

The Development and Implementation of an Instrument to Assess Students' Data Analysis Skills in Molecular Biology †

Brian J. Rybarczyk^{1*}, Kristen L.W. Walton², and Wendy Heck Grillo³

¹*Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599,*

²*Department of Biology, Missouri Western State University, St. Joseph, MO 64507,*

³*Department of Biology, North Carolina Central University, Durham, NC 27707*

Developing visual literacy skills is an important component of scientific literacy in undergraduate science education. Comprehension, analysis, and interpretation are parts of visual literacy that describe related data analysis skills important for learning in the biological sciences. The Molecular Biology Data Analysis Test (MBDAT) was developed to measure students' data analysis skills connected with scientific reasoning when analyzing and interpreting scientific data generated from experimental research. The skills analyzed included basic skills, such as identification of patterns and trends in data and connecting a method that generated the data, and advanced skills, such as distinguishing positive and negative controls, synthesizing conclusions, determining if data supports a hypothesis, and predicting alternative or next-step experiments. Construct and content validity were established and calculated statistical parameters demonstrate that the MBDAT is valid and reliable for measuring students' data analysis skills in molecular and cell biology contexts. The instrument also measures students' perceived confidence in their data interpretation abilities. As scientific research continues to evolve in complexity, interpretation of scientific information in visual formats will continue to be an important component of scientific literacy. Thus science education will need to support and assess students' development of these skills as part of students' scientific training.

INTRODUCTION

Science literacy encompasses a broad spectrum of knowledge, competencies, and skills (1, 8, 31). Research in science literacy is primarily focused on assessing students' conceptual knowledge and elucidating misconceptions as measured by concept inventories in areas such as natural selection (2), genetics (35), and molecular and cell biology (32). Other instruments have been developed to measure students' attitudes about science (30) and critical thinking and scientific reasoning skills (13, 15, 20, 36). A recent report highlighted an understudied but important part of scientific literacy, how undergraduate students learn from visualizations (33).

Visual literacy is a skill-based competency that involves the comprehension of concepts through visual representations (7). Aspects of visual literacy include describing, analyzing, and interpreting information in the form of graphs, tables, diagrams, and images at the macro level (e.g. photographs of organisms) as well as at the micro level

(e.g. histological sections, illustrations of molecules) (14). Students encounter visual representations in textbooks (9, 21), during classroom lectures, and while reading primary literature (19). A previous study showed that undergraduate biology textbooks are composed of different types of visuals compared with primary research journals and that introductory level textbooks contain significantly fewer visuals that depict experimental data compared to discipline-specific textbooks (27). Thus, students may have restricted opportunities to develop visual literacy skills in experimental research contexts in their undergraduate biology courses.

Studies of how undergraduate students interpret external representations of data are limited and primarily focus on how students interpret graphs and charts (6, 24) and diagrams of molecules (29). Challenges that students face with effective visual interpretation include unfamiliar or non-existent context/reference points and an assumed prior experience with visualizations which are more characteristic of expert interpreters (5). Effective data interpretation requires skills that are demonstrated by expert scientists who are more familiar with discipline-specific representations and conventions (28). Previous studies that investigated graphical interpretation skills of undergraduate students revealed challenges with their abilities to distinguish variables and how to interpret interactions among variables (24). The Test of Scientific Literacy Skills (TOSLS) instrument includes only a subset of questions focused on data interpretation

*Corresponding author. Mailing address: University of North Carolina at Chapel Hill, 211A West Cameron Ave., CB#5492, Chapel Hill, NC 27599. Phone: 919-962-2505. Fax: 919-962-5134. E-mail: brybar@unc.edu.

†Supplemental materials available at <http://jmbe.asm.org>

(15). Thus, additional validated tools are needed to determine how students interpret data representations situated in experimental research contexts.

We describe the development, validation, and implementation of a tool, the Molecular Biology Data Analysis Test (MBDAT). The purpose of the MBDAT is to assess how undergraduate biology major students analyze and interpret experimental data, to elucidate misinterpretations of data, and to measure students' self-perceived confidence in their data-analysis skills. We hypothesize that the MBDAT is a valid and reliable instrument which can differentiate between basic and advanced level data analysis skills among different types of student abilities.

