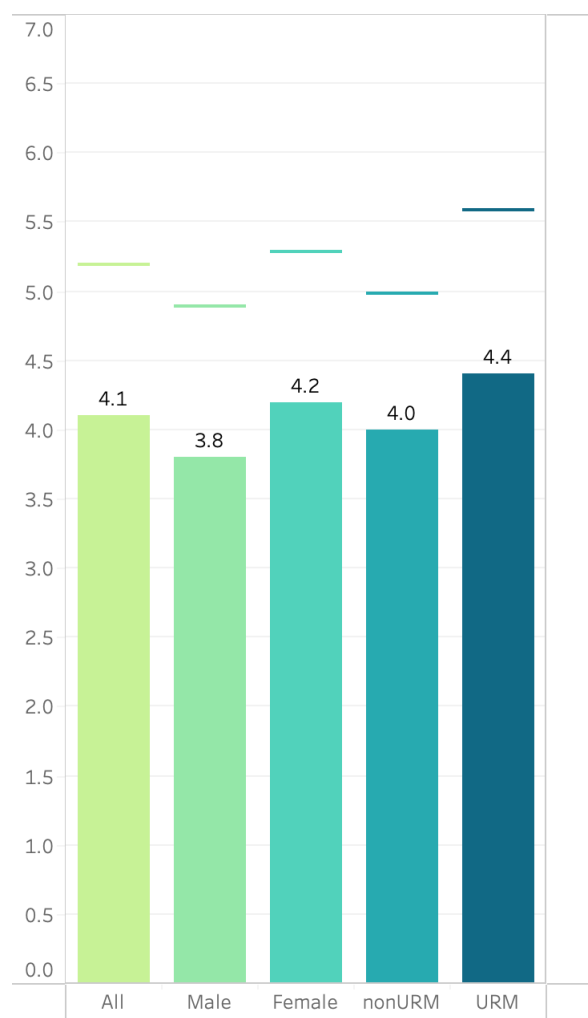


## Supporting Information

The assessments presented in this manuscript are available in the BioMolViz Library at [library.biomolviz.org](http://library.biomolviz.org). The assessment IDs presented here are the last two digits of the Library ID, and each ID is preceded by the numbers 100 (e.g., the item 13 Library ID = 10013). Items are reviewed and modified as needed after each phase of our assessment design and validation process; therefore, some items may appear differently in the Library than they are presented in this manuscript.



**Figure S1.** Performance data disaggregated by gender and race/ethnicity.

**Table S1.** Descriptive and inferential statistics, gender differences in performance (% correct) for students only for all items.

		Males		Females		Independent Samples t-Test
		Mean	SD	Mean	SD	
Students	00	0.81	0.25	0.83	0.22	ns
	01	0.62	0.49	0.60	0.49	ns
	02	0.05	0.22	0.13	0.34	ns
	03	0.71	0.46	0.94	0.25	$t(25) = -2.07, p < 0.05 *$
	04	0.82	0.28	0.77	0.32	ns
	06	0.69	0.48	0.31	0.47	$t(21) = 2.44, p < 0.05 *$
	09	0.33	0.48	0.35	0.48	ns
	10	0.61	0.39	0.33	0.32	$t(33) = 2.73, p < 0.05 *$
	11	0.46	0.52	0.43	0.50	ns
	12	0.77	0.44	0.46	0.51	$t(25) = 2.10, p < 0.05 *$
	13	0.33	0.48	0.47	0.50	ns

14	0.56	0.50	0.45	0.50	ns
15	0.46	0.52	0.51	0.51	ns
16	0.60	0.50	0.44	0.50	ns
<b>Overall % Correct</b>	0.55	0.25	<b>0.50</b>	0.22	ns

Scale: 0.00–1.00 (% correct). \*\*  $p < 0.01$ ; \*  $p < 0.05$ ,  $\pm p < 0.10$  (approaching significance). ns = not significant.

