Article

Seeing Eye to Eye? Comparing Faculty and Student Perceptions of Biomolecular Visualization Assessments

Josh T. Beckham [1], Daniel R. Dries [2], Bonnie L. Hall [3], Rachel M. Mitton-Fry [4], Shelly Engelman [5], Charmita Burch [6], Roderico Acevedo [7], Pamela S. Mertz [8], Didem Vardar-Ulu [9], Swati Agrawal [10], Kristin M. Fox [11], Shane Austin [12], Margaret A. Franzen [13], Henry V. Jakubowski [14], Walter R. P. Novak [15], Rebecca Roberts [16], Alberto I. Roca [17] and Kristen Procko [18,*]

[1] Freshman Research Initiative, The University of Texas at Austin, Austin, TX 78712, USA; josh.beckham@austin.utexas.edu

[2] Department of Chemistry & Biochemistry, Juniata College, Huntingdon, PA 16652, USA; dries@juniata.edu

[3] Department of Chemistry & Physics, Grand View University, Des Moines, IA 50316, USA; bhall@grandview.edu

[4] Department of Chemistry and Biochemistry, Denison University, Granville, OH 43023, USA; mittonfryr@denison.edu

[5] Custom EduEval LLC, Austin, TX 78749, USA; customedueval@gmail.com

[6] Department of Chemistry, Georgia Gwinnett College, Lawrenceville, GA 30043, USA; cburch@ggc.edu

[7] Department of Chemical and Physical Sciences, Westfield State University, Westfield, MA 01086, USA; racevedo@westfield.ma.edu

[8] Department of Chemistry and Biochemistry, St. Mary's College of Maryland, St. Mary's City, MD 20686, USA; psmertz@smcm.edu

[9] Department of Chemistry, Boston University, Boston, MA 02115, USA; dvardar@bu.edu

[10] Department of Biological Sciences, University of Mary Washington, Fredericksburg, VA 22401, USA; sagrawal@umw.edu

[11] Department of Chemistry, Union College, Schenectady, NY 12308, USA; foxk@union.edu

[12] Department of Biological and Chemical Sciences, The University of the West Indies Cave Hill Campus, Bridgetown BB11000, Barbados; shane.austin@cavehill.uwi.edu

[13] Center for Biomolecular Modeling, Milwaukee School of Engineering, Milwaukee, WI 53202, USA; franzen@msoe.edu

[14] Department of Chemistry, College of Saint Benedict & Saint John's University, Saint Joseph, MN 56374, USA; hjakubowski@csbsju.edu

[15] Department of Chemistry, Wabash College, Crawfordsville, IN 47933, USA; novakw@wabash.edu

[16] Department of Biology, Ursinus College, Collegeville, PA 19426, USA; rroberts@ursinus.edu

[17] DiverseScholar, Irvine, CA 92616, USA

[18] Department of Molecular Biosciences, The University of Texas at Austin, Austin, TX 78702, USA

[*] Correspondence: kristen.procko@austin.utexas.edu

Abstract: While visual literacy has been identified as a foundational skill in life science education, there are many challenges in teaching and assessing biomolecular visualization skills. Among these are the lack of consensus about what constitutes competence and limited understanding of student and instructor perceptions of visual literacy tasks. In this study, we administered a set of biomolecular visualization assessments, developed as part of the BioMolViz project, to both students and instructors at multiple institutions and compared their perceptions of task difficulty. We then analyzed our findings using a mixed-methods approach. Quantitative analysis was used to answer the following research questions: (1) Which assessment items exhibit statistically significant disparities or agreements in perceptions of difficulty between instructors and students? (2) Do these perceptions persist when controlling for race/ethnicity and gender? and (3) How does student perception of difficulty relate to performance? Qualitative analysis of open-ended comments was used to identify predominant themes related to visual problem solving. The results show that perceptions of difficulty significantly differ between students and instructors and that students' performance is a significant predictor of their perception of difficulty. Overall, this study underscores the need to incorporate deliberate instruction in visualization into undergraduate life science curricula to improve student ability in this area. Accordingly, we offer recommendations to promote visual literacy skills in the classroom.

1. Introduction

At a 2013 education-focused meeting of the American Society for Biochemistry and Molecular Biology, a group of instructors began informally discussing the importance of developing students' visual literacy. Defined as an individual's ability to extract meaning from visual representations, visual literacy has been described as a threshold concept in the molecular life sciences—one so foundational that a learner cannot progress to expertise in the field without developing it [1–3]. The group agreed that, while life science education is image-laden, educators do not always provide deliberate instruction on how students should approach the interpretation of such images. To complicate matters, there is a lack of pedagogical frameworks for visual literacy instruction and only a handful of studies exploring its assessment. The team of educators continued this discussion to improve biomolecular visualization instruction through online meetings following the society meeting, laying the foundation for the BioMolViz project.

Over several years and in collaboration with the life science education community, the group identified broad themes in visual literacy, subdividing them into learning goals and objectives [4]. With measurable student outcomes in hand, they considered how to craft assessments that would allow instructors to probe specific objectives. Envisioning a searchable online repository where educators could access assessment instruments to measure biomolecular visual literacy skills, the team planned workshops to engage life science instructors in writing and reviewing these assessments with support from the National Science Foundation.

These workshops transformed into sustained online working groups, where many different members of the expanding network contributed to the process, leading to the development of over 100 peer-reviewed assessment items. A method to validate the assessments was proposed, with the final phase of the process involving large-scale field testing of the items in a broad range of classrooms.

In this study, we describe a smaller-scale pilot used to evaluate our planned approach to classroom testing of biomolecular visualization assessments, where we collected student answers to questions, along with a mixture of quantitative and qualitative data. Intrigued by student comments and their perceptions of assessment difficulty, we expanded our study to capture faculty perspectives and aimed to answer the following research questions (RQ):

RQ1: Which assessment items exhibit statistically significant disparities or agreements in perceptions of difficulty between instructors and students?
RQ2: What differences in perceived difficulty persist between instructors and students even after controlling for race/ethnicity and gender?
RQ3: How does student perception of difficulty relate to performance on the assessment?
RQ4: What predominant themes related to visual problem solving emerge from open-ended feedback that could guide visual literacy instruction and assessment?

2. Literature Review

In biomolecular visualization, complex macromolecules are represented using multiple types of renderings to simplify and highlight various important characteristics of those structures. Visualizations are critical for students' understanding of molecular structure and how it relates to function, a core concept of scientific literacy [5]. Such images serve as a bridge for the communication of scientific ideas that underpin biological processes.

Three frameworks for the teaching and assessment of visual literacy describe overlapping but distinct skills, highlighting a lack of consensus in the community about what defines competence in molecular visualization [4,6–9]. In addition, there are no well-defined expectations of content proficiency based on prior experiences of the learners. To the best

of our knowledge, Mnguni's work is the only study that involved the administration of a series of biomolecular visualization assessments in conjunction with a framework [8].

There are many documented challenges in teaching and assessing biomolecular visual literacy. Student content knowledge and the degree to which a representation resembles a phenomenon are important factors influencing the understanding of a visual representation [9]. Another challenge lies in the translation of 2D images to understand a 3D molecule and its function, a critical skill in scientific literacy [5,10]. Differences in the level of prior exposure to various types of visualizations is also an important issue, and understanding visual narratives requires experience and practice with graphic images [11]. Student traits have also been tied to challenges associated with visual literacy. Emotional factors such as anxiety can influence performance in biology courses [12]. Student preconceptions can impact how they interpret a visualization [13]. This impact is also observed with gender differences using physical models [14]. Additionally, gender differences in spatial ability are commonly reported in STEM [15–20]. Finally, color blindness and other visual impairments can also influence the ability to decode information presented in a visualization [9,21–24].

Experts often gain much more information from a visualization than a novice, with novices' gains improving as their specific content knowledge and visualization-decoding skills progress [8,9]. Students face a high cognitive load as they learn to work with and understand these visualizations [25]. In instances where students are interpreting visualizations and learning specific content at the same time, their cognitive load becomes very high—considerably higher than that of an expert who already has the content knowledge and is able to compile individual concepts into larger units [26,27]. Additionally, mental processes used to interpret information and derive solutions to problems differ between novices and expert learners [28]. Moreover, the schematic approach to problem solving used by experts is rarely employed by novices in STEM fields [29], and this difference likely extends to the way in which novices extract information from an image during visual problem solving.

In addition to the differences in the cognitive loads and problem-solving approaches between novices and experts, consideration must also be given to perceptions of task difficulty. Divergence between instructor and student perceptions are often observed. Research findings support several reasons for this [30–32], including that an instructor's assessment of difficulty is dependent on the "granularity" with which a topic is examined. Student and instructor perceptions tend to diverge when diving deeper into the relevant details necessary to understand a concept [33].

