Supplemental Material

CBE—Life Sciences Education

Walck-Shannon *et al.*

Supplemental Material 1: Semi-structured Interview Example Questions.

Target time: 40 min.
Target n: 10-15 incoming Biocore 2016 students; 10-15 current Biocore 382 students; 10-15 previous Biocore 382 students

“Welcome, and thank you for agreeing to participate in this activity. I am conducting this study as part of my postdoc research. The purpose of this study is to find out how **you** think about within-species variation. I am going to ask you to think about some questions and tasks that you may never have even thought about before. That’s completely fine. I am not at all interested in whether you get the right or wrong answers; the answer is basically irrelevant to my work. What I really need from you is for you to verbalize your thoughts as we go through the various tasks. The more you can talk out loud about your thinking the better it is for me. It will help me understand how you think about these things and how we can use it in education and curriculum development. There is no pressure and no grading here, I just want to hear your thoughts as you explore each task.”

“During the interview, we would like you to ‘think aloud’. This is like talking to yourself, but we will be listening. This can be hard sometimes. Let’s try out an example with a graph (Appendix A). Talk me through what you see? How you connect the words to the graph?”

Example Question	Purpose, **Skill**	Potential Response
PART I: Think aloud and identify language	*Overall purpose: Identify the interviewee’s base language biological variation within a species*	
<u>Main Question 1:</u> “Recently, one of my friends showed me these examples of the same kind of animals that all look really different (Appendix B). Have you ever seen this in your own life? Can you provide a few examples?”	*Determine whether the interviewee can see evidence for variation in the world around them, and whether they can collate associated concepts into a concrete example.* ***Observe variation***	*Human traits (hair color, eye color, height)*
<u>Follow up</u>: Which words come to mind when you think about how this phenomenon when plants or animals of the same type look different (Appendix B)? (If necessary, redirect students to	*Expose the concepts that students associate with variation.* ***Graph/depict variation***	*Genetic differences, different kinds of behaviors, animals in the same species that look different, etc.*