METHODS

Construct and content validation of the instrument

The development process for the MBDAT was similar to previously published processes of development for concept inventories and scientific literacy instruments (15, 32, 36). Visual literacy was used as the theoretical framework for the development of the instrument (7). At least 11 abilities compose visual literacy focused on reading and interpreting skills (4). The most relevant skills from this construct that served as the framework for the MBDAT were critical viewing, visual reasoning, constructing meaning, and knowledge of visual conventions (4). As part of the construct validation process, these skills were validated by expert faculty ($n = 8$) via a survey. Expert reviewers were selected using the criteria that they are currently engaged in active research and have experience teaching college-level coursework in the biological and biomedical sciences. The experts were informed of the purpose of the instrument: to identify student challenges at the beginning of the semester and to measure growth in student abilities over a single semester. The experts were asked first to identify important skills, from a list supplied, that they deemed necessary for students to interpret data in visual formats from experimental contexts and then secondly to rank their top three selections from the skills they selected in step one. The skills which were most frequently selected and ranked among the most important were used to create categories (Table 1) as the basis for designing the MBDAT questions. Skills that were indicated with no or low frequency and subsequently not included in the MBDAT were: 1) construct a visual to explain a concept, 2) spatially manipulate a visual, and 3) visualize orders of magnitude and scale (28). Skills defined from previous studies (28) and additional science process skills connected with interpreting data (26) were used to refine the construct of the MBDAT (Table 1). Basic level skills were defined as those skills that are fundamental to *describing* a visual's surface level characteristics whereas advanced level skills involve *analysis*, *inference*, and *interpretation*, usually requiring synthesis of contextual information

with reasoning skills. The categorization of the defined skills into basic and advanced levels was based on previous studies (28, 39) and classification of educational learning objectives (3), where basic skills equated with lower level thinking skills and advanced skills equated with higher level thinking skills. Basic skills included in the MBDAT were identifying patterns and trends in data and connecting data with a method that generated the data. Advanced skills included in the MBDAT were distinguishing between positive and negative controls, proposing alternative/additional experimental controls, synthesizing conclusions from data, determining whether data supports a hypothesis, and proposing follow-up experiments. Each question was categorized by the authors as a basic or advanced skill level, and faculty experts confirmed that each question aligned with the skills intended in the construct and agreed on the assignment of questions among the two skill levels (Table 1) with an initial consensus of 90%. Discrepancies between the questions that did not initially receive consensus were resolved by discussion to ensure that all questions were categorized appropriately.

Four scientists who defined themselves as molecular and/or cell biologists independently evaluated the instrument and answer key. The four scientist-experts all indicated that the questions aligned with the predetermined set of skills that the instrument was intended to measure and agreed on each question's assignment to the specified categories (Table 1). The scientists had prior experience teaching both introductory and advanced undergraduate students and thus were familiar with students' abilities at these different levels. Experts were asked if they would consider the instrument to be a test of data analysis skills and all experts agreed and deemed the experimental descriptions/contexts to be scientifically accurate. The experts also provided suggestions for improvement that related to consistency, clarity, and precision in the presentation of data displayed in the visuals to ensure that controls indicated were appropriately aligned with the text of the experimental scenarios associated with the questions.

The questions in the instrument were designed in a multiple-choice format and consisted of a brief description of an experimental context along with each visual representation. The use of contexts related to research in human disease in the design of the MBDAT was intentional. These contexts aim to provide more meaningful connection with students' interests in applications of biological concepts to human health and medicine and may help to increase students' motivation (16). The multiple-choice format was chosen because this format can be readily used and graded with both small and large class sizes, and it provides an objective score that is not dependent on students' writing skills or on the grader's expertise. Questions in the initial pool that were not aligned with the skills described in the construct were discarded. Distractors for each question were generated from previous students' responses to

TABLE 1.
Data analysis skills, corresponding questions, and description of distractors.

Questions	Description	Challenges and Misinterpretations
Basic-Level Skills		
B1. Identify patterns and trends in data	2, 6, 9, 10, 11, 17, 18, 19 Describe linear and exponential changes in data as displayed in graphs Describe data presented in table format, data measurements that change over time Determine relationship between variables	Inability to distinguish between changes over time vs. dose dependence in visual Inability to match data results with pattern displayed in visual Describe relationship of variables displayed in visual
B2. Connect data with a method as the source of the data	1, 13 Identify methods that measure amount of macromolecules Identify methods that measure size of macromolecules Recognize conventions of data displays from methods	Distinguish methods that measure amount and size of macromolecules Inability to recognize conventions of data generated from specific methods
Advanced-Level Skills		
A1. Distinguish between positive and negative controls Propose other controls	3, 7 Determine controls inherent in experimental processes	Propose positive or negative controls that are irrelevant or unrelated to experimental contexts
A2. Synthesize conclusions from data within a study Determine whether data supports hypothesis	5, 8, 12, 14, 15 Match data display with experimental results Evaluate data to determine if supports hypothesis	Assume molecules interact if they display similar patterns of changes over time Overinterpret or incorrectly extrapolate data to draw conclusions Provide unsubstantiated conclusions about relationship between cell division, gene, and protein expression
A3. Propose follow-up experiments. Predict results	4, 16, 20 Propose appropriate and logical experiments aligned with experimental context	Propose experiments or variables irrelevant to context Extrapolation of data points does not follow trends predicted in results

open-ended exam questions which were similar in content and validated with student interviews. Questions at the basic and advanced levels were placed in a randomized order so that the instrument did not progress from assumed easier questions to more difficult questions toward the end of the instrument. The final set of questions focused on the interpretation of two digital images, one data table, and three sets of graphs in different formats. Questions that tested students' abilities to interpret graphs were similar in design to a previously developed instrument (22) but were modified and situated in a different context. The final instrument consisted of ten questions that assessed basic skills (B1 and B2) and ten questions that assessed advanced skills (A1, A2, A3, Table 1). Seventeen of the 20 questions included a choice of "I don't know" to provide students the opportunity to indicate that they do not know the correct answer instead of guessing among the given choices. In the calculations of item difficulty, the selection of "I don't know" was scored as an incorrect choice (see Appendix 2 for complete instrument).