While there is no body of literature comparing student and instructor difficulty perceptions for visualization tasks specifically, the level of detail required to interpret complex visualizations may lead to disparities between instructor and student perceptions. Furthermore, a given individual's perception of task difficulty can impact whether they perform well or not, even within a similar cohort [34].

The above discussion illustrates the complexity and challenges encountered in developing and assessing student visualization skills. To overcome these challenges and help students develop the necessary visual proficiency, visualization skills must be an explicit part of the curriculum [26,35]. Furthermore, a very intentional approach to assessing biomolecular visualization skills is required in order to promote visual literacy among *all* students. Herein, we explore our efforts to develop quality assessments and understand the differences in how they are perceived by students and instructors.

3. Materials and Methods

3.1. Assessment Development

Figure 1 outlines our team's five-stage process to develop and validate biomolecular visualization assessments. Items were initially created through three-day workshops, where participants were introduced to backward design using learning objectives from a Biomolecular Visualization Framework [4,36]. Assessments were crafted to target two different levels of learners: novices (students in introductory courses) and amateurs (college junior/seniors who are life science majors).

Figure 1. Workflow: overview of the assessment review and validation process.

Many of the assessments included original images generated using a 3D modeling program. Guided by a BioMolViz steering committee member, groups of participants created assessment figures and tested their accessibility using the Coblis color blindness simulator developed by Matthew Wickline and the Human–Computer Interaction Resource Network (https://www.color-blindness.com/coblis-color-blindness-simulator/ (accessed on 8 January 2024). To develop assessment prompts, participants were encouraged to use the most common terminology and include alternative terms when appropriate, avoiding jargon that may be challenging for students for whom English is a second language.

Over time, a group of workshop participants began regular online meetings to iteratively review and revise workshop-generated assessments, in addition to authoring original items. Following revision by this working group, assessments were advanced to the steering committee of the project for review. Items requiring further revision were returned to the working group in an iterative process until the steering committee was satisfied that no additional modifications were required. Next, to measure the content validity of the items, the assessments were distributed to a panel of expert reviewers composed of four educators with extensive experience using biomolecular visualization in research and teaching. The panel rated the items based on three dimensions: relevance to the primary learning objective, appropriateness for the learner level (novice or amateur), and overall clarity of the prompt and images. Using a four-point Likert scale ranging from "requiring major revisions" to "excellent, requiring no revision", experts evaluated batches of up to 16 items. Assessments with a mean score of less than 3.6 were returned to the steering committee for further revision. Those scoring higher than 3.6 in all dimensions were prepared for classroom testing, as described below in survey design and administration.

The assessments described in this study, in addition to others at various stages of the validation process, are available for instructor use in the BioMolViz Library (https://library.biomolviz.org/ accessed on 8 January 2024). Viewable items in this searchable database have undergone review by the BioMolViz steering committee and are categorized by their stage in the validation process. Assessments are considered fully validated once they have been tested at scale in a variety of classrooms and evaluated for actual item difficulty, discrimination, effectiveness of distractors, and consistency across different cohorts of students. We invite instructors to use these assessments in their current form or to revise them to suit the unique needs of their classrooms.

3.2. Survey Design and Administration

3.2.1. Student Survey

A total of 15 items from the BioMolViz Library assessment repository were distributed across three surveys. Each survey contained six items, three of which were administered in more than one survey. The majority of the items were multiple-choice questions (MCQs), two were multiple-select (multiple-answer) questions, and one was a short-answer question. For each item, students were also asked to rate its difficulty on a seven-point Likert scale, where 1 was "very easy" and 7 was "very difficult". Finally, each question contained an optional open-ended response: "If applicable, please share any additional feedback about this item that may be helpful". Surveys were administered as an extra credit or ungraded assignment, either in class or as a take-home activity.

To recruit student participants, faculty members were invited to administer visual literacy assessment surveys in their classrooms. The invitation was posted to the group website, emailed to recipients of the BioMolViz newsletter, and further publicized through

a poster at the 2022 American Society for Biochemistry and Molecular Biology annual meeting in Philadelphia, PA. Interested instructors were asked to complete a brief interest form, and once they committed to participating, they entered into an IRB agreement with the host university. Following IRB approval, they chose one or more surveys from the three possible options that best aligned with their course content. Student surveys were administered during the fall of 2022 and spring of 2023, near the end of the course.

Among all the returned surveys, 181 complete responses were collected, where students attempted all assessment items on their survey and rated the difficulty for at least 5 items. These surveys were analyzed as part of this study, while incomplete surveys were excluded. Student demographics are summarized in Table 1 [37]. The general location and institution types where field testing was carried out are shown in Figure 2. A total of seven institutions participated; institutions with common student demographics and geographical locations were grouped together.

Table 1. Student and instructor demographics. Under-represented minority groups (URMs) encompass individuals of Hispanic/Latinx, Black/African American, or Native American/Indigenous descent. Participants identifying as Asian and/or White (including Middle Eastern) were classified as non-URM [37]. Students: *n* = 181; instructors: *n* = 27.

Demographic	Category	Instructors	Students
Gender	Female	63.0%	66.3%
	Male	37.0%	29.8%
	Other	0.0%	0.6%
	Prefer to not say	0.0%	3.3%
URM Status	nonURM	85.2%	71.3%
	URM	14.8%	24.9%
	Prefer to not say	0.0%	3.9%
Year in School	First Year/Sophomore		14.4%
	Junior/Senior		85.6%
Years Teaching Experience	2–9	18.5%	
	10–20	44.4%	
	>20	37.0%	
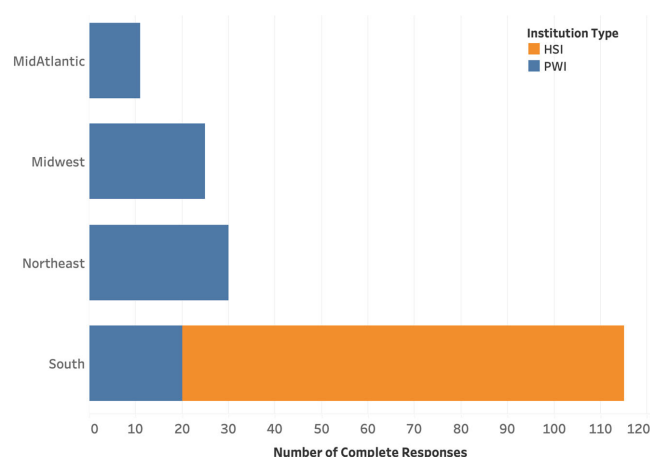
Figure 2. Student respondent data. Data include institution type (HSI = Hispanic-serving institution; PWI = primarily White institutions), geographic region in the United States, and number of complete responses.

3.2.2. Faculty Survey

The instructor survey was designed using the items administered to students. One assessment included in the student survey underwent significant revision following an initial field-testing analysis that explored item difficulty, discrimination, and the efficacy of distractors. As a result of the changes, this item was excluded from this analysis, resulting

in a total of 14 items evaluated across students and instructors. Instructors were provided the items with the correct answers and were asked to rate the difficulty of each on a seven-point Likert scale, analogous to the student survey. Instructors were also specifically asked to comment on their reasoning for their difficulty rating in an open-ended question.

We recruited instructors who had participated in workshops and/or the online assessment working group to complete the faculty survey. Instructor demographics are summarized in Table 1. Across the 27 instructors, 22 schools were represented: five from the south, nine from the northeast, seven from the midwest, and one international institution. Of the domestic schools, eighteen were PWIs, and three were HSIs. The instructor survey was distributed during the summer of 2023.

3.2.3. Data Analysis

Data analysis encompassed a series of statistical tests tailored to address the research questions. We conducted a power analysis to ascertain the adequacy of our sample size in detecting significant differences between these two groups. Assuming a medium effect size, as defined by a Cohen's d ranging from 0.5 to 0.8, our power analysis revealed that we achieved a substantial level of statistical power. The calculated power values fall within the range of 0.78 to 0.99, demonstrating that our study was well-equipped to detect significant differences and effects. This robust statistical power underscores the reliability of our findings and enhances the validity of our conclusions.

RQ1 sought to uncover differences in perceptions of assessment difficulty between instructors and students. We first employed descriptive statistics to capture overall ratings provided by both groups. To assess statistical significance between groups, independent-samples *t*-tests were conducted to determine whether noteworthy differences existed in how instructors and students perceived the overall difficulty of the assessments. We then analyzed each individual assessment item to identify those that displayed significant disparities or agreements in perceptions between instructors and students. For this item-wise analysis, we utilized a combination of descriptive statistics and *t*-tests for independent samples.