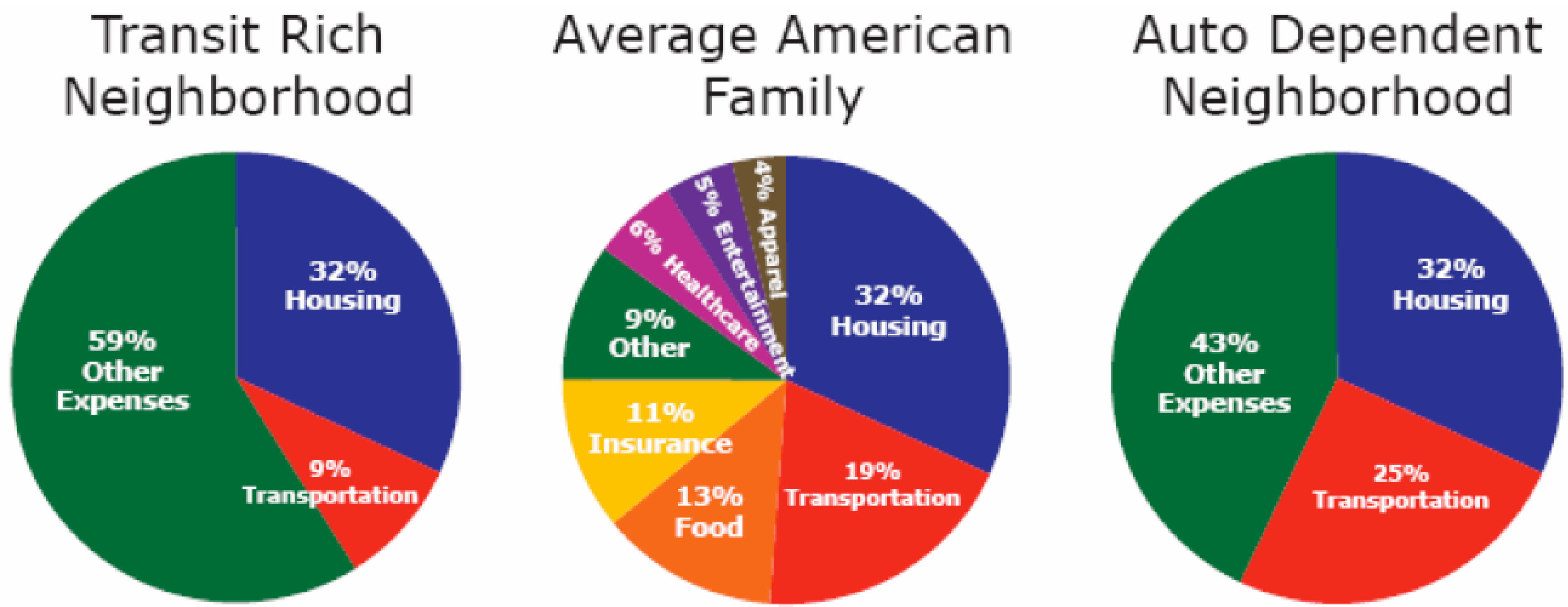
think about a new thought rather than additional concepts related to an earlier one.)		
PART II: Case study – Inquiry of variation.	*Overall purpose: Having all interviewees work on a common question, we can normalize their language and determine if it is consistent with the language used previously. Also, because these examples have not been used in the Biocore curriculum, we can observe how students are able to acclimate to a new problem.*	
"Next I would like to show you some stuffed birds. Please feel free to pick them up [pick one up so that they are not afraid], but do so gently."	*(Transition)* *We will then unveil 10-20 preserved specimens to the interviewee.The interviewee will also have access to a ruler, pencil, and paper.*	
<u>Main Question 2A:</u> "These are all the same kind of bird. What are some of the ways that they look different from each other?"	*Examine students as they observe an example of variation.* ***Observe variation***	*Some discrete traits, like spot number and some continuous traits, like beak color. We will limit students to 3-4.*
<u>Main Question 2B</u>: "Pick one of those differences. How would you measure this difference among the birds?"	*Determine students' methods to measure discrete and continuous variation.* ***Measure variation***	*Students will pick a trait to measure (X).*
<u>Main Question 2C:</u> "Could you put the birds in order from one extreme to another for [trait] 'X'?	*Determine students' ability to compare phenotypes among individuals.* ***Measure variation***	*Students will order the birds from one extreme to the other (for example, least spots to most spots)*
<u>Main Question 3</u>: "If you wanted to express to your friendthis difference in 'X' among the birds without showing them the birds themselves, how would you do it? Make a sketch and talk about your procedure as you carry it out."	*Examine how students depict variation.* ***Graph variation*** *Determine whether students choose to depict biological variation in a statistical way.* ***Statistical variation***	*Histogram, box plot, or bar graph of average with error bars.*
Follow Up: "Ok I think that I understand. Just to make sure, could you pick a different way to display differences in 'X'?"	*To determine whether students are able to form an alternative depiction.*	*Mean points with standard error bars at each temperature.*
<u>Transition Question</u>: "It is just crazy to me to	*Transition: Reset student thinking*	*Mutation, genetic differences, age, sex.*

think that all of these are the same type of bird! Do you have any idea what makes the birds look so different from each other?	*from variation detection to variation etiology*.	
<u>Main Question 4:</u> "If you think about these two individuals [pick up specimens] that differ for 'X' how would you expect the contents of their cellsto compare?"	*To ask if students can identify genetic origins for phenotypic variation.* ***Genetic origins of variation, Predict variation***	*They would have different genes (incorrect use of the word). They would have different alleles of the same genes.*
Follow Up: "In my other interviews, one respondent was confident that several genes were involved in the difference of 'X', while another was confident that only one gene was involved. What do you think? Why would it matter?"	*Knowledge Evaluation: Genetic basis for continuously variable traits.*	*Yes expect more than one gene because these traits are continuous and have many more possibilities than alleles of a single gene could provide.*
Follow Up:"In one of my other interviews, the respondent said that each bird has different genes. Do you agree? Why or why not?" One said gene another said gene variants. Can you help me understand these differences? What do you think?	*Knowledge Evaluation: Genetic terminology.*	*No. They have the same genes but different alleles.*
Follow Up: "Another interview respondent said that there might be fewer spots because the birds were gathered at a different time of the year. Why do you think this may or may not be the case?"	*Knowledge Evaluation:* ***Environmental variation.***	*Yes it is possible that environmental variation is contributing to spot variation.*
"Let's imagine that these birds reproduced and that their chicks grew up to look like this [unveil new set of specimens]."	*We will unveil pictures of a second set of specimens that have less, more, or equal levels of variation relative to the parental generation.*	
<u>Main Question 5A:</u> "How do the differences among the birds compare between the chicks and the parents?"	*To determine if students can detect changes in the amount of variation at the population level.* ***Observe variation***	*The offspring have less variation in spot number.*