Student interviews and implementation

The instrument was piloted with advanced undergraduate Biology major students ($n = 32$) and brief interviews were conducted to probe students about confusing word choices, to help determine whether the questions were appropriate for biology majors, and to verify the classification of the questions as basic or advanced level skills (Table 1). Seventeen students who had participated in at least one semester of hands-on laboratory-based research experiences and/or who were currently engaged in experiential research completed the instrument. Four of these students were asked to review and provide additional feedback on the MBDAT to identify questions that were particularly challenging, identify confusing language, and describe whether the content was aligned with concepts in molecular and cell biology that Biology majors are expected to know.

Students' interviews confirmed the validity of distractors since students identified more than one possible answer. Comments such as "doesn't choice (b) also support

the conclusion, but it may not be the best choice” and “I was stuck between two choices” and “can any of these conclusions be made from the data?” support this claim. Students also confirmed that the questions related to describing patterns in data required more basic skills and the other questions that asked for higher level, science reasoning skills were correctly designated as advanced level skills. In contrast to other instruments (35), some wording such as positive and negative controls, terminology such as mRNA, and wording related to scientific inquiry were maintained in the questions since these aspects of science literacy were defined in the construct of the MBDAT and students confirmed that these were known aspects related to scientific experimentation.

The instrument was administered in paper or online format during the first week of a 15-week semester as a pretest and during the last week of the semester as a posttest. Time on task for both formats averaged 20 minutes for completion. The study participants included students from a large public research extensive university, a minority-serving institution, and a public, primarily undergraduate university. A total of 127 upper-level and 73 introductory-level students consented to participate in the study. All upper-level undergraduate students were majoring in science or health sciences at the junior and senior level and completed prerequisite courses that included content in cell biology, molecular biology, and genetics. Only data from students who completed both the pre- and posttests were included in the analysis (94 upper-level students and 40 introductory-level students). Significance was assessed with paired *t*-test using $\alpha < 0.05$. To determine reliability, the coefficient of stability was calculated using pretest scores from two consecutive semesters at the same institution, which resulted in a coefficient of stability of $r = 0.93$ and comparing pretest scores from two different institutions, which resulted in a coefficient of stability of $r = 0.96$. Secondly, the distribution of responses for each question was analyzed using a χ^2 test and was not found to be significantly different between the two consecutive semesters of pretest administration ($p > 0.05$) on 17 of the 20 questions. Differences in the distribution of responses were noted on Question 12 ($p = 0.012$), Question 14 ($p = 0.014$), and Question 19 ($p = 0.038$). The overall internal reliability of the instrument was 0.66 as assessed by Cronbach's alpha coefficient, an acceptable level (4).

Item difficulty, item discrimination index, and reliability analyses were performed as previously described (2, 12, 32, 35). Many of the questions in the basic level (BI) revealed item difficulty measurements above 90%. It was justifiable to maintain these items in the final iteration of the instrument since these questions were related to students' abilities to describe patterns and trends in data from a variety of different types of visual displays beyond standard line graphs (24, 25) and thus provide additional evidence and an opportunity to measure the skill of identifying patterns in data as depicted in different types of visual representations not previously studied.

Based on the advanced level of the student population tested and the distribution of pretest scores, we defined high-performing students as the top 25% and low-performing students as the bottom 25% on each of the instruments' implementations and calculated the item discrimination index as $D = (N_H - N_L)/(N/4)$ where N_H is the number of high performing students who scored correctly, N_L is the number of low performing students who scored correctly, and N is the total number of students who completed the test (12). Using the 25%-25% calculation reduces the chances of underestimating the *D* values for this population (11).

For the 17 questions that included “I don't know” as an answer choice, rates of uncertainty were calculated for each individual question by counting the number of “I don't know” responses and dividing by the total number of students ($n = 94$). A total rate of uncertainty was also calculated for the group of 10 basic- and group of 10 advanced-level questions. Since seven of the basic-level questions had the option of “I don't know,” this number was multiplied by the total number of students, 94, to give a total of 658 possible responses of “I don't know.” The number of “I don't know” responses for these seven basic-level questions was counted and divided by 658, which is the total number of “I don't know” responses possible. All ten advanced-level questions had an option for a response of “I don't know”; thus the total number of “I don't know” responses possible for the group of advanced-level questions was 940. The number of “I don't know” responses for these ten advanced-level questions was counted and divided by 940 to give a rate of uncertainty for the group of advanced-level questions.

