Current Updates in Bioinformatics

Research Article Open Access

Developing and Assessing an Instructional Definition of Bioinformatics [Version 1, 1 Approved, 1 Approved with Reservations]

William E Tapprich[1], Letitia Reichart[2], Dawn M Simon[2], Garry Duncan[3], William McClung[4], Nealy F Grandgenett[5], and Mark A Pauley[6*]

[1]Department of Biology, University of Nebraska at Omaha, USA
[2]Department of Biology, University of Nebraska at Kearney, USA
[3]Biology Department, Nebraska Wesleyan University, USA
[4]Mathematics and Computer Science Department, Nebraska Wesleyan University, USA
[5]Department of Teacher Education, University of Nebraska at Omaha, USA
[6]School of Interdisciplinary Informatics, University of Nebraska at Omaha, USA

[*]**Corresponding author:** Mark A Pauley, 6001 Dodge Street, School of Interdisciplinary Informatics, University of Nebraska at Omaha, Omaha, NE 68182-0116, USA, Tel: (402) 403-0361, Fax: (402) 554-3284; Email: mpauley@unomaha.edu

Original Submission

Received: April 02, 2017
Accepted: April 18, 2017
Published: April 28, 2017
Last Updated: October 13, 2017

Open Peer Review Status: 1 Approved, 1 Approved with Reservations

How to cite this article: William E Tapprich, Letitia Reichart, Dawn M Simon, Garry Duncan, William McClung, Nealy F Grandgenett, Mark A Pauley. Developing and Assessing an Instructional Definition of Bioinformatics [Version 1, 1 Approved, 1 Approved with Reservations]. Curr Updates Bioinform (2017) 1: 1.1

Acknowledgments: This work was supported by National Science Foundation award #1122971. The authors acknowledge the help of students at Nebraska Wesleyan University, the University of Nebraska at Kearney, and the University of Nebraska at Omaha for providing responses to the “What is Bioinformatics?” prompt. We further acknowledge the help of scorers at all three institutions.

Abstract

The lack of a common instructional definition for bioinformatics impedes progress within the discipline and delays the effective integration of bioinformatics into biology coursework. Using an iterative process, our team of biologists, a mathematician/computer scientist, and a bioinformatician together with an educational evaluation and assessment specialist, has developed an instructional definition of the discipline: Bioinformatics is "an interdisciplinary field that is concerned with the development and application of algorithms that analyze biological data to investigate the structure and function of biological polymers and their relationships to living systems." The field is defined in terms of its two primary foundational disciplines (biology and computer science) and reflects its interdisciplinary nature. We have also created a rubric for assessing open-ended responses to a prompt about what bioinformatics is and how it is used. The rubric has been shown to be reliable in successive rounds of testing. The instrument's final interrater reliability was demonstrated using a score percent agreement procedure (89.7%) and an intraclass correlation procedure (0.620). We offer the instructional definition and associated assessment rubric to life sciences instructors to help further integrate bioinformatics into biology instruction, as well as for fostering further educational research projects.

Keywords

Bioinformatics; Computational Biology; Definition; Undergraduate Life Sciences Education; Undergraduate Biology Education; Assessment Rubric

Introduction

The widespread use of molecular sequences to address questions across the life sciences has triggered a rapid change in the tools needed to conduct research. While this change is not yet reflected in many undergraduate curricula [1], it is likely to become a central component in the future. Such a change will face multiple challenges, not the least of which is assessment.

Historically, teaching disciplinary content knowledge and skills without students having a clear conceptual understanding of the field as a whole has been a problem in science, technology, engineering, and mathematics (STEM) disciplines. For example, algebra and other mathematics courses have long suffered from definitional problems in an instructional context [2,3]. A parallel situation exists in the evolution of computer science as a discipline—the definition of the field is still a problem, with the "What is Computer Science" definitional context itself limiting student understanding of the field's implications, applications, and importance [4,5]. Indeed, the lack of a clear understanding of what computer science is by school administrators has been pointed to as one of the key reasons for its slow adoption into K-12 (kindergarten through high school) education [6]. Similar problems exist in other disciplines. For example, evidence shows students have little understanding of agreed-upon definitions in environmental and geographic education [7] and evolutionary biology [8]. In short, understanding definitional contexts is critical to effective STEM instruction in general [9,10].