Main Question 5B: "Can you make any predictions about which parents mated (or which did not) to give rise to this group of chicks?"	*Ask students to make predictions about the parental contributions towards variation.* ***Interpret variation, Predict outcome based on variation***	*Only a subset of animals (those with fewer spots) reproduced*
Main Question 5C: "What is one plausible scenario that would result in the situation that you see here in a natural environment?"	*Observe students as they reason about causes for changes in variation across generations.* ***Predict context based on variation***	*[Selective pressures] against spots. Perhaps spots make the birds more visible to a predator.*
Main Question 5D: "Can you make any inferences about the differences in [trait X] between birds in the next generation?	*Ask students to predict how variation might be inherited in the next generation.* ***Predict variation in future generations.***	*If the selective pressure continued to be applied, any new variation that arose might be selected against.*
Follow Up: "What if [the selective pressure] no longer existed? What would the next generation look like?"		*The next set of offspring would look like roughly like the parental set if all interbred, though there could be some differences due to genetic drift.*
Follow Up: "Would your answer change if the birds were in an environment [opposite of that described]?"		*If the selective pressure was opposite, new phenotypic variation might be favored.*
"Next I would like to show you a graph that depicts the same kind of animal that looks different."	*Transition* *We will then show the interviewee the graph of eyespot size versus temperature (seasonal dimorphism) (Appendix A).*	*N/A*
Main Question 6A: "Tell me about the relationship that you see in the graph. Ask as many clarifying questions as you would like."	*To observe students as they interpret biological variation based on a graph.* ***Interpret graph of variation***	
Main Question 6B: "At any one temperature, can you give a biological explanation about why the points might vary?"	*To ask if students can interpret the graph and appreciate that genetic differences will yield a range of phenotypes.*	*Genotype*

	Interpret graph, Genetic variation	
Main Question 6D: "Trace one line over the series of temperatures. What do you see?"	*To examine students' abilities to explain environmental origins of variation.* ***Interpret graph, Environmental variation, Phenotypic plasticity***	*For each family, the eyespot index varies at different temperatures. Biologically, this could be caused by changes in gene expression.*
Main Question 6E: " How could this be happening biologically?"		*Epigenetic changes, etc.*
Follow Up: What questions do you have or insights can you share about how environment might be affecting spot size?	*To examine students' interrogation of new data.* ***Question variation***	
PART III: Personal Experiences		
Main Question 7: "You seem to have thought a lot about this. Before this interview, did you ever think about differences within the same kind of plant or animal?	*To ask which experiences students perceive as being most valuable to their intellectual reasoning about variation.*	*Curriculum from previous courses*
Follow up: Can you tell me about experiences helped you answer these questions?" (If necessary, prompt to classes.)		

Appendix A: Think aloud prompt.



Source: Center for TOD Housing + Transportation Affordability Index, 2004 Bureau of Labor Statistics