RQ2 explores whether the perception of difficulty was influenced by gender and race/ethnicity. We used regression analysis with participant status (instructor or student) as the predictor variable and controlled for gender and race/ethnicity.

To investigate RQ3, concerning the relationship between student performance (actual difficulty as % correct) and perceptions of assessment difficulty (scale: 1 = very easy to 7 = very difficult), we conducted regression and correlation analyses. In our regression model, we included the percentage of total correct responses as an additional predictor variable, allowing us to explore how students' performance on the assessment related to their perceptions, while simultaneously considering the effects of gender and race/ethnicity. For the correlation analysis, we used the Pearson correlation coefficient to indicate a positive or negative correlation between students' overall performance on the assessments and their perceptions of their difficulty; we applied this same approach in an item-wise analysis. A *p* value less than 0.05 indicates statistical significance when calculated with a 2-tailed approach.

Finally, to address RQ4 and uncover predominant themes related to visual literacy/visual problem solving, we evaluated open-ended feedback from the surveys. Our qualitative analysis began with a structural coding approach [38]. One of the project leads, serving as the codebook editor [39], identified five broad areas for coding: image-related considerations, assessment prompt, topic-related conceptual considerations, description of difficulty, and discussion of item quality. Our second round of coding took a more descriptive approach, identifying two to three child codes under each of the aforementioned parents. Both student and instructor surveys were collaboratively coded [38,40], with a minimum of three team members coding the responses and discussing and refining subcategories. The 5 broad areas were ultimately divided into 14 subcategories to capture themes emerging from the comments (e.g., image interpretation considerations vs. the choice of visual representation). Our third round of coding included a final discussion and revision of child codes, followed by review of each item by a pair of team members.

Through our thematic analysis, we uncovered three visualization-based emergent themes: (a) expectations about images guide student performance, (b) disparities exist in visual literacy problem-solving approaches, and (c) content knowledge can be both a help and a hindrance in visualization. The thematic analyses are presented as embedded observations within the results of RQ1, RQ2, and RQ3.

These analytical approaches collectively provided a nuanced understanding of the perceived difficulty of the assessments, offering insights into overall perceptions, specific item-level differences, and the influence of demographic and performance-related factors on participants' perceptions. Analysis of these results in conjunction with thematic coding allowed us to comment on trends in biomolecular visual literacy assessment.

4. Results and Discussion

Macromolecular visualization in life sciences courses often focuses heavily on proteins, with the inclusion of a few carbohydrates and nucleic acids. Our team's "macromolecule-agnostic" approach prompted us to include structures of diverse macromolecules and small molecules across the surveys. The type of molecule displayed in the assessment image, the primary learning objective targeted from the Biomolecular Visualization Framework, and the question type are summarized in Table 2 [4].

Table 2. Visual literacy assessment item overview and student performance data.

Item ID	Molecule Type(s)	Primary Learning Objective	Question Type	% Correct	*n*
00	Nucleic acid (DNA)	AR2.03: Students can identify or create a suitable rendering or combination of renderings for a specific purpose.	MA/MS	82 *	61
01	Protein, saccharide (carbohydrate)	MA1.01: Students can identify individual biomolecules in a macromolecular assembly.	MCQ	59	120
02	Tetrapeptide	MB1.01: Given a rendered structure of a biological polymer, students are able to identify the ends of a biological polymer.	MCQ	11	70
03	Dipeptide	MB1.03: Given a rendered structure, students can identify the individual building blocks.	MCQ	87	70
04	Carbohydrate (trisaccharide)	TC1.02: Students can use appropriate terms to describe the linkages/bonds/interactions that join individual building blocks together in a macromolecule or macromolecular assembly.	MA/MS	78 †	70
06	Carbohydrate (trisaccharide)	TC1.02: Students can use appropriate terms to describe the linkages/bonds/interactions that join individual building blocks together in a macromolecule or macromolecular assembly.	MCQ	42 ‡	50
09	Carbohydrate (trisaccharide)	LM1.02: Students can visually identify non-protein chemical components in a given rendered structure.	MCQ	36	111
10	Protein	MB1.03: Given a rendered structure, students can identify the individual building blocks.	Short answer	44 ¶	61
11	Protein	SF2.03: Students can identify functionally relevant features of a macromolecule.	MCQ	44	50
12	Small-molecule drug	MI1.03: Students can predict whether a functional group (region) is a hydrogen bond donor or acceptor.	MCQ	54	50
13	Small-molecule drug	MI1.03: Students can predict whether a functional group (region) is a hydrogen bond donor or acceptor.	MCQ	43	70
14	Amino acids, heme group	MI1.02: Students can identify the different non-covalent interactions given a 3D structure.	MCQ	50	131
15	Oligopeptides	MB1.03: Given a rendered structure, students can identify the individual building blocks.	MCQ	50	50
16	Protein, lipid bilayer	MA1.01: Students can identify individual biomolecules in a macromolecular assembly.	MCQ	51	61

Item IDs are a truncated version of the deposition ID from the BioMolViz Library assessment repository and are not necessarily consecutive (see further details in the Supplementary Material). For the categorization of question type, MA/MS is a multiple-answer (multiple-select) question, and MCQ is a multiple-choice question. The short-answer question required students to enter the names of six amino acids shown in an image. * Average of percentages correct for each individual choice. 44.3% of students provided fully correct answers. † Average of percentages correct for each individual choice. 54.3% of students provided fully correct answers. ‡ This assessment presented a carbohydrate with an N-acetyl group, and students were asked to identify an amide functional group. Another 38% chose the incorrect answer, "peptide bond". ¶ Average of percentages correct answers for each individual choice. 14.8% of students provided fully correct answers.

Overview of Student Performance

The percentage of students that responded correctly to each item varies widely from question to question (Table 2). As we prepared for a deeper analysis of performance and perceived difficulty, we first disaggregated the percentage of correct answers by gender and race/ethnicity to note any demographic differences emerging in performance. When examining the overall assessment, no statistically significant differences in overall performance were observed between these groups (Figure S1, Supplementary Material). However, some statistically significant differences were revealed through an item-wise analysis. Males outperformed females on assessments 06, 10, and 12 (Table S1, Supplementary Material), while females outperformed males on item 03. In our evaluation of race/ethnicity by item, statistically significant differences were observed for three items, with non-URM students outperforming URMs on items 02, 10, and 16 (Table S2, Supplementary Material). Open-ended feedback did not reveal any indication of the origin of performance disparities due to gender or race/ethnicity.

Having obtained insights into student performance, we transitioned to exploring the differences between novice and expert perspectives on these visual literacy assessments. To address RQ1, we analyzed Likert-scale ratings provided by both students and instructors.

RQ1: Which assessment items exhibit statistically significant disparities or agreements in perceptions of difficulty between instructors and students?

Using a combination of descriptive statistics and independent-samples *t*-tests, we looked for statistically significant agreements and deviations in students' and instructors' perceptions of item difficulty. Collapsing across all assessment items, the bottom row of Figure 3 shows that, overall, students (mean = 4.08; SD = 1.09) perceived the assessments as being significantly more difficult than instructors (mean = 3.32; SD = 0.69; t(48) = 4.94; $p < 0.01$) (see also Table S3, Supplementary Materials). Examining individual assessment items, Figure 3 further reveals that among the 14 evaluated items, 8 exhibited statistically significant differences between students and instructors at $p < 0.05$. Additionally, two items approached significance at $p < 0.10$, while four others displayed no significant distinctions, suggesting similarity in difficulty perceptions (Table S3, Supplementary Materials).
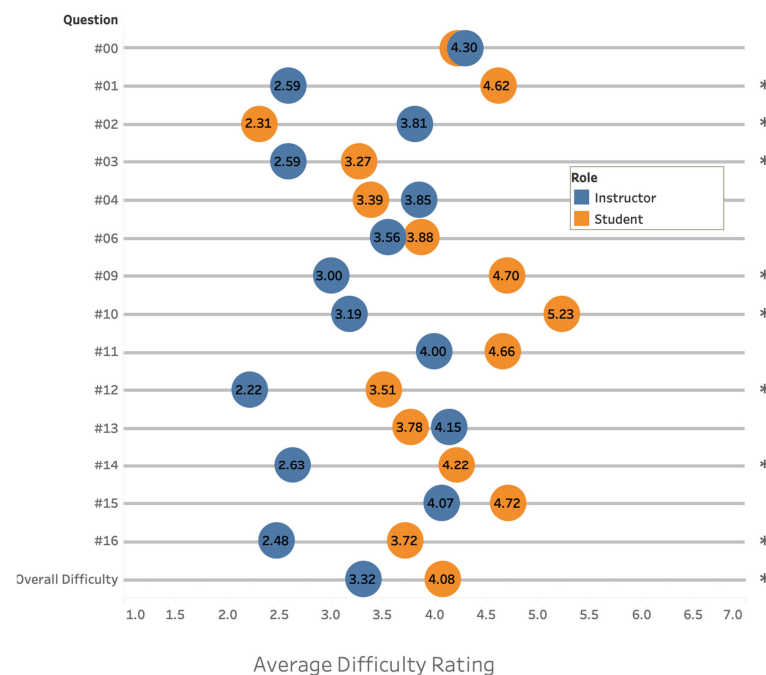

Figure 3. Perceived difficulty of assessments. Average perceived difficulty of the 14 assessment items as reported on a 7-point Likert scale by both students (orange circles) and instructors (blue circles). The stars on the right axis indicate a statistically significant *p* value (*p* < 0.05) between the perception of difficulty of the respondents on that question (see Table S3 for the numbers of respondents).