**Table S2.** Descriptive and inferential statistics, race/ethnicity differences in performance (% correct) for students only for all items.

		nonURM		URM		Independent Samples t-Test
		Mean	SD	Mean	SD	
Students	00	0.81	0.25	0.85	0.17	ns
	01	0.61	0.49	0.60	0.50	ns
	02	0.14	0.35	0.00	0.00	$t(50) = 2.82, p < 0.01$ **
	03	0.88	0.33	0.86	0.36	ns
	04	0.80	0.28	0.67	0.40	ns
	06	0.41	0.50	0.44	0.51	ns
	09	0.34	0.48	0.45	0.51	ns
	10	0.49	0.38	0.24	0.24	$t(30) = 2.85, p < 0.01$ **
	11	0.41	0.50	0.56	0.51	ns
	12	0.56	0.50	0.50	0.52	ns
	13	0.39	0.49	0.50	0.52	ns
	14	0.49	0.50	0.48	0.51	ns
	15	0.56	0.50	0.44	0.51	ns
	16	0.58	0.50	0.23	0.44	$t(21) = 2.50, p < 0.05$ *
	<b>Overall % Correct</b>	0.53	0.24	0.49	0.20	ns

Scale: 0.00–1.00 (% correct). \*\*  $p < 0.01$ ; \*  $p < 0.05$ ,  $\pm p < 0.10$  (approaching significance). ns = not significant.

**Table S3.** Descriptive and inferential statistics, difficulty perceptions and mean comparison between instructors and students.

Assessment ID#	Students		Instructors		Independent Samples t-Test	Mean Comparison
	Mean	SD	Mean	SD		
00	4.23	1.19	4.30	1.41	$t(43) = -0.21, ns$	-0.07
01	4.62	1.40	2.59	1.25	$t(42) = 7.44, p < 0.01$ **	2.03
02	2.31	1.46	3.81	1.14	$t(60) = -5.34, p < 0.01$ **	-1.50
03	3.27	1.34	2.59	1.47	$t(44) = 2.08, p < 0.05$ *	0.68
04	3.39	1.54	3.85	1.46	$t(50) = -1.39, ns$	-0.46
06	3.88	1.64	3.56	1.48	$t(58) = 0.89, ns$	0.32
09	4.70	1.29	3.00	1.33	$t(39) = 6.00, p < 0.01$ **	1.70
10	5.23	1.71	3.19	1.27	$t(66) = 6.23, p < 0.01$ **	2.04
11	4.66	1.61	4.00	1.30	$t(64) = 1.95, p = 0.056 \pm$	0.66
12	3.51	1.72	2.22	1.19	$t(70) = 3.84, p < 0.01$ **	1.29
13	3.78	1.40	4.15	1.79	$t(39) = -0.97, ns$	-0.37
14	4.22	1.54	2.63	0.93	$t(60) = 7.14, p < 0.01$ **	1.59
15	4.72	1.76	4.07	1.47	$t(62) = 1.72, p = 0.091 \pm$	0.65
16	3.72	1.44	2.48	1.40	$t(51) = 3.80, p < 0.01$ **	1.24
<b>Overall Perceived Difficulty Average</b>	<b>4.08</b>	<b>1.09</b>	<b>3.32</b>	<b>0.69</b>	<b><math>t(48) = 4.94, p &lt; 0.01</math> **</b>	<b>0.76</b>

A negative value indicates that the students perceive it is easier. Scale: 1, very easy to 7, very difficult.

\*\*  $p < 0.01$ ; \*  $p < 0.05$ ,  $\pm p < 0.10$  (approaching significance). ns = not significant.

**Table S4.** Descriptive and inferential statistics, gender differences in difficulty perceptions for instructors and students.