A misinterpretation index (MI), as previously described, was calculated according to the following formula, where MFIA = most frequent incorrect answer (23):

$$MI = 1 - \frac{(\# \text{ of students selecting any incorrect answer} - \# \text{ of students selecting MFIA})}{\# \text{ of students selecting any incorrect answer}}$$

The original implementation of this formula was used to indicate a prominent misconception among distractors provided on concept inventories. In the context of the MBDAT, the formula was used to identify the strength of misinterpretations and distractors used in the questions which had item difficulty measures ≤ 0.7 .

This study was approved by the Institutional Review Boards at UNC Chapel Hill (study #05-0148), Missouri Western State University (study #783), and North Carolina Central University (study #1200888). All research has complied with relevant federal and institutional policies.

RESULTS

Upper-level students earned an average score of 74.7% on the pretest and an 81.7% on the posttest (Table 2) indicating that the MBDAT can detect learning gains over the course of a semester. A majority of students, 83%, demon-

TABLE 2.
Mean pretest and posttest scores for tested groups.

	<i>n</i>	Mean Pretest (\pm SE), %	Mean Posttest (\pm SE), %
Upper-level students	94	74.7 (\pm 1.3)	81.7 (\pm 1.1)
Research students	17	85.0 (\pm 2)	89.0 (\pm 1)
Introductory Biology students	40	55.5 (\pm 1.9)	50.0 (\pm 1.6)

strated equal or positive learning gains comparing pre- to posttest scores. Students who had at least one semester of research experience and who were currently involved in a research experience were selected to complete the MBDAT as a comparison group. The research students' average pretest score was 85%, which was higher and statistically different (Student's *t*-test, $p = 0.002$) than the pretest scores for the upper-level students group and showed additional gains on the posttest (Table 2). As expected, a comparison group of introductory biology students scored lower on the pretest (55.5%) and posttest (50%), which were significantly lower as compared to the pretest ($p < 0.001$) and posttest ($p < 0.001$) scores of the upper-level students. These results suggest that the instrument was more difficult for introductory-level students, most likely since they did not have the prerequisite content and were sampled from a course that did not emphasize data analysis skills. Together, these results demonstrate that the MBDAT can measure different abilities of different types of students.

For the upper-level students, item difficulty (*P*) values for all questions fell between 0.35 and 0.99 (Fig. 1). This range of *P* values is similar to the accepted range of 0.3 to 0.9, indicating that the questions were reliable measures for different students' abilities (12). Learning gains for individual questions ranged from 0 to 18 percentage points (Fig. 1). Analysis of these gains on a per-question basis revealed that seven questions had significant gains comparing pre- to posttest results (Fig. 1).

We anticipated that upper-level students should already possess several well-developed data analysis skills. Students demonstrated high proficiency with basic level skills indicated by *P* values greater than 0.7 on both pre- and posttest questions (Fig. 1, Questions 6, 9, 10, 11, 17, 18, 19). Since these questions tested students' abilities to identify patterns and trends in data displayed in different formats, the results suggest that the students included in our study are adept at interpreting patterns in data as represented in a variety of formats including line graphs, bar graphs, and tables. Similarly, students demonstrated high proficiency with some skills in the advanced level (Fig. 1, Questions 3, 5, 12, 20). These advanced-level questions tested students' abilities to propose an appropriate control for an experiment and to evaluate whether data supported a hypothesis or a conclusion.

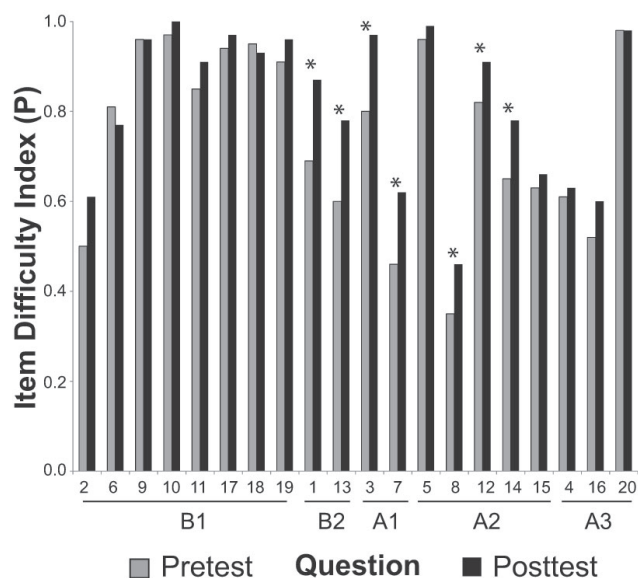


FIGURE 1. Students' performance on pre- and posttest questions as measured by item difficulty (*P*) for each question. Gray bars represent average pretest *P* and black bars represent average posttest *P*. $n = 94$. * $p < 0.05$.