Consistent with our expectations, overall, students perceived the assessments as significantly more challenging than instructors. This finding is consistent with prior research indicating students often find assessments more difficult due to factors including differences in perspective, emotions and expectations regarding assessments, and the cognitive demand required to complete them [12,32,41,42]. Specifically, students find assessments that involve the transfer of information from one subject to another more difficult [32]; visual problem solving is characterized by such transfer. The work of van de Watering and van der Rijt also suggests that instructors tend to underestimate the complexity of certain items.

Items 04, 06, and 09 present an interesting comparison because the assessments use a similar visualization of an N-acetylated modified carbohydrate adapted from PDB ID 1gya (Figure 4). Students found assessment 04 marginally less difficult than instructors. This multiple-choice question required them to select glycosidic linkages from arrows in a labeled version of the structure, and it was perceived as easier than identifying a circled amide functional group within the structure from a multiple-choice list (item 06). Of the three visually related assessments, item 09—in which students were asked to identify the molecule type as an oligosaccharide from a list of choices—was the only one that showed statistically significant differences in perceptions, with students rating it as much more difficult than instructors.
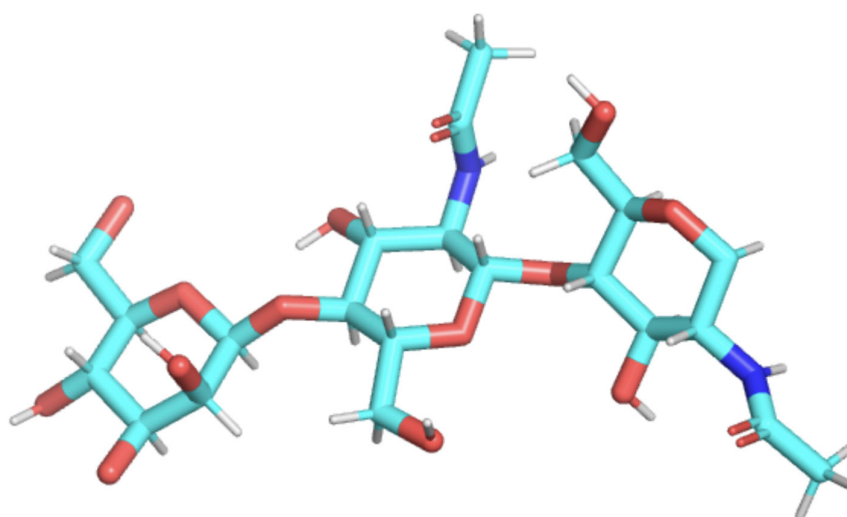
Figure 4. Modified carbohydrate image used in item 09; items 04 and 06 present a similar image with additional labels. The structure is displayed using CPK coloring with cyan carbon atoms.

Several instructors commented on the presentation of the carbohydrate in items 04, 06, and 09 as a 3D molecular structure: "This seems like a moderately difficult question because students are not often used to looking at carbohydrates in this representation". A student corroborated this in their open-ended feedback: "It is different viewing the molecules in this 3D structure. We do not usually see structures this way in lecture."

Typically, carbohydrates are presented in biochemistry courses as chemical structures—often Fischer or Haworth projections. Despite the popular use of these representations, it is notoriously difficult for students to extract stereochemical information from them [43], which suggests that 3D molecular representations should be more widely used in instruction.

Across these three assessments, the question being asked about the image contributed more to the perception of difficulty than the visualization itself, which remained nearly identical. If students were familiar with a glycosidic linkage, the prompt for item 04 revealed the molecule as a carbohydrate and only required students to focus on finding the bond. Item 06 required recognition of a single functional group in the structure.

Identification of the molecule type represented the largest challenge for students overall, in both percent correct and perceived difficulty. Based on instructor quotes, the

carbohydrate modification may have contributed to the challenge: "Glycopeptide might be a common error for this question, since the carbohydrate contains nitrogen atoms". Indeed, this aspect of the visualization appears to have led a student to the incorrect answer.

> It has no phosphorus and contains amine. It's also very ringy so I'm guessing it's glycopeptide but I'm not sure.

This student possessed the visual literacy needed to identify features of the molecule but lacked the ability to discern an N-acetyl carbohydrate modification from the presence of an amino acid. The perceived difficulty rating appears to originate from a combination of the visualization and assessment task.

Within the subset of the eight items displaying significant differences in perceived difficulty, students viewed seven of these items as more challenging than instructors. However, it is noteworthy that one assessment (item 02) stood out as an exception, with instructors rating it as significantly more demanding than students did. The student and instructor open-ended feedback illuminated the first of three emerging themes related to visual literacy assessment.

Theme 1: Expectations about images guide student performance (RQ4).

Assessment 02 required students to identify the N terminus of a peptide. The inclusion of an N-terminal asparagine placed the N terminus in a position that was not on the far left of the image; instead, the asparagine side chain was on the left (Figure 5). This assessment was designed to test students' visual literacy by requiring them to follow the peptide backbone to correctly identify the N terminus.
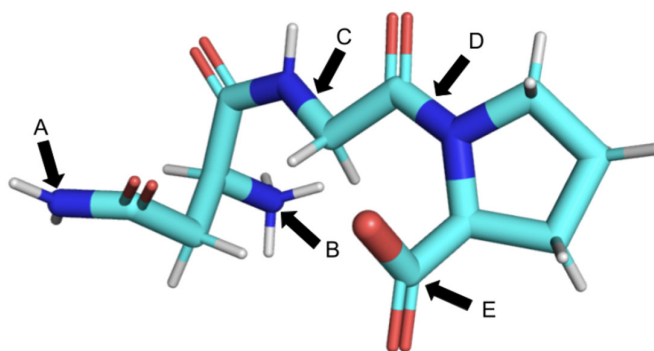
Figure 5. The tetrapeptide image from assessment 02, where students are asked to identify the N terminus from choices A–E. The structure is displayed using CPK coloring with cyan carbon atoms.

Instructors commented on the molecule's orientation as a positive.

> This is a great assessment; I actually like that the N-terminal end isn't at the far left side of the image because I think that students would automatically go there. This item really tests to see if students can map two-dimensional drawings onto a 3D image.

A total of 8 of the 26 instructors who provided open-ended feedback made predictions about student performance, with two stating they expected about 50% of students to answer correctly. This proved an overestimation, with only 11% of students selecting the correct response.

Indeed, student comments alluded to their expectations about the representation, which led them to choose the incorrect response: "I found the end of the chain, and so one side had the COO-, so I chose the other end". Even a student that answered correctly commented that it was "tricky" to locate the N terminus in this assessment. Interestingly, one student presented a definition analyzing features that would enable them to answer the question correctly, yet still selected the incorrect answer.

> N-terminus does not contain another carbonyl group that often characterizes amino acids. It is a lone amino group.

Despite the knowledge that the N terminus should lack a carbonyl group, the student selected an answer *containing* a carbonyl, demonstrating a lack of the visual literacy required to identify the N terminus based on that definition. This student chose the group with the leftmost nitrogen—the convention that guided many students' expectations for the visualization and prevented them from engaging in a deeper analysis of the structure.

These assessments were administered in several biochemistry courses, where many instructors place a strong emphasis on learning the biochemical "alphabet" of the 20 standard amino acids. Students found this task challenging; the assessments that students perceived as most difficult both involved recognition of amino acid side chains (items 10 and 15). Assessment 10, a short-answer question that required students to name amino acids based on the displayed side chains, also corresponded to the largest *disparity* in student and instructor perceptions, with a mean difference of 2.04, indicating that students found it much more difficult than faculty.

