Supplemental Material

CBE—Life Sciences Education

Deane *et al.*

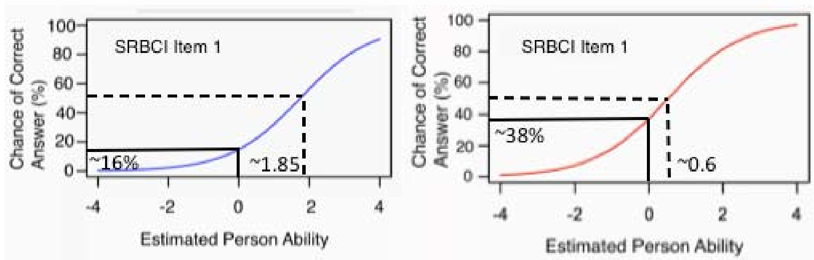

Figure S1: ICC Curves for SRBCI Item 1 in the Biology-first-year level (blue curve, left) and the Biology-third-year level (red curve, right) populations. Two common ways of using ICC curves to provide quantitative information about item difficulty as a function of an individual's ability in the trait being assessed are highlighted by the solid and dashed lines. Researchers can assess: A) the chance (%) of answering an item correctly for a person with a mean ability in the trait being assessed (0 on the x-axis, ~16% and ~38% on the y-axis for the Biology-first-year and Biology-third-year level populations respectively, solid lines); B) the ability of a person in the trait being assessed that is required for that person to have a 50% chance of answering the item correctly (50 on the y-axis, ~1.85 and ~0.6 on the x-axis for the Biology first-year and Biology-third-year level populations respectively, dashed lines).

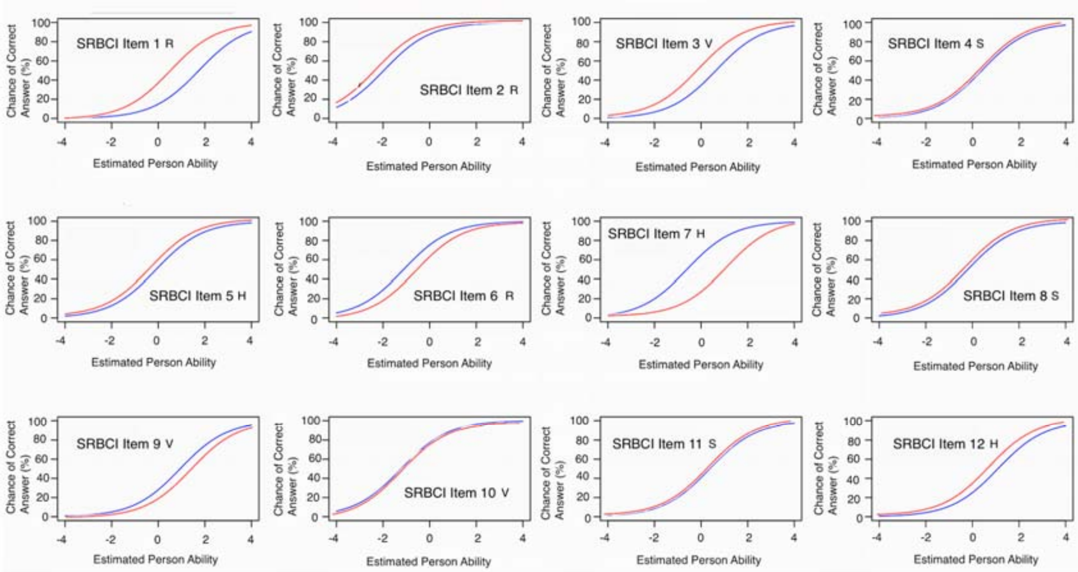

Figure S2: ICC Curves for each of the SRBCI's 12 items, based on responses provided by students in the Biology-first-year (blue) and Biology-third-year (red) populations. Letters after items refer to the conceptual grouping to which these items belong (R = Repeatability of Results, V = Variation in Data, H = Hypotheses and Predictions and S = Sample Size).

Table S1: Student quotes taken from interviews that indicate examples of statistical reasoning students used when choosing different answers to four of SRBCI's items (one from each core conceptual grouping: Repeatability of Results, Variation in Data, Hypotheses and Predictions, and Sample Size). Bold text indicates expert-like reasoning.

Item 1: This item presents bar graphs showing salmon growth (cm) in different treatment groups (different food types given in normal and extra amounts). Students are asked which food they would feed in extra amounts to enhance salmon growth. Only in Group D (Water Fleas) do salmon grow significantly more in the extra food group in both trials, as indicated by non-overlapping 95% confidence intervals. In Group A different results occur in the two trials, in Group B the 95% confidence intervals overlap with wide variation, whereas in Group C the 95% confidence intervals overlap with smaller variation.