Although bioinformatics is frequently described in the literature, to the best of our knowledge only two definitions of the field have been formally proposed (Table 1). In 2000 a National Institutes of Health (NIH) working group provided definitions of both bioinformatics and the closely related discipline of computational biology [11]. Then, in their 2001 paper "What is bioinformatics? A proposed definition and overview of the field" Luscombe, et al. [12] proposed a formal definition of the field. These were clearly research-level definitions and not well-suited for use in instruction, particularly at the undergraduate level. Furthermore, in a rapidly-changing field like bioinformatics, these definitions may be inadequate as they predate recent scientific breakthroughs in genomics, transcriptomics, and proteomics, etc. Although definitions of bioinformatics can be found in expected places like dictionaries and websites such as Wikipedia, these definitions are often simplistic and often stress widely different aspects of the field (Table 1). Recently, the Curriculum Task Force of the International Society for Computational Biology's Education Committee [13,14] defined different bioinformatics "personas" and specified core competencies for bioinformatics. Although this helped narrow the understanding gap, it did not provide a concise definition. To summarize, and as exemplified by recent publications by Vincent and Charette [15] and Smith [16], there is persistent confusion about both what bioinformatics is and what bioinformaticians do.

As mentioned above, historically it has been a problem when a STEM discipline is not well-defined, and this is currently a situation that bioinformatics faces. In addition, many have expressed the need to integrate bioinformatics into life sciences education [17,18]—which will be difficult to accomplish without a common understanding of what is meant by the term, particularly in an educational context. To address this need, we propose here an instructional definition collaboratively developed by the authors. The instructional focus of the definition is important since lengthy, detailed, and typically specialized research-oriented definitions are often problematic for student learners. These contexts usually require a more general definition where knowledge transmission is "scaffolded" to the level of the student [19,20,21]. Instructors often need to modify or enhance research-oriented definitions for effective classroom use. Our definition is intended to guide bioinformatics instruction by having a more classroom-ready definition available to instructors.

Table 1: Select Definitions of Bioinformatics in the Literature.

Definition	Source
Research, development, or application of computational tools and approaches for expanding the use of biological, medical, behavioral, or health data, including those to acquire, store, organize, archive, analyze, or visualize such data.	Huerta M, et al. NIH Working Definition of Bioinformatics and Computational Biology. The Biomedical Information Science and Technology Initiative Consortium (BISTIC) Definition Committee of National Institutes of Health (NIH). 2000. http://www.bisti.nih.gov/docs/CompuBioDef.pdf
Bioinformatics is conceptualising biology in terms of molecules (in the sense of Physical chemistry) and applying "informatics techniques" (derived from disciplines such as applied maths, computer science, and statistics) to understand and organise the information associated with these molecules, on a large scale. In short, bioinformatics is a management information system for molecular biology and has many practical applications.	Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. Methods Inf Med. 2001; 40: 346-358.
Bioinformatics is an interdisciplinary field that develops methods and software tools for understanding biological data. As an interdisciplinary field of science, bioinformatics combines computer science, statistics, mathematics, and engineering to analyze and interpret biological data.	Bioinformatics. (n.d.). In Wikipedia. Retrieved 27 October 2016 from https://en.wikipedia.org/wiki/Bioinformatics
Any use of computers to characterize the molecular components of living things (in Bioinformatics as a biological science).	Bioinformatics. (n.d.). In Bioinformatics.org, Retrieved 27 October 2016.
Information technology as applied to the life sciences, especially the technology used for the collection and analysis of genomic data.	Bioinformatics. (n.d.). In The American Heritage Science Dictionary. Retrieved October 27, 2016 from http://www.dictionary.com/browse/bioinformatics
The branch of science concerned with information and information flow in biological systems, esp. the use of computational methods in genetics and genomics.	Bioinformatics. OED Online. Retrieved 27 October 2016 from http://www.oed.com/view/Entry/255935?redirectedFrom=bioinfor matics
The collection, classification, storage, and analysis of biochemical and biological information using computers especially as applied to molecular genetics and genomics.	Bioinformatics. (n.d.). In Merriam-Webster.com. Retrieved 27 October 2016 from http://www.merriam-webster.com/dictionary/bioinformatics.
The retrieval and analysis of biochemical and biological data using mathematics and computer science, as in the study of genomes.	Bioinformatics. (n.d.). In Dictionary.com Unabridged. Retrieved 27 October 2016 from http://dictionary.reference.com/browse/bioinformatics