Interestingly, despite the similar content of items 10 and 15, instructors and students *converged* on their evaluation of difficulty for item 15, a multiple-choice question involving amino acid side-chain recognition (Figure 6).
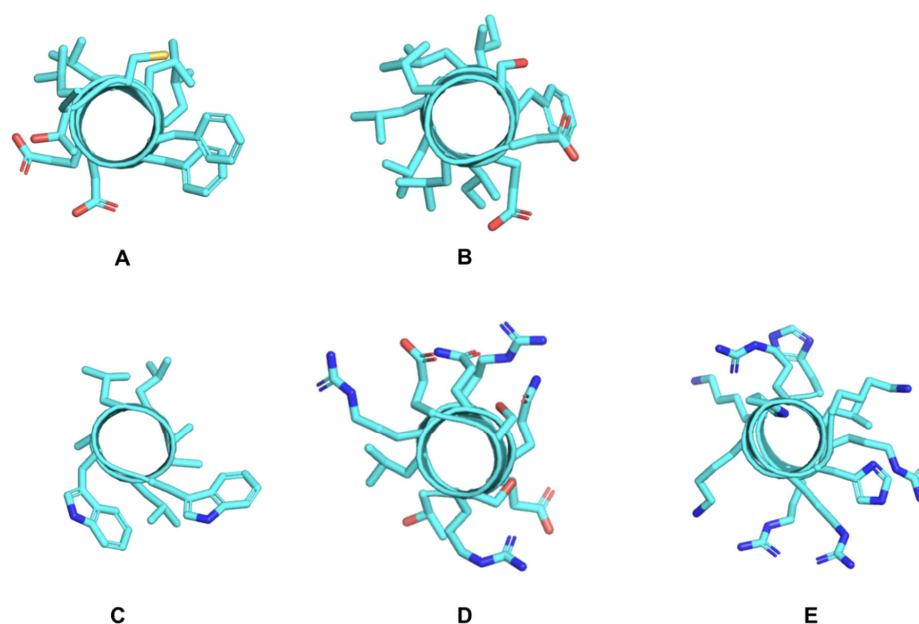
Figure 6. Assessment 15, presenting a series of five oligopeptide images where students were asked to determine which multiple-choice question option (**A**–**E**) matched the following sequence: Glu-Ser-Ser-Leu-Gln-Gln-Arg-Arg-Arg-Glu-Thr. The structures are displayed using CPK coloring with cyan carbon atoms.

Instructors commented that the difficulty of this assessment would vary, depending on the problem-solving method students used, which underscores our next emergent theme.

Theme 2: Disparities exist in visual literacy problem-solving approaches (RQ4).

Similar to extraneous values in mathematical problems and force and motion problems in physics [44], visual assessments require relevant information to be extracted from the image. In their open-ended feedback for many of the items, instructors described how students might approach solving visual literacy problems—in some cases, outlining strategies students could use to eliminate distractors. Instructors anticipated disparities in the way students would approach item 15, leading to a relatively higher difficulty rating of this assessment by instructors.

> Without being able to rotate the structures to see the linear view of the amino acids, some students would be stumped. Others could look at the side chains and realize that some of the amino acid residues in question are not in some of

the structures and determine which structure contains them all. This question relies on student recognition of amino acid side chains and has sort of a "puzzle component" that requires learners to realize that they do not need to view the molecule as a linear structure to answer the question. This question depicts the protein helix beautifully and will help students understand how helices interact.

Students expressed a contrasting view about the depiction of the helix. Of the six students who provided open-ended feedback, three expressed confusion or difficulty with the representation of the helix.

This was very hard to visualize. It honestly looks like a wobbly corkscrew pasta and feels like an ineffective way to model these structures.

For students unfamiliar with the helical wheel, this type of representation may introduce additional challenges. Of the six students who provided open-ended feedback, five commented on various aspects of the way the image was displayed, while only one indicated a strategy of looking for "nitrogen rich" amino acids. However, more than half of the instructors described the potential process of elimination that could be applied to this problem based on recognizing the colors of atoms present in the side chains that were visible. These open-ended comments reinforce differences in expert thinking and approaches to problem solving [28,29], which were rarely present in the novice responses.

Students' perceptions of difficulty regarding such assessments involving amino acid side-chain recognition are likely to depend on whether the instructor presents molecular structures generated using a 3D modeling program with the atom identity indicated by a standard color or only chemical structures, where the atoms are represented by their text-based chemical symbols. Moreover, familiarity with the side chains is related to whether students are required to memorize amino acids or allowed to use an amino acid chart on assessments. In future field-testing studies, we intend to probe course content through an instructor survey and specifically examine the use of 3D representations and whether amino acid memorization is required.

Student's perceptions of difficulty may also depend on whether they felt that the assessment required them to turn the structure in their mind. Given previous findings that male students tend to outperform female students in tasks that require 3D mental rotations [45], we wondered if this type of task has the potential to be perceived as more difficult by female students. We next analyzed whether other demographic features played a role in perceptions in RQ2.

RQ2: What differences in perceived difficulty persist between instructors and students even after controlling for race/ethnicity and gender?

To account for the potential impact of gender and race/ethnicity on the evaluation of assessment difficulty, regression models were employed. These models examined the influence and proportion of variance explained by demographic variables in elucidating disparities between instructors and students. Initially, descriptive statistics and *t*-tests were conducted to investigate potential differences between males and females and between underrepresented minorities (URMs) and non-URMs.

Among instructors, no significant differences in difficulty perceptions were revealed between males and females (Table 3; see also Table S3, Supplementary Material). However, within the student group, females (mean = 4.21, SD = 1.06) perceived the overall assessment as significantly more challenging than males (mean = 3.84, SD = 1.21; $t(97) = -2.09$, $p < 0.05$). In particular, items 01, 06, 12, and 16 displayed statistically significant differences, with female students rating them as more difficult. Returning to assessment 15, which brought up questions about mental rotation, female students did, indeed, rank the item as slightly more difficult than males, but this result was not statistically significant. It is worth noting that males outperformed females on items 06 and 12; however, on items 01 and 16, there were no statistically significant performance differences. All four of these items include a non-protein component in the image, and three involved identifying the molecule type or a functional group; perhaps, identifying features of these less familiar structures was

perceived as more difficult by female students in this study. Several other assessments with non-protein components were not found to be more difficult by females, suggesting that this effect is minor.

Table 3. Difficulty perceptions and gender. Key descriptive and inferential statistics and gender differences in difficulty perceptions between instructors and students.

		Male		Female		Independent-Samples *t*-Test
		Mean	SD	Mean	SD	
Instructors	Overall Difficulty Average	3.40	0.65	3.27	0.72	ns
Students	01	4.09	1.38	4.89	1.33	t(60) = −2.88, *p* < 0.01 **
	06	3.23	1.17	4.14	1.77	t(33) = −2.07, *p* < 0.05 *
	12	2.77	1.17	3.79	1.84	t(34) = −2.27, *p* < 0.05 *
	16	3.25	1.55	4.00	1.36	t(34) = −1.83, *p* < 0.10 ±
	Overall Difficulty Average	3.84	1.12	4.21	1.06	t(97) = −2.09, *p* < 0.05 *

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

In our analysis of the correlation of race/ethnicity with difficulty perception, some disparities emerged between URM and non-URM students in their perceptions (Table 4; see also Table S4, Supplementary Material). In general, the data indicated that URMs tended to perceive the overall assessment as slightly more difficult than non-URMs, and this was reflected in higher mean difficulty perception scores across most items. However, in most cases, these distinctions only approached significance at *p* < 0.10. Yet, for item 11, significant differences were observed.

Table 4. Difficulty perceptions and race/ethnicity. Descriptive and inferential statistics and race/ethnicity differences in difficulty perceptions for items approaching statistical significance.

		Non-URM		URM		Independent-Samples *t*-Test
		Mean	SD	Mean	SD	
Students	06	3.50	1.52	4.50	1.75	t(27) = −1.95, *p* < 0.10 ±
	11	4.25	1.70	5.44	1.21	t(40) = −2.78, *p* < 0.01 **
	15	4.38	1.86	5.31	1.49	t(37) = −1.88, *p* < 0.10 ±
	Overall Difficulty Average	4.00	1.04	**4.35**	**1.18**	t(65) = −1.76, *p* < 0.10 ±

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.

Interestingly, although URM students perceived three of the items (06, 11, and 15; Table S5, Supplementary Material) to be more difficult than the non-URMs, there were no differences in performance (Table S2). Conversely, for the three items where there was a significant difference in performance between URM and non-URM students (02, 10, and 16; Table S2), there were no differences in perceived difficulty.

These results introduce some thought-provoking questions concerning variations in the perception of difficulty for molecular visualization assessments and whether it truly represents a significant learning barrier, especially for females or URMs. In an upcoming large-scale field-testing effort, we will explore if these perceived difficulty and/or performance trends persist and if the reasons behind differences in perceived difficulty corresponding with race/ethnicity or gender can be uncovered through student interviews and/or focus groups.