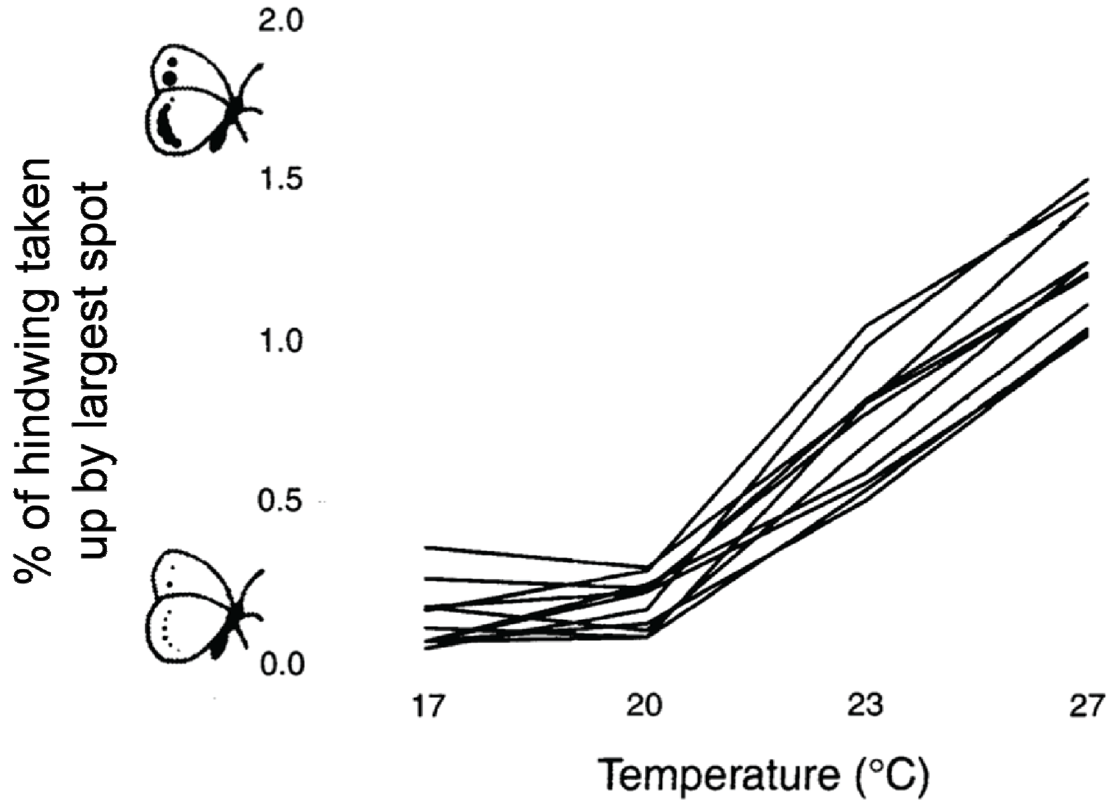
Appendix B: Examples of Within-Species Variation:

http://mar-bengaltigers.blogspot.com/

Westernwildlife.org

Appendix C:

Average genotype morphology across a range of temperatures for the butterfly *B. anynana* (Modified from Brakefield *et al.* 1996).



— Lines represent genetic clones raised at different temperatures

Supplemental Material 2: Below are detailed descriptions of rubric development for the four threshold concept dimensions, which is summarized in Table 1 of the main text:

Discursive Dimension#

<u>First stage coding</u>: In the process of performing the analysis above of all disciplinary words uttered by respondents, we decided that we needed a more specific measure. Therefore, we focused on respondents' language use as they described variation specifically. We used respondents' explanations of the variation in the contents of cells between birds (Question 4) for this analysis, as the stem asked respondents to describe variation itself. Additionally, we supposed that we would see a lot of disciplinary word use in these responses because the question focuses on cellular contents, which are challenging to describe using vernacular terminology. We then took a grounded approach to explore the disciplinary word use in respondents' explanations of variation and found that respondents' descriptions of the cellular differences used terminology on many scales (see Table 1). During this first stage of coding, the number of scales mentioned using disciplinary language within the description were simply counted up. Respondents' responses to question 4 ranged from having 0-4 types of descriptions for variation that used disciplinary language.#

#

<u>Second stage coding</u>: Ultimately, we binned respondents into whether or not they used disciplinary language at any scale to describe the variation that expected within the birds' cells (Table 1).#