Our recognition of the need for an instructional definition for bioinformatics originates in a joint project undertaken by the authors. This project, *Integrating Bioinformatics into the Life Sciences*, is focused on the development of bioinformatics curricular material for undergraduate biology education at three diverse institutions in Nebraska: Nebraska Wesleyan University (NWU), the University of Nebraska at Kearney (UNK), and the University of Nebraska at Omaha (UNO). (See Acknowledgments.) The specific intent of our project is to permanently modify the biology curriculum at the three participating institutions by vertically integrating a set of bioinformatics-focused laboratories through several layers of complexity, ranging from high school to undergraduate seniors. The seven-member project team is diverse with complementary expertise and consists of four biologists with varying specialties—genetics, molecular biology, evolutionary biology, ecology, virology, and environmental science—a mathematician/computer scientist, a bioinformatician, and an educational evaluation and assessment specialist. The evaluation/assessment specialist has significant experience with assessing student learning in STEM contexts, with more than 130 publications associated with STEM program evaluation. As the project progressed, we realized each team member had his/her own understanding of bioinformatics and that our goal of integrating bioinformatics into the biology curricula at the three participating institutions would be hindered without an organizing definition. We agreed that an instructional definition that reflected a consensus understanding of the field was needed for us to proceed effectively. Furthermore, we felt that a common definition would help us better assess our own work and would provide a focus for developing instructional materials at different levels. Such a definition could also be used by others working to integrate bioinformatics into life sciences education [18] and in future educational research projects.

As described in detail below (see Methods), development of our definition began with a lengthy group discussion. In this discussion, we considered the facets of bioinformatics we felt were both central to the field as a stand-alone discipline, and crucial to its application in biology. In the end, we agreed that bioinformatics should be defined in terms of its two primary foundational disciplines (biology and computer science) but also acknowledged that bioinformatics is a discipline in and of itself. The definition we arrived at is:

"An interdisciplinary field that is concerned with the development and application of algorithms that analyze biological data to investigate the structure and function of biological polymers and their relationships to living systems"

The definition highlights what we believe are the five fundamental aspects of bioinformatics. These five aspects, as follows, are reflected in an assessment rubric that was developed in parallel with the definition (Table 2).

- **Interdisciplinary nature:** Although biology and computer science are clearly of central importance to bioinformatics, bioinformatics also adopts concepts from other disciplines such as mathematics and chemistry. It is thus interdisciplinary in nature and is a field into and of itself.

- **Development and use of computer algorithms:** Although computer algorithms are important to bioinformatics, it is not necessary to have the ability to develop new tools to apply bioinformatics effectively. In this respect, our definition takes into account the recent work by Welch, et al. [13,14] involving the development of different bioinformatics personas.

- **Analysis of data:** Implicit in our definition is that bioinformatics generally involves the analysis of large quantities of data; i.e., bioinformatics is inherently "big data."