Considering the perception variations noted above, it is crucial to consider potential gender and race/ethnicity differences when exploring demographic influences on perceptions (Table 5). The results of our regression model suggest that, even with the inclusion of race/ethnicity (β = 0.10, non-significant) and gender (β = 0.11, non-significant) in the model, the role of the participant—either instructor or student (β = −0.24, *p* < 0.01)—continues to emerge as a robust predictor of perceived difficulty (the outcome variable).

Table 5. Regression model for perceived difficulty vs. demographics. The regression model accounts for role, race/ethnicity, and gender.

Outcome Variable = Overall Difficulty Average	Unstandardized Coefficients		Standardized Coefficients	t	*p*-Value
	B	Std. Error	β		
(Constant)	3.86	0.14		27.81	0.000
Role (instructor, 1; student, 0)	−0.74	0.21	−0.24	−3.43	0.001 **
Race/Ethnicity (URM, 1; non-URM, 0)	0.25	0.17	0.10	1.45	0.147
Gender (female, 1; male, 0)	0.25	0.16	0.11	1.57	0.119

R-squared = 0.09, ** *p* < 0.01. β values range from 0 to 1 or 0 to −1, depending on the direction of the relationship. Values closer to 1 or −1 represent a stronger correlation.

In summary, despite observed disparities in perceptions associated with race/ethnicity and gender, the impact of these demographic variables does not diminish the significance of participant type (instructor vs. student), revealing this latter characteristic as a much more important predictor of perceived difficulty. With role established as the key predictor of difficulty perceptions, we examined the student cohort to evaluate whether their performance had an influence on their perception.

RQ3: How does student perception of difficulty relate to their performance on the assessment?

A regression analysis was conducted exclusively within the student cohort to investigate this RQ. The primary aim was to ascertain whether variations in students' performance (% correct) were related to their perception of the difficulty of the assessment. The regression model revealed a significant and negative relationship between students' performance and their perception of overall assessment difficulty (Table 6, β = −0.40, *p* < 0.01), suggesting that students who perceived the assessments as less challenging performed better on the survey.

Table 6. Regression model for perceived difficulty vs. demographics and performance. The regression model accounts for race/ethnicity, gender, and performance for students.

Outcome Variable = Overall Difficulty Average	Unstandardized Coefficients		Standardized Coefficients	t	*p*-value
	B	Std. Error	β		
(Constant)	4.84	0.23		20.99	0.000
Gender (female, 1; male, 0)	0.22	0.17	0.09	1.31	0.191
Race/Ethnicity (URM, 1; non-URM, 0)	0.24	0.18	0.10	1.38	0.171
Overall % Correct	−0.02	0.00	−0.40	−5.61	0.000 **

R-squared = 0.19, ** *p* < 0.01.

Furthermore, the analysis considered the impact of gender and race/ethnicity on students' perceptions of assessment difficulty. The inclusion of these demographic variables as covariates in the model allowed for an examination of whether they contributed to the observed relationship. However, the findings indicate that neither gender (β = 0.09, *p* = 0.191) nor race/ethnicity (β = 0.10, *p* = 0.171) had a significant effect on students' perceived assessment difficulty after accounting for their performance.

The results of this regression analysis suggest that students' performance shapes their perceptions of assessment difficulty, with improved performance associated with decreased perceived difficulty. Additionally, it appears that students' perceptions are not solely influenced by the inherent difficulty of assessments but are also shaped by their anticipated performance outcomes. This aligns with research that strongly correlates the role of self-efficacy with achievement in students' perceptions of academic tasks [46,47]. It underscores the importance of addressing not only the content and design of assessments but also students' confidence and self-perception as learners.

It is important to note that students did not receive information about their performance while completing the survey or even immediately after; therefore, they had no

evidence of achievement that would influence difficulty perceptions. As we delve into our third theme, this relationship provides a compelling backdrop for a more in-depth examination of the intricate interplay between content knowledge and visualization abilities.

Theme 3: Content knowledge can be both a help and hindrance in visualization (RQ4).

To further explore individual assessments, we performed item-wise correlations of average percent correct and students' perceived difficulty, which we established is connected to their anticipated performance. This relationship between judgment of performance and actual performance on each individual item is known as relative accuracy [48]. Generally, students exhibited an ability to accurately determine their knowledge of concepts covered by most of the items, which is represented by a negative correlation in Figure 7. However, for items 02 and 13, there was a positive correlation. In these two instances, we see an inability of students to differentiate between what they know and do not know about the visualization of the biomolecules in these two items. Students were overconfident in their abilities [49], although to a lesser extent for item 02, which had a non-significant *p* value.
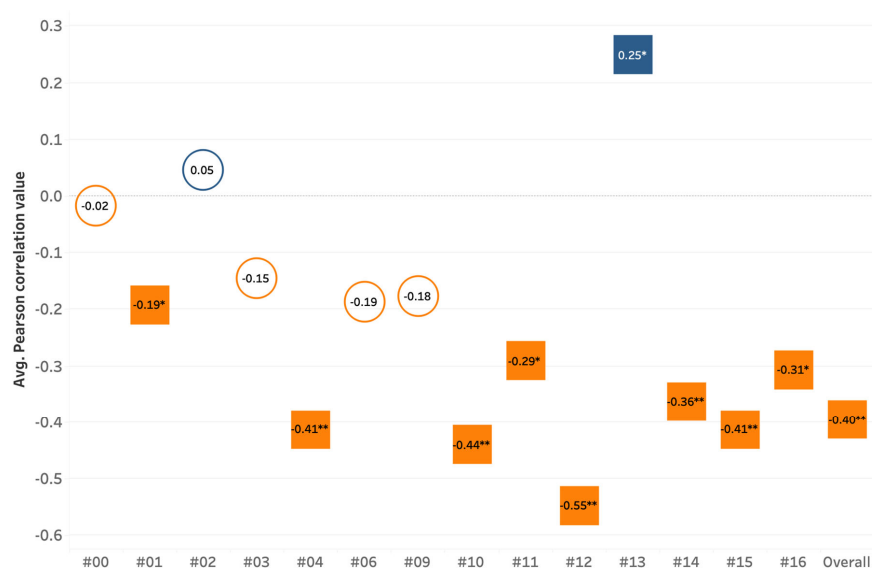
Figure 7. Perceived difficulty vs. performance. Pearson correlation of student perception of difficulty compared to average percent correct on individual survey items. Orange circles and squares show a negative correlation, while blue circles and squares highlight a positive correlation. Open circles show a non-significant *p* value, while filled squares demonstrate a significant *p* value (see correlation value for threshold information). Correlation significance (2-tailed) at the 0.01 level is indicated by ** and by * at the 0.05 level.

Item 13, our archetypal example to illustrate Theme 3, exhibited a statistically significant positive correlation between perceived difficulty and percent correct. This assessment requires students to evaluate the hydrogen-bonding ability of an atom, presenting a nitrogen atom with a lone pair that is engaged in resonance and contributes to the aromaticity of the compound, making it unlikely to serve as a hydrogen bond acceptor (Figure 8).

Item 13 presented a tale of caution through our thematic analysis: students' visual literacy needs to be calibrated to understand the nature of their answers—in some instances, even when they are correct. Analysis of open-ended feedback revealed that some students responded correctly because of a lack of content knowledge: "Not really sure, but I believe that that is a nitrogen with a hydrogen". Observing the atoms, the student selected the correct answer without a deeper analysis of the hydrogen-bonding capability of the atom given its environment.

Conversely, some students who demonstrated a better understanding of nitrogen's ability to serve as a donor and acceptor through open-ended feedback answered incorrectly. One student who described the resonance in the molecule misinterpreted the significance

of that effect and selected the incorrect answer: "The nitrogen can accept a hydrogen, but I was thinking that it could also donate the hydrogen it has because it could be resonance stabilized by the nearby carbonyl".
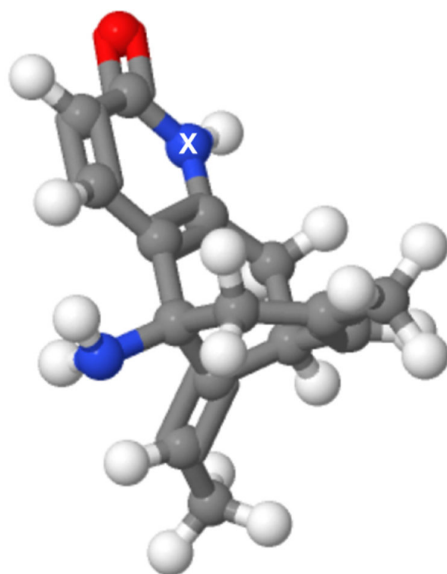
Figure 8. Item 13 image. A ball-and-stick representation of huperzine A from assessment 13; students were asked to identify atom X as a hydrogen bond donor, acceptor, both a donor and acceptor, neither a donor nor acceptor, or whether the donor/acceptor ability cannot be predicted from the information given. The structure is displayed using CPK coloring with gray carbon atoms.