Answer	Example Student Quote
A (Worms)	*"… even when there was a lot of difference in the salmon – look at the bars – the extra food group still grew a lot more than the others."*
B (Slugs)	*"I would say slugs because the results were similar between the experiments and the salmon getting extra food grew a lot bigger than the normal ones."*
C (Larvae)	*"The variation was more consistent in all the groups and the salmon getting extra food seemed to be a bit bigger, so I would feed them larvae."*
D (Water Fleas)	***"The bars that show the variation don't overlap between the normal group and the group getting the extra food, and that's not true for any of the other foods."***

Item 5: This item asks students to interpret results with reference to null hypotheses (x will have no effect on y) and alternate hypotheses (x will have an effect on y). Students are shown bar graphs of results from two trials; in both of these, raccoon growth significantly increases when individuals are given more food, as shown by non-overlapping 95% confidence intervals.

Answer	Example Student Quote
A (Fail to reject null hypothesis)	*"… you can never reject a null hypothesis because you can't prove anything in science so it must be this one."*
B (Reject null hypothesis)	***"I would reject my null hypothesis and find support for my alternate hypothesis because the results support that extra food is affecting growth in both the trials."***
C (Prove alternate hypothesis)	*"The alternate hypothesis is right in both of them [the trials] so you have proved it in both… as long as the experiment was done scientifically, which I assume it was…"*
D (Support alternate hypothesis in trial 1 and prove it in trial 2)	*"I would choose this answer… you need to be careful about saying you prove anything in experiments but this is OK because your results were backed up in the second trial."*

Item 8: This item asks students what effect they would expect a larger sample size to have on i) the average variation around sample means (95% confidence intervals) and ii) the range of values contributing to these sample means.

Answer	Example Student Quote
A (Increase, Increase)	*"The average variation around the mean is basically the same as the range so the more animals you get data from, the more that both will increase."*
B (Increase, Decrease)	*"The average variation will increase… the more of anything you measure, the more different it'll be, and then I guess the range would decrease."*
C (Decrease, Increase)	***"The average variation will decrease – that's why you should always get loads of data – but the range would increase as you'd be more likely to get the outliers too."***
D (Decrease, Decrease)	*"I'm pretty sure both would decrease. You should always get as many measurements as possible so that the error bars are smaller, and the range should also be smaller too."*

Item 9: This item provides mean values for sunflower growth (with 95% confidence intervals) for plants grown in three different temperature treatment groups. Students are asked whether temperature affected growth (and which interpretation of the results is the most accurate). The sunflowers grew significantly more in the intermediate temperature group, as indicated by 95% confidence intervals not overlapping with those from the cooler and hotter groups.

Answer	Example Student Quote
A (Yes – sunflowers grew taller in one of the groups)	***"The sunflowers grew significantly taller in one of the groups, so you can say that temperature affects their growth because the bars don't overlap with the others."***
B (No – sunflowers grew taller in the middle group)	*"… no, you can't say temperature has an effect because if it did you'd expect either the cooler or hotter temperature to show the biggest difference."*
C (Yes – mean growth was not the same in each group)	*"I think it's this one… if everything was kept the same apart from temperature but mean growth rates weren't exactly the same then it can only be the temperature causing it."*
D (No – sunflowers grew similarly in the extreme groups)	*"It wouldn't make sense that the temperature is the reason if plants in the coolest and hottest groups didn't differ significantly from each other…"*

Table S2: Examples of statistical reasoning used by students when choosing different answers to the 12 different items (four core conceptual groupings) of SRBCI. Examples help to delineate student reasoning along the novice to expert spectrum.
#

Core Conceptual Grouping	Example Reasoning		
	Novice	**Intermediate**	**Expert**
Repeatability of Results	*> If an experiment is controlled, the results will be reliable (you do not need to repeat it)* *> A significant finding proves the dependent variable has an effect*	*> Replicated patterns in data enhance the confidence of an interpretation*	*> Replicated patterns in data enhance the confidence of an interpretation* *> Replicated patterns do not prove the dependent variable has an effect*
Variation in Data	*> Absolute differences (e.g. means) indicate significant differences between groups* *> Significant differences must be seen across all groups if the dependent variable has an effect* *> Effects should be unidirectional*	*> Average variation around sample means is important in indicating significant differences* *> Effects can be complex (they don't have to be unidirectional)*	*> Average variation around sample means is important in indicating significant differences* *> Significant differences can occur between any/all treatment groups for a dependent variable to have an effect* *> Effects can be complex (they don't have to be unidirectional)*
Hypotheses and Predictions	*> Hypotheses and predictions are the same* *> A hypothesis cannot be supported unless a prediction is also supported by the results*	*> A prediction is not the same thing as a hypothesis*	*> A prediction is not the same thing as a hypothesis* *> A hypothesis can therefore be supported even if a prediction is not*
Sample Size	*> Sample size has no effect on variation around sample means or the range of data values* *> Larger sample sizes increase variation around sample means*	*> Sample size has an effect on variation around sample means* *> This effect is always unpredictable*	*> Sample size has an effect on variation around sample means and on the range of data values* *> Usually, larger samples reduce the average variation (e.g. 95% confidence intervals) around sample means*