- **Molecular-level analysis:** Bioinformatics most often involves data about molecules, in particular, the central biological polymers of DNA, RNA, and proteins. Although it can, bioinformatics does not normally involve other types of biopolymers such as complex carbohydrates or lipids.

- **Enhancing the understanding of living systems:** At its core, bioinformatics attempts to provide insight into biological function. For example, annotation of organismal genomes can lead to a better understanding of an entire range of biological characteristics, from cellular biochemistry and metabolism to ecosystem interactions and evolution. Importantly, bioinformatic analyses can also lead to a deeper knowledge of genetic diseases, as well as a more thorough comprehension of the agents of infectious diseases and the cellular responses of the host.

In addition to the definition, we developed an assessment rubric for testing student open-ended responses to a prompt about what bioinformatics is and how it is used. Scoring rubrics have frequently been used in formal K-16 (kindergarten through college) curricula as a tool for helping to inform and assess learning objectives and can be particularly useful in examining student understanding of "big picture" concepts or ideas [22]. Student awareness and understanding of bioinformatics as a field may well depend on whether they recognize bioinformatics as something they are experiencing as they move through instructional activities. The scoring rubric we developed targets the consistency of student-generated definitions (such as in a short response or essay) within an instruction-related definition of bioinformatics. It should allow instructors to "spot-check" students' overall awareness of what bioinformatics is based on the definition and to consider their "big picture" understanding of bioinformatics as they pursue their

Table 2: Assessment Rubric.

Criteria	0	2	2	3
Computer Science: Development and application of algorithms	Does not mention computers	Mentions use of computers	Mentions application of algorithms/tools	Mentions the development and application of algorithms/tools
Computer Science: Recognition of data analysis	Does not mention data	Mentions data	Mentions mass data	Mentions mass data analysis
Biology: Recognition of analysis at the molecular level	Does not mention molecules or molecular level	Mentions molecules	Mentions a biological polymer	Generalizes to multiple biological polymers
Biology: Recognition that bioinformatics enhances the understanding of living systems	Does not mention relationship to biological function	Mentions biological function	Mentions biological function with an example	Generalizes the integration of biological function to living systems
Bioinformatics: Recognition of the interdisciplinary nature of bioinformatics	Does not recognize that bioinformatics uses concepts of other disciplines	Recognizes that bioinformatics uses concepts from different disciplines	Recognizes that bioinformatics is a discipline	Recognizes that bioinformatics generates new interdisciplinary concepts

instructional coursework. Because student understanding will change throughout a course or program, the definition and rubric should give instructors a sense of how student perceptions of bioinformatics as a field change over time.

Methods

Our definition of bioinformatics was developed through an iterative and collaborative process. In particular, we began with a definition that the bioinformatician on the project team had used with his students. This definition was a synthesis of definitions from the literature (Table 1). As subject matter experts, the team agreed that this definition did not capture all of the fundamental aspects of bioinformatics discussed in the Introduction. In a lengthy and sometimes animated discussion, this initial definition was refined and expanded. At the same time, we began the development of the rubric to score student narrative responses when they were asked to describe bioinformatics and how it is used. This initial rubric was then used to score a group of student written responses to this prompt. Based on these results, the project team revised and refined the rubric along with the definition, which we found to be incomplete. This process was repeated once more before we arrived at the definition above and the rubric presented in Table 2.

Choosing an Assessment Format and Context

The process we used to develop our definition and rubric is modeled on work in the rapidly emerging field of technological pedagogical content knowledge (TPACK) [23, 24]. In P-16 (preschool through college) educational contexts, TPACK research, among other applications, strives to contribute to student knowledge-building, instructional materials, and assessment, while directly considering the technology, pedagogy, and targeted disciplinary content of instruction. Educational researchers in this field tend to suggest a learning-by-design approach in which educators, content experts, and technology specialists collaboratively design instruction and support engaging and immersive student activities [25, 26]. Rubrics have been successfully developed to examine TPACK within the context of classroom lesson plans, video observations, and teacher interviews [27]. Bioinformatics learning by students resides within the context of TPACK since bioinformatics depends on: 1) the effective use of technology; 2) inquiry-based or problem-based learning scenarios; and 3) focused content knowledge, as students engage in learning tasks associated with bioinformatics, such as comparing genetic data from two different sources to determine possible relationships.