Students who demonstrated a stronger understanding of the concepts in their open-ended feedback were able to identify the nitrogen as a hydrogen bond donor only—this time, based on a deep understanding of both the presented visualization and content knowledge.

Atom X can serve as a hydrogen bond donor because it is sufficiently electronegative and its bonded hydrogen can then interact with other molecules. However, the lone pair is not involved as a hydrogen bond acceptor because it participates in resonance.

This assessment suggests that analysis of the student thought process is needed to calibrate visual literacy. Only when students clearly articulated their understanding of resonance in the open-ended feedback could we ascertain that they were answering correctly because of their proficiency.

Interestingly, for item 02, the quintessential example illustrating *Theme 1: Expectations about images guide student performance*, students found the item easier than instructors by 1.5 points; however, only 11% answered correctly. Again, for this assessment, the slightly positive correlation between perceived difficulty and performance indicates that the percent correct for item 02 did not correspond with the assigned difficulty rating (Figure 7). Although performance on the assessment was a strong predictor of difficulty perceptions, in general, there were items where students' inflated assurance in their answer prevented them from identifying the actual difficulty of the assessment. This observation is consistent with the Dunning–Kruger effect, a cognitive bias whereby individuals with lower abilities tend to overestimate their competence, potentially hindering critical self-reflection and contributing to underperformance [50]. This, in turn, suggests that when instructors over rely on conventions—such as drawing a polypeptide with the terminal amine on the left side and the terminal carboxylic acid to the far right—they may inadvertently encourage students to overlook pertinent information.

5. Conclusions

The main outcome of our study is evidence of a significant disparity in the perception of assessment difficulty between students and instructors; as educators, we must acknowledge this incongruence to better meet students "where they are" to provide superior visual literacy instruction. In the classroom, we recommend increased variety in images, molecule type, and three-dimensional perspectives presented to students. While multiple-choice questions were predominantly used for quantitative analysis in our study, it was the qualitative analysis that revealed the nuances of students' approaches to interpreting visualizations. In some cases, a correct answer on an MCQ can be insufficient to effectively probe understanding. We therefore recommend that instructors use iterative formative assessments to teach visual literacy in order to identify both student assumptions and misconceptions. Discussing students' rationales for incorrect and correct responses to MCQs as a formative evaluation could bridge the gap between instructors' and students' approaches and address student expectations that lead to overconfidence. Furthermore, the common misconceptions revealed through these discussions would provide superior distractors for the development of future MCQs.

Moreover, we call for a consensus on visual literacy assessment and encourage the use of common frameworks as we develop teaching tools and assessments. We humbly offer the peer-reviewed items within the BioMolViz Library as a resource for assessment, and encourage the repurposing of images curated for this open-educational resource as a starting point for formative assessments. Finally, we encourage educators to be more deliberate in their overall instruction and assessment to ensure that biomolecular visual literacy becomes an integral part of the learning process in the life sciences.

Author Contributions: Conceptualization, J.T.B., D.R.D., S.E., H.V.J., M.A.F., P.S.M., W.R.P.N., R.R., A.I.R., K.P.; Methodology, J.T.B., D.R.D., S.E., H.V.J., P.S.M., K.P.; Validation, S.E., R.M.M.-F., R.R., K.P.; Formal Analysis, J.T.B., S.A. (Swati Agrawal), C.B., S.E., B.L.H., P.S.M., R.M.M.-F., D.V.-U., K.P.; Investigation, J.T.B., B.L.H., P.S.M., R.M.M.-F., K.P.; Resources, D.R.D., H.V.J., M.A.F., W.R.P.N., R.R., A.I.R., K.P.; Data Curation, J.T.B., B.L.H., P.S.M., R.M.M.-F., D.V.-U., K.P.; Writing—Original Draft Preparation, J.T.B., S.A. (Swati Agrawal), S.A. (Shane Austin), C.B., S.E., B.L.H., P.S.M., R.M.M.-F., D.V.-U., K.P.; Writing–Review & Editing, J.T.B., D.R.D., R.A., S.A. (Swati Agrawal), S.A. (Shane Austin), C.B., S.E., K.M.F., B.L.H., H.V.J., M.A.F., P.S.M., R.M.M.-F., W.R.P.N., R.R., A.I.R., D.V.-U., K.P.; Visualization, S.E., B.L.H., K.P.; Project Administration, J.T.B., K.P.; Funding Acquisition, J.T.B., D.R.D., H.V.J., M.A.F., W.R.P.N., R.R., A.I.R., K.P. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: All subjects gave their informed consent for inclusion before they participated in the study. The study was conducted in accordance with the Declaration of Helsinki, and the protocol was approved by the Institutional Review Board (IRB) at The University of Texas at Austin under IRB ID: 2019070083-MOD01 on September 30th, 2022 for the student surveys. Each participating faculty had their institution enter into a Memorandum Of Understanding (MOU) with the University of Texas in order to participate in the study. The faculty survey was approved under 2019070083-MOD02.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

References

1. Loertscher, J.; Green, D.; Lewis, J.E.; Lin, S.; Minderhout, V. Identification of Threshold Concepts for Biochemistry. *CBE—Life Sci. Educ.* **2014**, *13*, 516–528. [CrossRef]
2. Schönborn, K.J.; Anderson, T.R. The importance of visual literacy in the education of biochemists. *Biochem. Mol. Biol. Educ.* **2006**, *34*, 94–102. [CrossRef]
3. Meyer, J.H.; Land, R.; Baillie, C. *Threshold Concepts and Transformational Learning*; Brill Publishers: Brill, The Netherlands, 2010; Volume 42.
4. Dries, D.R.; Dean, D.M.; Listenberger, L.L.; Novak, W.R.P.; Franzen, M.A.; Craig, P.A. An expanded framework for biomolecular visualization in the classroom: Learning goals and competencies. *Biochem. Mol. Biol. Educ.* **2017**, *45*, 69–75. [CrossRef]
5. Brewer, C.A.; Smith, D. *Vision and Change in Undergraduate Biology Education: A Call to Action*; American Association for the Advancement of Science: Washington, DC, USA, 2011.
6. Arneson, J.B.; Offerdahl, E.G. Visual Literacy in Bloom: Using Bloom's Taxonomy to Support Visual Learning Skills. *CBE—Life Sci. Educ.* **2018**, *17*, ar7. [CrossRef]
7. Bateman, R.C.; Craig, P.A. A Proficiency Rubric for Biomacromolecular 3D Literacy. *PDB Newsl. Educ. Corner* **2010**, *45*, 5–7.
8. Mnguni, L.; Schönborn, K.; Anderson, T. Assessment of visualisation skills in biochemistry students. *S. Afr. J. Sci.* **2016**, *112*, 1–8. [CrossRef]
9. Offerdahl, E.G.; Arneson, J.B.; Byrne, N. Lighten the Load: Scaffolding Visual Literacy in Biochemistry and Molecular Biology. *CBE—Life Sci. Educ.* **2017**, *16*, es1. [CrossRef]
10. Jaswal, S.S.; O'Hara, P.B.; Williamson, P.L.; Springer, A.L. Teaching structure: Student use of software tools for understanding macromolecular structure in an undergraduate biochemistry course: Teaching Structure in Undergraduate Biochemistry. *Biochem. Mol. Biol. Educ.* **2013**, *41*, 351–359. [CrossRef]
11. Cohn, N. Visual narrative comprehension: Universal or not? *Psychon. Bull. Rev.* **2020**, *27*, 266–285. [CrossRef]
12. England, B.J.; Brigati, J.R.; Schussler, E.E.; Chen, M.M. Student Anxiety and Perception of Difficulty Impact Performance and Persistence in Introductory Biology Courses. *CBE—Life Sci. Educ.* **2019**, *18*, ar21. [CrossRef]
13. Stylianidou, F. Analysis of science textbook pictures about energy and pupils' readings of them. *Int. J. Sci. Educ.* **2002**, *24*, 257–283. [CrossRef]
14. Forbes-Lorman, R.M.; Harris, M.A.; Chang, W.S.; Dent, E.W.; Nordheim, E.V.; Franzen, M.A. Physical models have gender-specific effects on student understanding of protein structure-function relationships. *Biochem. Mol. Biol. Educ. A Bimon. Publ. Int. Union Biochem. Mol. Biol.* **2016**, *44*, 326–335. [CrossRef]
15. Hoffman, M.; Gneezy, U.; List, J.A. Nurture affects gender differences in spatial abilities. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 14786–14788. [CrossRef]
16. Lee, Y.; Capraro, R.M.; Bicer, A. Gender difference on spatial visualization by college students' major types as STEM and non-STEM: A meta-analysis. *Int. J. Math. Educ. Sci. Technol.* **2019**, *50*, 1241–1255. [CrossRef]
17. Maeda, Y.; Yoon, S.Y. A Meta-Analysis on Gender Differences in Mental Rotation Ability Measured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R). *Educ. Psychol. Rev.* **2013**, *25*, 69–94. [CrossRef]