#

Table S3a: Item-item correlation comparisons were calculated to check for local independence of all possible item combinations in SRBCI, using data provided by students in the Biology-first-year level population; bold values (> ± 0.20) indicate correlations above the recommended range.

SRBCI Item	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	-	-	-	-	-	-	-	-	-	-	-	-
2	-0.07	-	-	-	-	-	-	-	-	-	-	-
3	-0.01	-0.19	-	-	-	-	-	-	-	-	-	-
4	-0.06	-0.08	-0.04	-	-	-	-	-	-	-	-	-
5	-0.10	-0.13	-0.14	-0.06	-	-	-	-	-	-	-	-
6	-0.11	-0.09	**-0.21**	-0.10	-0.01	-	-	-	-	-	-	-
7	-0.17	-0.13	-0.11	-0.14	-0.01	-0.07	-	-	-	-	-	-
8	-0.01	**-0.23**	0.09	0.01	-0.16	-0.19	-0.14	-	-	-	-	-
9	-0.08	-0.09	-0.03	-0.15	-0.07	-0.14	-0.06	-0.14	-	-	-	-
10	0.03	-0.02	-0.07	-0.16	-0.06	-0.05	-0.18	-0.14	-0.13	-	-	-
11	0.02	-0.12	-0.10	-0.16	-0.19	-0.08	-0.05	-0.01	-0.03	-0.09	-	-
12	-0.10	-0.01	0.02	-0.01	-0.13	-0.19	-0.05	-0.11	0.02	-0.19	-0.09	-

Table S3b: Item-item correlation comparisons were calculated to check for local independence of all possible item combinations in SRBCI, using data provided by students in the Biology-third-year level population; bold values (> ± 0.20) indicate correlations above the recommended range.

SRBCI Item	1	2	3	4	5	6	7	8	9	10	11	12
1	-	-	-	-	-	-	-	-	-	-	-	-
2	-0.19	-	-	-	-	-	-	-	-	-	-	-
3	-0.07	0.05	-	-	-	-	-	-	-	-	-	-
4	-0.03	-0.03	-0.18	-	-	-	-	-	-	-	-	-
5	-0.06	-0.18	-0.10	-0.12	-	-	-	-	-	-	-	-
6	**-0.21**	0.09	-0.15	-0.12	-0.10	-	-	-	-	-	-	-
7	-0.13	-0.03	-0.06	-0.18	-0.06	0.13	-	-	-	-	-	-
8	-0.17	0.15	-0.16	0.14	0.02	-0.16	**-0.21**	-	-	-	-	-
9	0.10	-0.13	0.01	-0.17	-0.18	-0.19	-0.09	-0.08	-	-	-	-
10	-0.09	0.02	-0.04	0.14	-0.09	-0.09	-0.16	-0.16	-0.13	-	-	-
11	-0.10	-0.15	**-0.22**	-0.04	-0.08	-0.04	-0.15	-0.15	-0.02	-0.17	-	-
12	-0.12	-0.10	0.09	-0.19	-0.06	0.04	0.01	0.01	0.13	-0.18	-0.04	-

Table S4: Infit and Outfit MSQ statistics for each of the SRBCI's 12 items, as answered by the two populations of students assessed in these analyses; all values were well within the suggested range of 0.6 – 1.4.

SRBCI Item and Concept Grouping	**Biology-first-year level**		**Biology-third-year level**	
	Infit MSQ	**Outfit MSQ**	**Infit MSQ**	**Outfit MSQ**
1 R	0.877	0.933	1.081	1.201
2 R	0.957	0.962	0.757	0.861
3 V	0.924	0.920	0.934	0.838
4 S	0.972	0.994	1.047	0.986
5 H	0.988	0.964	1.054	1.214
6 R	0.928	0.928	1.010	1.153
7 H	0.995	0.978	0.818	0.832
8 S	0.996	1.006	1.009	1.007
9 V	1.000	1.073	0.998	1.235
10 V	0.894	0.863	0.866	0.825
11 S	0.941	0.937	0.949	0.924
12 H	1.022	1.161	0.858	0.821