Questions with *P* values equal to or less than 0.7 on both the pre- and posttests were designated as skills that students found challenging. As predicted, the results revealed greater skill deficiencies in the advanced-level category (five questions) as compared to the basic-level category (one question) (Fig. 1). The calculated *P* was less than 0.7 on one basic level question (B1 Question 2) on both the pre- and posttests. This question tested students' ability to analyze changes in gene expression levels changing in a dose response pattern. *P* was equal to or less than 0.7 on five of the ten questions in the advanced-level skills category on both the pre- and posttests (Fig. 1, Questions 4, 7, 8, 15, 16). These questions tested students' abilities to propose a follow-up experiment, propose an appropriate experimental control, and draw conclusions from data.

All calculated *D* values ranged from 0 to 0.67 with an average of 0.32 (Fig. 2). An item discrimination index of ≥ 0.3 suggests that overall, the questions provided good discrimination within this population (12). Since *D* is dependent on item difficulty (*P*), several questions revealed lower discriminatory power, as expected, particularly the basic-level questions with an average *D* value of 0.27 on the pretest and 0.20 on the posttest. For the advanced-level questions, average *D* values were higher, 0.36 on the pretest and 0.30 on the posttest (Fig. 2). This range of *D* values indicates that some questions were correctly answered by a majority of only the high-performing students (high *D* value on both the pre- and posttests), while other questions were correctly answered on the posttest by lower-performing students (high *D* value on pretest and lower on posttest), and there were questions correctly answered by all students (low *D* value on both pre- and posttests).

For each question, the proportion of students selecting the “I don’t know” response was used as an indicator of students’ confidence in their ability to identify the correct answer and represents a level of uncertainty. The overall rate of uncertainty was greater across the group of basic level questions on both the pretest (16.3%) and the posttest (16.9%) as compared to the group of advanced-level skills on the pretest (4.8%) and on the posttest (1.6%) (Fig. 3A). Analysis of individual questions indicated a range of uncertainty of 1% to 17% across all 17 questions (Fig. 3B). These results suggest that students had a high level of confidence

indicated by relatively low rates of selecting the “I don’t know” option, particularly with questions that measured advanced-level skills. However, students’ actual performance on the advanced-level questions did not match the perceptions of their skills as indicated by lower item difficulty measures on the advanced-level skills as compared with the basic-level skills (Fig. 1).

Many students’ misconceptions in science have been identified in disciplinary areas related to specific concepts (2, 18, 23). However, little is known about what misconceptions or misinterpretations undergraduate biology major students may exhibit while analyzing data. A misinterpretation index (MI) was calculated for the six questions which had a calculated item difficulty of 0.7 or less on both the pre- and posttests (Fig. 1, Questions 2, 4, 7, 8, 15, 16). A strong MI was calculated for Question 2 on both the pretest (0.86) and the posttest (0.97) (Table 3). These results indicate that most students selected the same incorrect answer which included wording about changes in gene expression patterns *over time* rather than the correct answer which related changes in gene expression patterns with a *dose response*. For the five advanced-level questions, the calculated MIs ranged between 0.37 and 0.71, indicating that although there may have been a predominant wrong answer, students selected among all of the wrong answers (Table 3). These data also provide additional evidence for the quality of distractors used in these questions since students selected options from among all of the answer choices. One skill that students were challenged with was identifying an inappropriate control for a described experiment (Question 7). The MFIA that students selected described a correct control that measures amounts of an unrelated reference protein. A second skill that challenged students was proposing a conclusion given an experimental

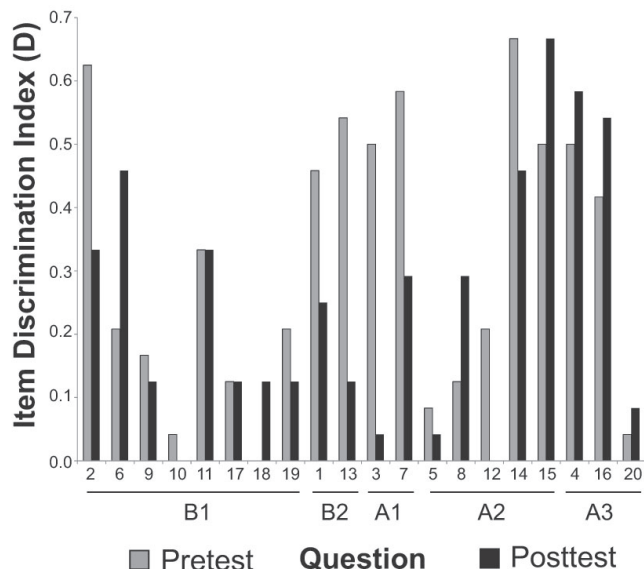


FIGURE 2. Item discrimination index. Gray bars represent pretest D and black bars represent posttest D. $n = 94$.

