students know. Furthermore, our findings suggest that these inconsistencies may indicate differences in the ideas and experiences students draw upon when responding to items about natural selection. As a result, assessments that rely on animals alone for their item contexts—such as the Conceptual Inventory of Natural Selection (Anderson et al. 2002)—may not be sensitive to the entire range of student ideas activated by plant contexts. Therefore, diagnostic assessment developers should construct natural selection assessments with items featuring plants and animals to provide a more complete picture of student understanding of this disciplinary core idea.

While our study was framed around diagnostic assessment, it also has implications for large-scale assessments. In the current K-12 educational climate in which standardized tests are used to make high stakes decisions about students, school funding, and instructional quality, state tests must cover a great deal of content. As a necessity, these tests may feature only a few items about any given concept. Our study indicates that student understanding of natural selection might be over-estimated if all of these items are written about animals or, conversely, student understanding might be under-estimated if all these items are written about plants. Only larger pools of items featuring a variety of animal and plant contexts will provide a full picture of student understanding of natural selection.

Limitations

Given the small set of items with which we worked, we were not able to look for interaction effects that might be attributed to the type of organism used in the item stem. For example, blue jays are able to fly long distances and migrate and as such students may perceive these organisms differently than animals that are not as able to change location. Similarly, students might be more likely to apply intuitive ideas when asked about carnivorous plants, such as the Venus flytrap. Carnivorous plants are able to move and capture other organisms, therefore students may apply an animacy bias to these plants and respond more similar to questions that contain an animal as a target organism. Future studies might more systematically explore the interaction of the organism used in the item stem with the intuitive ideas students draw upon when answering them. In addition, our study explored only a few dimensions of natural selection. Expanded assessments might more comprehensively assess all of

the dimensions of natural selection identified by Mayr (1982) or in other representations of this concept.

Conclusion

Educators should be aware of how different contexts influence which ideas students draw on to explain their understanding of natural selection. Greater efforts should be made on the part of teachers and curriculum developers to integrate examples of evolution across both plants and animals. The canonical examples of *Geospiza fortis* finches on the Galapagos Islands and *Biston betularia* moths during the Industrial Revolution may be illustrative of how natural selection happens in animals, but our results indicate they should be complemented by examples of natural selection in plants. By drawing out student ideas using a range of examples during instruction, educators may learn more about what students know in order to inform instruction. Furthermore, students should be encouraged to develop universal explanations for natural selection that can be applied not only to animals, but also to plants and other organisms.

Authors' contributions

SCH carried out the data analysis and drafted the manuscript. EMF and DM generated and piloted the assessment used in the study and contributed to writing of the manuscript. All authors read and approved the final manuscript.

Author details

¹ University of North Carolina at Greensboro, Greensboro, NC, USA. ² University of Colorado at Boulder, Boulder, CO, USA. ³ University of Washington, Seattle, WA, USA.

Acknowledgement

Funding was provided by National Science Foundation (grant no. 0953375).

Competing interests

The authors declare that they have no competing interests.

Received: 20 November 2015 Accepted: 20 November 2016

Published online: 01 December 2016

References

- Anderson DL, Fisher KM, Norman GJ. Development and evaluation of the conceptual inventory of natural selection. *J Res Sci Teach*. 2002;39(10):952–78.
- Anggoro FK, Waxman SR, Medin DL. Naming practices and the acquisition of key biological concepts: evidence from English and Indonesian. *Psychol Sci*. 2008;19(4):314–9.
- Ayala CC, Yin Y, Shavelson RJ, Vanides J. Investigating the cognitive validity of science performance assessment with think alouds: technical aspects. In: Paper presented at the American Educational Research Association, New Orleans. 2002.
- Bennett RE. Formative assessment: a critical review. *Assess Educ Princ Policy Pract*. 2011;18(1):5–25.
- Bishop BA, Anderson CW. Student conceptions of natural selection and its role in evolution. *J Res Sci Teach*. 1990;27(5):415–27.
- Carey S. *Conceptual change in childhood*. Cambridge: MIT Press; 1985.
- Carey S. Science education as conceptual change. *J Appl Dev Psychol*. 2000;21(1):13–9.
- Dagher ZR, Boujaoude S. Students' perceptions of the nature of evolutionary theory. *Sci Educ*. 2005;89:378–91.