While developing the definition and rubric, we spent considerable time discussing the big picture of bioinformatics instruction and systematically refining a useful definition of bioinformatics for life sciences instructors. During this process, we studied current research, asked students at different levels for their insights, and worked systematically to refine and validate our assessment rubric for the definition.

The assessment rubric itself was developed in three distinct phases: 1) team development of rubric criteria and definition refinement; 2) team reliability testing and refinement with student responses; and 3) external instructor reliability testing with student responses. These three development phases are described in detail below. At each phase of development, the team conducted focused discussion sessions. The entire development process took approximately eighteen months.

Phase 1 – Team Development of Rubric Criteria and Definition Refinement

In the first phase of our efforts, discussions were spread across several meetings in which the team discussed a tentative definition of bioinformatics for use in classroom instruction. The team also considered the utility of a scoring rubric that would be linked to the definition, and that would be used to score students' responses to a prompt about what bioinformatics is and how it is used. Rubrics successfully developed from other instructional venues, such as TPACK, were examined and discussed. Eventually, a pilot definition of bioinformatics emerged, as did a scoring rubric. The rubric included five rows (for criteria), one for each of the important aspects of bioinformatics discussed above, with four levels for scoring student understanding, ranging from low to high (0,1,2,3). The rubric included two criteria from computer science (development and application of algorithms; analysis of data), two from biology (recognition of analysis at the molecular level; recognition that bioinformatics enhances the understanding of living systems), and one for bioinformatics as a discipline (recognition of the interdisciplinary nature of bioinformatics). A five-by-four rubric was chosen due to advantages of relatively quick scoring and enough diversity in scoring levels for meaningful reliability testing of the rubric. Such a structure is consistent with well-established educational assessment research [27-29].

Phase 2 – Team Reliability Testing and Refinement with Student Responses

After the pilot definition and rubric had been drafted, the members of the team asked students in a wide range of courses at each of their institutions to respond to the simple prompt: "Please describe bioinformatics and how it is used." Students were also asked their gender, class standing (freshman, sophomore, etc.), the number of biology courses taken, the number of computer science courses taken, and their age, major, ethnicity, and race as well as how long they spent on the response and what suggestions they had for improving the task. (The response sheet is included as a Supplementary Document.) A total of fifty-seven responses across introductory and advanced coursework in the areas of biology, computer science, and bioinformatics were collected. From this, the team's evaluation/assessment specialist selected thirteen representative responses for the remaining team members to independently score in order determine the reliability of the rubric. After reliability testing, the team reviewed the overall set of student responses, as well as the demographics and student suggestions for refinement of the task. After reviewing the reliability statistics (described later) and the overall student responses, the team discussed possible refinements to the definition, rubric, and student response instrument and made small modifications to each.

Phase 3 – External Instructor Reliability Testing with Student Responses

After making small refinements to the definition, rubric, and student response request sheet, the team then solicited input from a new set of students to establish a new response set. These responses were solicited at approximately the same time—the beginning of the Fall 2014 semester. A total of eighty-six responses were evaluated from a demographically diverse group of students from different courses. From these, the project's evaluation/assessment expert selected a diverse subset of twenty-one responses to test the rubric further. Five colleagues of the team members, who had not been a part of the development process, were asked to score these responses. (The scoring sheet is included as a Supplementary Document.) The reliability of the scoring was again reviewed along with the full set of student responses and judge scores to determine whether the rubric scoring and definition needed further refinement. The Results section (see below) describes the general evolution of the instrument and instrument reliability indicators.