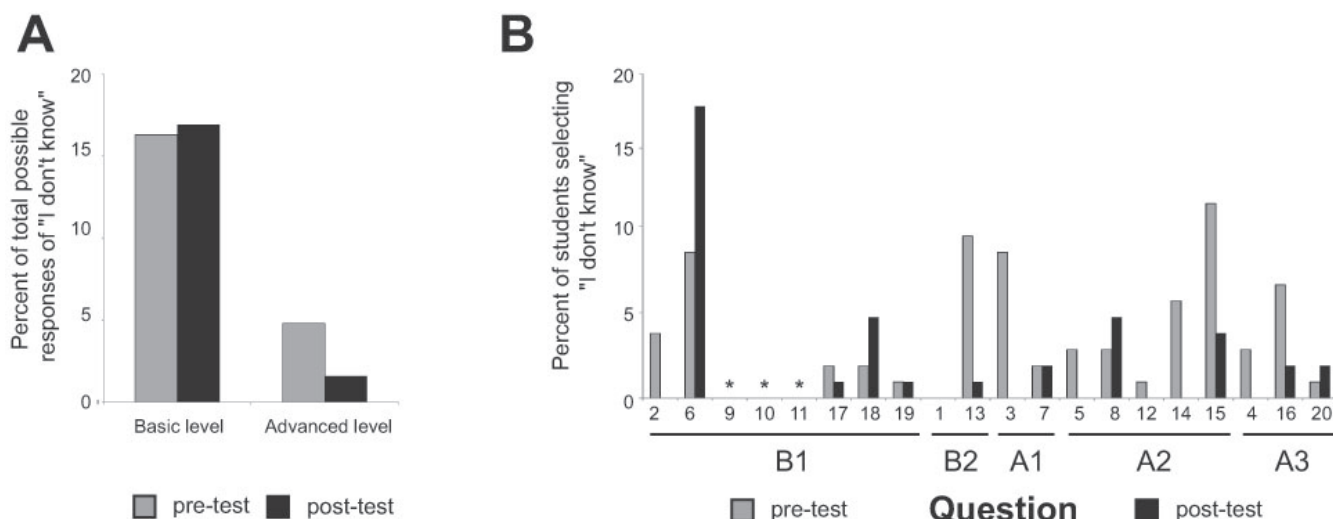


FIGURE 3. Students’ confidence in their data analysis skills. A) percent of overall responses of “I don’t know” on basic- and advanced-level questions on pre- and posttest. B) percent of students who responded “I don’t know” on each question of the pre- and posttest. * Indicates that these questions did not have an “I don’t know” response option. Gray bars represent pretest percentages and black bars represent posttest percentages. $n = 94$.

TABLE 3.

Analysis of most frequent incorrect answer (MFIA) for six challenging questions. Misinterpretation Index (MI) on pre- and posttest are indicated.

Question	Skill	Answer	MFIA	MI pre	MI post
2	describe patterns in data	D	B	0.86	0.97
4	propose next experiment	A	B to D*	0.38	0.46
7	propose experimental control	A	B	0.57	0.68
8	generate conclusions	C	B	0.74	0.83
15	generate conclusions	C	E to D*	—	0.52
16	propose next experiment	A	D	0.40	0.42

* Indicates that the MFIA was different between pre- and posttest.

context (Questions 8 and 15). For Question 8, students selected a MFIA which stated that these two proteins interact with each other; however, the data presented were not a measurement of protein interaction. An MI was not calculated for Question 15 on the pretest since the predominant answer was “I don’t know” (Table 3). The predominant wrong answer selected for Question 15 on the posttest included only part of the original hypothesis being supported by the data presented. This change from “I don’t know” on the pretest to the wrong answer on the posttest indicated that students increased their confidence in answering the question but they still selected an incorrect answer. The third skill that challenged students was proposing next-step experiments (Questions 4 and 16). The calculated MIs for these two questions were lower than the MIs for other questions analyzed (Table 3), suggesting that students selected various incorrect responses that were not logical to propose as follow-up experiments in the context given.

DISCUSSION

The MBDAT was developed and tested with undergraduate biology major students resulting in a range of item difficulty and item discrimination index measures, suggesting that it is a valid tool for use with this student population of biology majors. This investigation adds value to the field of visual literacy since it provides an instrument to measure how upper-level biology major students interpret experimental data. This study used both domain-specific visualizations from molecular biology (i.e. images of gels) (37) and non-domain specific visualizations such as graphs

and tables, adding complexity to the types of visualizations used in the instrument. Thus, the MBDAT provides a broader assessment of scientific reasoning skills and a way to measure students’ data analysis abilities within experimental contexts, likening it to contexts found in primary literature (8, 10, 17, 19, 38).