- Darwin C. On the origin of species by means of natural selection. London: John Murray; 1859.
- Dobzhansky T. Nothing in biology makes sense except in the light of evolution. *Am Biol Teach*. 1973;35:125–9.
- Duncan RG, Rogat A, Yarden A. A learning progression for deepening students' understanding of modern genetics across the 5th–12th grades. *J Res Sci Teach*. 2009;46(6):655–74.
- Ferrari M, Chi MTH. The nature of naive explanations of natural selection. *Int J Sci Educ*. 1998;20(10):1231–56.
- Furtak EM. The problem with answers: an exploration of guided scientific inquiry teaching. *Sci Educ*. 2006;90(3):453–67.
- Furtak EM. Toward learning progressions as teacher development tools. In: Learning progressions in science conference, Iowa City, IA. 2009.
- Furtak EM, Morrison D, Kroog H. Investigating the link between learning progressions and classroom assessment. *Sci Educ*. 2014;98(4):640–73.
- Garvin-Doxas K, Klymkowsky M. Understanding randomness and its impact on student learning: lessons learned from building the biology concept inventory (BCI). *Life Sci Educ*. 2008;7(2):227–33.
- Geraedts CL, Boersma KT. Reinventing natural selection. *Int J Sci Educ*. 2006;28(8):843–70.
- Gregory TR. Understanding natural selection: essential concepts and common misconceptions. *Evol Educ Outreach*. 2009;2(2):156–75.
- Hammer D, Elby A, Scherr RE, Redish EF. Resources, framing, and transfer. *Information age*. 2005. p. 1–26.
- Hatano G, Siegler RS, Richards DD, Inagaki K, Stavy R, Wax N. The development of biological knowledge: a multi-national study. *Cogn Dev*. 1993;8(1):47–62. doi:10.1016/0885-2014(93)90004-O.
- Leddon EM, Waxman SR, Medin DL. Unmasking “alive:” children's appreciation of a concept linking all living things. *J Cogn Dev*. 2008;9(4):461–73. doi:10.1080/15248370802678463.
- Mayr E. The growth of biological thought: Diversity, evolution, and inheritance. Cambridge: The Belknap Press of Harvard University Press; 1982.
- Muller HJ. Further studies on the nature and causes of gene mutations. In: Proceedings of the 6th international congress of genetics, vol 1. 1932. p. 213–55.
- Nehm RH, Reilly L. Biology majors' knowledge and misconceptions of natural selection. *Bioscience*. 2007;57(3):263–72.
- Nehm RH, Kim SY, Sheppard K. Academic preparation in biology and advocacy for teaching evolution: biology versus non-biology teachers. *Sci Educ*. 2009;93(6):1122–46.
- Nehm RH, Ha M. Item feature effects in evolution assessment. *J Res Sci Teach*. 2011;48(3):237–56.
- Nehm RH, Beggrow EP, Opfer JE, Ha M. Reasoning about natural selection: diagnosing contextual competency using the ACORNS instrument. *Am Biol Teach*. 2012;74(2):92–8.
- Nettle D. Understanding of evolution may be improved by thinking about people. *Evol Psychol*. 2010;8(2):205–28.
- NGSS Lead States. Next generation science standards: for states, by states. Washington, DC: The National Academies Press; 2013.
- Odom AL, Barrow LH. Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction. *J Res Sci Teach*. 1995;32(1):45–61.
- Opfer JE, Nehm RH, Ha M. Cognitive foundations for science assessment design: knowing what students know about evolution. *J Res Sci Teach*. 2012;49(6):744–77.
- Opfer JE, Siegler RS. Revisiting preschoolers' living things concept: a micro-genetic analysis of conceptual change in basic biology. *Cogn Psychol*. 2004;49(4):301–32. doi:10.1016/j.cogpsych.2004.01.002.
- Pellegrino JW, Chudowsky N, Glaser R, editors. Knowing what students know: the science and design of educational assessment. Washington, DC: National Academies Press; 2001.
- Ruiz-Primo MA, Shavelson RJ, Li M, Schultz SE. On the validity of cognitive interpretations of scores from alternative concept-mapping techniques. *Educ Assess*. 2001;7:99–141.
- Sadler PM. Psychometric models of student conceptions in science: reconciling qualitative studies and distractor-driven assessment instruments. *J Res Sci Teach*. 1998;35(3):265–96.
- Shtulman A. Qualitative differences between naïve and scientific theories of evolution. *Cogn Psychol*. 2006;52:170–94.
- Shtulman A, Schulz L. The relation between essentialist beliefs and evolutionary reasoning. *Cogn Sci*. 2008;32(6):1049–62. doi:10.1080/03640210801897864.
- Strike KA, Posner GJ. A revisionist theory of conceptual change. In: Duschl R, Hamilton R, editors. Philosophy of science, cognitive psychology, and educational theory and practice. Albany: SUNY Press; 1992.
- Treagust DF. Diagnostic assessment of students' science knowledge. In: Glynn SM, Duit R, editors. Learning science in the schools: research reforming practice. Mahwah: Erlbaum Associates; 1995.
- White BT, Yamamoto S. Freshman undergraduate biology students' difficulties with the concept of common ancestry. *Evol Educ Outreach*. 2011;4:680–7.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com