		Male		Female		Independent Samples t-Test
		Mean	SD	Mean	SD	
Instructors	00	3.80	1.03	4.59	1.54	ns
	01	2.50	1.08	2.65	1.37	ns
	02	4.10	1.37	3.65	1.00	ns
	03	2.60	1.43	2.59	1.54	ns
	04	4.10	1.10	3.71	1.65	ns
	06	3.50	1.51	3.59	1.50	ns
	09	3.20	1.55	2.88	1.22	ns
	10	3.20	1.14	3.18	1.38	ns
	11	4.00	1.63	4.00	1.12	ns
	12	2.40	1.26	2.12	1.17	ns
	13	4.20	1.62	4.12	1.93	ns
	14	2.80	0.63	2.53	1.07	ns
	15	4.60	1.43	3.76	1.44	ns
	16	2.60	1.71	2.41	1.23	ns
	<b>Overall Perceived Difficulty Average</b>	<b>3.40</b>	<b>0.65</b>	<b>3.27</b>	<b>0.72</b>	<b>ns</b>
Students	00	4.05	1.43	4.31	1.08	ns
	01	4.09	1.38	4.89	1.33	t(60) = -2.88, $p < 0.01$ **
	02	2.19	1.54	2.40	1.45	ns
	03	3.52	1.50	3.17	1.27	ns
	04	3.48	1.97	3.36	1.31	ns
	06	3.23	1.17	4.14	1.77	t(33) = -2.07, $p < 0.05$ *
	09	4.55	1.25	4.77	1.34	ns
	10	4.70	1.92	5.51	1.57	ns
	11	4.23	1.59	4.80	1.66	ns
	12	2.77	1.17	3.79	1.84	t(34) = -2.27, $p < 0.05$ *
	13	4.00	1.69	3.76	1.25	ns
	14	4.15	1.85	4.34	1.34	ns
	15	4.15	1.68	4.86	1.78	ns
	16	3.25	1.55	4.00	1.36	t(34) = -1.83, $p < 0.10$ ±
	<b>Overall Perceived Difficulty Average</b>	<b>3.84</b>	<b>1.12</b>	<b>4.21</b>	<b>1.06</b>	<b>t(97) = -2.09, <math>p &lt; 0.05</math> *</b>

Scale: 1, very easy to 7, very difficult. \*\*  $p < 0.01$ ; \*  $p < 0.05$ , ±  $p < 0.10$  (approaching significance). Ns = not significant. Due to the limited number of instructors who identified as URM ( $n < 5$ ), meaningful distinctions between URM and non-URM instructors could not be reliably interpreted.

**Table S5.** Descriptive and inferential statistics, race/ethnicity differences in difficulty perceptions for students only for all items.

		nonURM		URM		Independent Samples t-Test
		Mean	SD	Mean	SD	
Students	00	4.35	1.16	3.77	1.24	ns
	01	4.47	1.32	4.93	1.57	ns
	02	2.18	1.23	3.00	2.15	ns
	03	3.12	1.21	3.57	1.65	ns
	04	3.25	1.26	3.71	2.13	ns
	06	3.50	1.52	4.50	1.75	t(27) = -1.95, $p < 0.10$ ±
	09	4.65	1.21	4.76	1.48	ns
	10	5.25	1.71	5.15	1.77	ns
	11	4.25	1.70	5.44	1.21	t(40) = -2.78, $p < 0.01$ **
	12	3.44	1.72	3.40	1.72	ns
	13	3.69	1.36	4.14	1.46	ns
	14	4.24	1.51	4.33	1.66	ns
	15	4.38	1.86	5.31	1.49	t(37) = -1.88, $p < 0.10$ ±
	16	3.73	1.44	3.69	1.49	ns
	<b>Overall Perceived Difficulty Average</b>	<b>4.00</b>	<b>1.04</b>	<b>4.35</b>		<b>t(65) = -1.76, <math>p &lt; 0.10</math> ±</b>

Scale: 1, very easy to 7, very difficult. \*\*  $p < 0.01$ ; \*  $p < 0.05$ , ±  $p < 0.10$  (approaching significance). Ns = not significant.