We identified areas of scientific reasoning connected to visual literacy that are challenging for upper-level students. Students were particularly challenged with the advanced-level skill of proposing follow-up experiments as indicated by low P values (high difficulty) for Questions 4 and 16 (Fig. 1), as expected for students who do not explicitly practice these skills as part of a course (34). Taken together with the previously described low rates of “I don’t know” responses for these two questions (Fig. 3B), these data suggest that students are confident in their abilities but demonstrate these skills unsuccessfully and may need more time than one semester to develop these skills.

The current formulation of the MBDAT does have several limitations. First, it includes data representations in the context of molecular and cell biology. Although the MBDAT’s structure can be useful in creating similar instruments that measure visual literacy, it may not be as useful or valid to measure students’ skills outside of these disciplinary contexts. Another limitation is that the MBDAT’s construction also assumes that students will self-identify with the “I don’t know” option if they do not know an answer rather than guessing an answer from the options provided. Lastly, the MBDAT does not have the power to measure every skill necessary for complete visual literacy (4). The contexts in the current iteration of this MBDAT focus on only those aspects of visual literacy defined in the construct (Table 1).

The results from this study identify several scientific reasoning skills connected to visual literacy that instructors could focus on and provide opportunities for students to develop during instruction related to analyzing and interpreting data generated from experimental research. Upper-level undergraduate students demonstrated proficient skills at the defined basic level but need additional support in developing advanced-level skills. Thus, creating learning environments and opportunities to help students develop visual literacy are important educational goals for undergraduate science courses.

SUPPLEMENTAL MATERIALS

Appendix I: Molecular biology data analysis test – instrument and answer key

ACKNOWLEDGMENTS

Special thanks to students and faculty (especially J. Tenlen, S. Alvarez, J. Shaffer, K. Monahan) who provided input in the development and modification of the instrument.

The authors declare that there are no conflicts of interest. The authors acknowledge that there were no other individuals who provided writing or other assistance. The

investigators disclose that there were no potential conflicts to study participants. There were no other study sponsors involved in the study design; in the collection, analysis, and interpretation of the data; in the writing of the report; and in the decision to submit the report for publication.