18. Reilly, D.; Neumann, D.L.; Andrews, G. Gender Differences in Spatial Ability: Implications for STEM Education and Approaches to Reducing the Gender Gap for Parents and Educators. In *Visual-Spatial Ability in STEM Education: Transforming Research into Practice*; Khine, M.S., Ed.; Springer International Publishing: Berlin/Heidelberg, Germany, 2017; pp. 195–224. [CrossRef]

19. Sharobeam, M.H. The Variation in Spatial Visualization Abilities of College Male and Female Students in STEM Fields Versus Non-STEM Fields. *J. Coll. Sci. Teach.* **2016**, *46*, 93–99. [CrossRef]

20. Tsigeman, E.S.; Likhanov, M.V.; Budakova, A.V.; Akmalov, A.; Sabitov, I.; Alenina, E.; Bartseva, K.; Kovas, Y. Persistent gender differences in spatial ability, even in STEM experts. *Heliyon* **2023**, *9*, e15247. [CrossRef]

21. Garrison, L.; Bruckner, S. Considering best practices in color palettes for molecular visualizations. *J. Integr. Bioinform.* **2022**, *19*, 20220016. [CrossRef]

22. Goodsell, D.S.; Jenkinson, J. Molecular Illustration in Research and Education: Past, Present, and Future. *J. Mol. Biol.* **2018**, *430*, 3969–3981. [CrossRef]

23. Hattab, G.; Rhyne, T.-M.; Heider, D. Ten simple rules to colorize biological data visualization. *PLoS Comput. Biol.* **2020**, *16*, e1008259. [CrossRef]

24. Jambor, H.; Antonietti, A.; Alicea, B.; Audisio, T.L.; Auer, S.; Bhardwaj, V.; Burgess, S.J.; Ferling, I.; Gazda, M.A.; Hoeppner, L.H.; et al. Creating clear and informative image-based figures for scientific publications. *PLoS Biol.* **2021**, *19*, e3001161. [CrossRef]

25. Van Merriënboer, J.J.G.; Sweller, J. Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educ. Psychol. Rev.* **2005**, *17*, 147–177. [CrossRef]

26. Linenberger, K.J.; Holme, T.A. Biochemistry Instructors' Views toward Developing and Assessing Visual Literacy in Their Courses. *J. Chem. Educ.* **2015**, *92*, 23–31. [CrossRef]

27. Osborne, R.J.; Wittrock, M.C. Learning science: A generative process. *Sci. Educ.* **1983**, *67*, 489–508. [CrossRef]

28. Kahneman, D. *Thinking, Fast and Slow*; Macmillan: New York, NY, USA, 2011.

29. Gick, M.L. Problem-solving strategies. *Educ. Psychol.* **1986**, *21*, 99–120. [CrossRef]

30. Impara, J.C.; Plake, B.S. Standard Setting: An Alternative Approach. *J. Educ. Meas.* **1997**, *34*, 353–366. [CrossRef]

31. Pérez, E.V.; Santos, L.M.R.; Pérez, M.J.V.; de Castro Fernández, J.P.; Martín, R.G. Automatic classification of question difficulty level: Teachers' estimation vs. students' perception. In Proceedings of the 2012 Frontiers in Education Conference Proceedings, Seattle, WA, USA, 3–6 October 2012; pp. 1–5. [CrossRef]

32. Van de Watering, G.; van der Rijt, J. Teachers' and students' perceptions of assessments: A review and a study into the ability and accuracy of estimating the difficulty levels of assessment items. *Educ. Res. Rev.* **2006**, *1*, 133–147. [CrossRef]

33. Chavan, P.; Mitra, R.; Srree Murallidharan, J. Multiscale nature of student and teacher perceptions of difficulty in a mechanical engineering lecture. *Eur. J. Eng. Educ.* **2022**, *47*, 910–929. [CrossRef]

34. Maynard, D.C.; Hakel, M.D. Effects of Objective and Subjective Task Complexity on Performance. *Hum. Perform.* **1997**, *10*, 303–330. [CrossRef]

35. Anderson, T.R. Bridging the educational research-teaching practice gap: The importance of bridging the gap between science education research and its application in biochemistry teaching and learning: Barriers and strategies. *Biochem. Mol. Biol. Educ. A Bimon. Publ. Int. Union Biochem. Mol. Biol.* **2007**, *35*, 465–470. [CrossRef]

36. Procko, K.; Engelman, S.; Jakubowski, H.; Beckham, J.T.; Dean, D.M.; Franzen, M.A.; Novak, W.R.P.; Roberts, R.; Roca, A.I.; Shor, A.C.; et al. Meeting report: BioMolViz workshops for developing assessments of biomolecular visual literacy. *Biochem. Mol. Biol. Educ.* **2021**, *49*, 278–286. [CrossRef]

37. Ballen, C.J.; Wieman, C.; Salehi, S.; Searle, J.B.; Zamudio, K.R. Enhancing Diversity in Undergraduate Science: Self-Efficacy Drives Performance Gains with Active Learning. *CBE—Life Sci. Educ.* **2017**, *16*, ar56. [CrossRef]

38. Saldaña, J. *The Coding Manual for Qualitative Researchers*; Sage: Thousand Oaks, CA, USA, 2021.

39. MacQueen, K.M.; McLellan-Lemal, E.; Bartholow, K.; Milstein, B. Team-based codebook development: Structure, process, and agreement. *Handb. Team-Based Qual. Res.* **2008**, *119*, 119–135.

40. Auerbach, C.; Silverstein, L.B. *Qualitative Data: An Introduction to Coding and Analysis*; NYU Press: New York, NY, USA, 2003; Volume 21.

41. Broekkamp, H.; Van Hout-Wolters, B.H.A.M.; Van den Bergh, H.; Rijlaarsdam, G. Teachers' task demands, students' test expectations, and actual test content. *Br. J. Educ. Psychol.* **2004**, *74*, 205–220. [CrossRef]

42. Krell, M.; Khan, S.; van Driel, J. Analyzing Cognitive Demands of a Scientific Reasoning Test Using the Linear Logistic Test Model (LLTM). *Educ. Sci.* **2021**, *11*, 9. [CrossRef]

43. Mak, C.H. A Simple Paper Model Illustrates How To Cyclize Monosaccharides from Fischer Projections to Haworth. *J. Chem. Educ.* **2018**, *95*, 1336–1339. [CrossRef]

44. Kryjevskaia, M.; Heron, P.R.L.; Heckler, A.F. Intuitive or rational? Students and experts need to be both. *Phys. Today* **2021**, *74*, 28–34. [CrossRef]

45. Linn, M.C.; Petersen, A.C. Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis. *Child Dev.* **1985**, *56*, 1479–1498. [CrossRef]

46. Bates, R.; Khasawneh, S. Self-efficacy and college students' perceptions and use of online learning systems. *Comput. Hum. Behav.* **2007**, *23*, 175–191. [CrossRef]

47. Penzar, E.C.; Shea, M.; Edwards, C.N. College Students' Academic Achievement: Exploring the Role of Hope and Academic Self-Efficacy. *Int. Dialogues Educ. J.* **2021**, *8*, 1. [CrossRef]

48. Morphew, J.W. Changes in metacognitive monitoring accuracy in an introductory physics course. *Metacogn. Learn.* **2021**, *16*, 89–111. [CrossRef]
49. Kolić-Vehovec, S.; Pahljina-Reinić, R.; Rončević Zubković, B. Effects of collaboration and informing students about overconfidence on metacognitive judgment in conceptual learning. *Metacogn. Learn.* **2022**, *17*, 87–116. [CrossRef]
50. Ehrlinger, J.; Johnson, K.; Banner, M.; Dunning, D.; Kruger, J. Why the Unskilled Are Unaware: Further Explorations of (Absent) Self-Insight Among the Incompetent. *Organ. Behav. Hum. Decis. Process.* **2008**, *105*, 98–121. [CrossRef]