#

Troublesome Dimension#

<u>First stage coding</u>: Our rubric for the troublesome dimension was informed by literature. Namely, Land and colleagues describe a threshold concept as troublesome if it contains one of the following attributes: it is counterintuitive, ritualized, inert, foreign, tacit, or conceptually difficult (Land et al. 2005, Land et al. 2010). Therefore, we aimed to look for examples of these attributes within respondent explanations, in particular (1) the overapplication of intuitive knowledge, (2) the ritualized understanding without reasoning, and (3) conceptually inaccuracies, respectively. To help guide our coding of intuitive thinking, we used the categories set forth by Coley and Tanner: teleological (everything has a purpose), essentialist (i.e. everything within a group is the same), and anthropocentric (human-centered) (Coley and Tanner, 2012, 2015). We especially suspected that the essentialist category would exist in our sample due to its prevalence in explanations of evolution from children, undergraduates, and adults (Shtulman and Schulz 2008, Emmons and Kelemen 2015, Richard et al. 2017). For ritualized knowledge and inaccuracies, we used a grounded approach to examine a subset of responses and look for potential examples based on our data, which are summarized in Table 1. During this first stage of coding, the number comments that represented overapplication of intuitive thinking, ritualized knowledge, or inaccuracies were simply counted up. #

#

We focused on the same question (explaining variation in cell contents) for troublesome coding as we did for discursive coding because it prompted students to describe the molecular origins of variation, which has been reported to be especially difficult for students to master (Speth et al. 2015).

__Second stage coding__: Ultimately, we binned respondents based on whether or not they exhibited any of the troublesome attributes described in the rubric above (Table 1).

Liminality Dimension

__First stage coding__: Again, our rubric for the liminal dimension was based on literature. As outlined by Land and colleagues, evidence that the learner is in a liminal space can be observed by oscillation between more than one response; mimicry of understanding; or feelings of discomfort, confusion, anxiety or humility (Land et al. 2010, Meyer et al. 2010). In fact, liminality has been observed using these guidelines in reflective interviews from computer science learners (McCartney et al. 2009). Therefore, we looked for these three types of evidence in our respondents' explanations of variation within the cells of different birds (Question 4), relying mainly on self-reporting directly from the respondent. During this first stage of coding, the number comments that suggested the respondent was in a liminal state were simply counted up. Notably, being tentative to accept a new assertion or wanting further testing (as observed in Halmo et al. 2018) was not coded as being within a liminal space. Instead, liminality was coded for respondents who observably oscillated between responses, or expressed self-proclaimed confusion or uncertainty while responding.

__Second stage coding__: Ultimately, we binned respondents based on whether or not they exhibited any evidence of liminality based on the explanation of variation with cells, as described in the rubric above (Table 1).

Integrative Dimension

__First stage coding__: Our rubric for the integrative dimension was generated using a grounded approach. For this dimension, we also turned to a different question stem that invited more integration. After respondents described the variation of a trait of interest in preserved specimens, they were presented with hypothetical offspring of the specimens. Then, respondents were asked to describe a plausible scenario that could account for their observations between the two generations (Question 5C). Upon reading a subset of these responses, we realized that respondents varied along how many biological scales they integrated into their responses (as discussed in Batzli et al. 2016). Therefore, we generated a list of potential scales based on respondents' descriptions (see Table 1). During this first stage of coding, the number of biological scales that were integrated into a single scenario were simply counted up. Because the stem itself mentioned the respondents' phenotype of interest varying across generations, further mention of phenotypic variation in the scenario did not count as a scale.

__Second stage coding__: For the second stage integrative code, we did not *a priori* know the number of scales that had to be integrated for the respondent to be considered "integrative." Therefore, we looked for a natural break in the data. Ultimately, we categorized respondents as integrative if they brought together two or more biological scales (in addition to phenotype) in their scenario, and non-integrative if they integrated one or no further biological scales (Table 1).

Please see main text for references.

Supplemental Material 3: Tablesdetailing patterns of co-occurrence across threshold concept dimensions. P-values (Fisher Exact tests) associated with each pairwise analysis shown to the right of each table.

	Troublesome	Non-Troublesome
Discursive	13	9
Non-Discursive	10	0

p=0.03

	Pre-Liminal	Liminal	Post-Liminal
Discursive	8	9	5
Non-Discursive	7	3	0

p=0.15

	Non-Discursive	Discursive
Integrative	0	9
Non-Integrative	10	13

p=0.03

	Pre-Liminal	Liminal	Post-Liminal
Integrative	2	5	2
Non-Integrative	13	7	3

p=0.18

	Non-Integrative	Integrative
Non-Troublesome	3	6
Troublesome	20	3

p=0.006

	Pre-Liminal	Liminal	Post-Liminal
Non-Troublesome	0	4	5
Troublesome	15	8	0

p<0.0001