REFERENCES

1. **American Association for the Advancement of Science.** 2011. Vision and change in undergraduate biology education: a call to action: a summary of recommendations made at a national conference organized by the American Association for the Advancement of Science, July 15–17, 2009. Washington, DC.
2. **Anderson, D. L., K. M. Fisher, and G. J. Norman.** 2002. Development and evaluation of the conceptual inventory of natural selection. *J. Res. Sci. Teach.* **39**:952–978.
3. **Anderson, L. W. E., et al.** 2001. A taxonomy for learning, teaching and assessing: a revision of Bloom's Taxonomy of educational outcomes: complete edition. Longman, New York.
4. **Avgerinou, M. D.** 2007. Towards a visual literacy index. *J. Visual Literacy* **27**(1):29–46.
5. **Bowen, G. M., and W.-M. Roth.** 2002. Why students may not learn to interpret scientific inscriptions. *Res. Sci. Educ.* **32**:303–327.
6. **Bowen, G. M., W.-M. Roth, and M. K. McGinn.** 1999. Interpretations of graphs by university biology students and practicing scientists: toward a social practice view of scientific representation practices. *J. Res. Sci. Teach.* **36**:1020–1043.
7. **Brill, J. M., K. Dohun, and R. M. Branch.** 2007. Visual literacy defined—the results of a Delphi study: can IVLA (operationally) define visual literacy. *J. Visual Literacy* **27**:47–60.
8. **Coil, D., M. P. Wenderoth, M. Cunningham, and C. Dirks.** 2010. Teaching the process of science: faculty perceptions and an effective methodology. *CBE Life Sci. Educ.* **9**:524–535.
9. **Cook, M.** 2008. Students' comprehension of science concepts depicted in textbook illustrations. *Electronic J. Sci. Educ.* **12**:1–14.
10. **DeBurman, S. K.** 2002. Learning how scientists work: experiential research projects to promote cell biology learning and scientific process skills. *Cell Biol. Educ.* **1**:154–172.
11. **Ding, L., R. Chabay, B. Sherwood, and R. Beichner.** 2006. Evaluating an electricity and magnetism assessment tool: brief electricity and magnetism assessment. *Phys. Rev. Spec. Topics—Phys. Educ. Res.* **2**:010105.
12. **Doran, R.** 1980. Basic measurement and evaluation of science instruction. NSTA, Washington, DC.
13. **Facione, P. A.** 1991. Using the California Critical Thinking Skills Test in research, evaluation, and assessment. California Academic Press, Millbrae, CA.
14. **Gilbert, J.** 2008. Visualization: an emergent field of practice and enquiry in science education. In Gilbert, J., Reiner, M., and Nakhleh, M. (ed.), *Visualization: Theory and Practice in Science Education*, Vol. 3. Springer, Netherlands.
15. **Gormally, C., P. Brickman, and M. Lutz.** 2012. Developing a Test of Scientific Literacy Skills (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE Life Sci. Educ.* **11**:364–377.
16. **Guilford, W. H.** 2005. "Shrink wrapping" lectures: teaching cell and molecular biology within the context of human pathologies. *Cell Biol. Educ.* **4**:138–142.
17. **Hoskins, S. G., D. Lopatto, and L. M. Stevens.** 2011. The C.R.E.A.T.E. approach to primary literature shifts undergraduates' self-assessed ability to read and analyze journal articles, attitudes about science, and epistemological beliefs. *CBE Life Sci. Educ.* **10**:368–378.
18. **Klymkowsky, M. W., K. Garvin-Doxas, and M. Zeilik.** 2003. Bioliteracy and teaching efficacy: what biologists can learn from physicists. *Cell Biol. Educ.* **2**:155–161.
19. **Krontiris-Litowitz, J.** 2013. Using primary literature to teach science literacy to introductory biology students. *J. Microbiol. Biol. Educ.* **14**:66–77.
20. **Lawson, A. E.** 1978. The development and validation of a classroom test of formal reasoning. *J. Res. Sci. Teach.* **15**:11–24.
21. **Leivas Pozzer, L., and W.-M. Roth.** 2003. Prevalence, function, and structure of photographs in high school biology textbooks. *J. Res. Sci. Teach.* **40**:1089–1114.
22. **McKenzie, D., and M. Padilla.** 1986. The construction and validation of the Test of Graphing in Science (TOGS). *J. Res. Sci. Teach.* **23**:571–579.
23. **Nazario, G. M., P. A. Burrowes, and J. Rodriguez.** 2002. Persisting misconceptions. *J. Coll. Sci. Teach.* **31**:292–296.
24. **Picone, C., J. Rhode, L. Hyatt, and T. Parshall.** 2007. Assessing gains in undergraduate students' abilities to analyze graphical data. *Teach. Issues Exper. Ecol.* **5**:1–54.
25. **Roth, W.-M.** 2002. Reading graphs: contributions to an integrative concept of literacy. *J. Curriculum Studies* **34**:1–24.
26. **Roth, W.-M., and A. Roychoudhury.** 1993. The development of science process skills in authentic contexts. *J. Res. Sci. Teach.* **30**:127–152.
27. **Rybarczyk, B. J.** 2011. Visual literacy in biology: a comparison of visual representations in textbooks and journal articles. *J. Coll. Sci. Teach.* **41**:106–114.
28. **Schonborn, K. J., and T. R. Anderson.** 2010. Bridging the educational research-teaching practice gap. *Biochem. Mol. Biol. Educ.* **38**:347–354.
29. **Schonborn, K. J., and T. R. Anderson.** 2009. A model of factors determining students' ability to interpret external representations in biochemistry. *Int. J. Sci. Educ.* **31**:193–232.
30. **Semsar, K., J. K. Knight, G. Birol, and M. K. Smith.** 2011. The Colorado Learning Attitudes about Science Survey (CLASS) for use in biology. *CBE Life Sci. Educ.* **10**:268–278.
31. **Shamos, M.** 1995. *The myth of scientific literacy.* Rutgers University Press, New Brunswick, NJ.
32. **Shi, J., W. B. Wood, J. M. Martin, N. A. Guild, Q. Vicens, and J. K. Knight.** 2010. A diagnostic assessment for introductory molecular and cell biology. *CBE Life Sci. Educ.* **9**:453–461.

33. **Singer, S. R., N. R. Nielsen, and H. Schweingruber.** 2012. Discipline-based education research: understanding and improving learning in undergraduate science and engineering. The National Academies Press, Washington, DC.
34. **Sirum, K., and J. Humburg.** 2011. The Experimental Design Ability Test (EDAT). *Bioscene* **37**:8–16.
35. **Smith, M. K., W. B. Wood, and J. K. Knight.** 2008. The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. *CBE Life Sci. Educ.* **7**:422–430.
36. **Sundre, D.** 2008. The Scientific Reasoning Test, Version 9 (SR-9). Center for the Advancement and Research Studies, Harrisonburg, VA.
37. **Tibell, L. A. E., and C.-J. Rundgren.** 2010. Educational challenges of molecular life science: characteristics and implications for education and research. *CBE Life Sci. Educ.* **9**:25–33.
38. **Wenk, L., and L. Tronsky.** 2011. First-year students benefit from reading primary research articles. *J. Coll. Sci. Teach.* **40**:60–67.
39. **White, B., M. Stains, M. Escriu-Sune, E. Medaglia, L. Rostamjad, C. Chinn, and H. Sevian.** 2011. A novel instrument for assessing students' critical thinking abilities. *J. Coll. Sci. Teach.* **40**:102–107.