Rubric Validity and Reliability

As mentioned previously, the rubric consists of five rows. Each corresponds to one of the five fundamental aspects of bioinformatics described in the Introduction. All were considered during reliability testing. Note that the first two rows of the rubric deal with computer science aspects of bioinformatics and the third and fourth deal with the biological aspects of bioinformatics. The fifth row is the recognition of bioinformatics as a stand-alone discipline.

Both validity and reliability of the rubric were considered. Construct validity of an assessment relates to an instrument meaningfully measuring a construct of particular interest, which in our case was a contextual definition of bioinformatics for coursework instruction, as developed by the expert team. The construct validity efforts of the rubric assessment aligned with strategies that are recommended for rubric validity [30,31]. Construct validity was refined by the project team which consisted of four biologists (two of which are current or former department chairs), one mathematician/computer scientist (a former department chair), a bioinformatician, and an educational evaluation/assessment specialist with a strong background in STEM learning. All members of the project team have Ph.D.'s and have significant teaching and research experience. The team has been involved in several projects on bioinformatics instruction, including funded work from the National Science Foundation, and had been previously working on lecture materials, labs, and general curriculum development efforts. The team members sequentially reviewed the literature, developed the definition and rubric, and engaged various colleagues in related discussions associated with both the definition and rubric.

As described earlier, the reliability analysis of the rubric was conducted via two successive trials of independent student groups, across all three participating universities. Student samples at each institution were chosen purposefully by the evaluation/assessment specialist in order to get a diversity of student experience levels and course contexts, from beginning to more advanced bioinformatics experiences.

Using the data generated as described above, interrater reliability of the assessment rubric was determined for each location using two different calculations: 1) common percent agreement; and 2) intraclass correlation coefficient (ICC). These methods were selected in consultation with a statistician at the Center for Research on Youth, Families, and Schools at the University of Nebraska—Lincoln. The different calculations provide different insights into the reliability of the instrument. Percent of scorer agreement was used to determine the extent of interrater reliability associated with the pairing of scores from judges taken two at a time on each student response. The mean percent of agreement across judges was then computed. Adjacent scoring was used to represent judge agreement and was defined as two scores with no more than one rubric category difference. For example, rubric scores of 3 and 4 were considered to be in agreement, while scores of 2 and 4 were seen as out of agreement. Since percent of agreement has long been used for criterion-referenced scoring [27,29], its usefulness to check interrater reliability in a rubric context was clear. This technique also has the advantage of being able to generate a detailed spreadsheet of pairings, including the ability to easily review scoring patterns. ICC is also a well-known statistic that flexibly examines relationships among members of a class [32-34]. In this study, the instructors scoring the responses were essentially designated as a class, with rubric scores considered to be random effects and the instructors considered to be fixed effects for the ICC procedure. This procedure helps to investigate any scorer outliers and also allows for a more overall investigation of scorer "absolute agreement" tendencies.

IRB

This project was reviewed by the University of Nebraska Medical Center/University of Nebraska at Omaha Institutional Review Board (IRB) and given exempt status. (See IRB 372-11-EX.) As long as data was reported only in aggregate and with no possible links to individual students, informed consent was not required.

Results

As mentioned previously, the rubric reliability scoring effort was first conducted with five members of the project team on a sample of thirteen student responses chosen from a total of fifty-seven (blinded for scoring) from various courses at the three participating institutions (UNO, UNK, and NWU). Those on the project team have expertise in biology, computer science, and bioinformatics. The percent agreement of the scores from the five-person team of "expert judges" on the same set of student responses was computed for the rubric. Percent agreement procedures are known to be less sensitive to the "direction" of how judges' scores align (whether trends between judges are higher or lower than each other) and instead primarily considers how "close" judges' scores are to each other. The percent agreement for the first version of the rubric was computed to be 95.8%, representing strong agreement for members of the project team. An ICC score of 0.633 was also calculated for the first version of the rubric. This ICC level is generally adequate for reliability for a rubric of this short length, especially since this indicator is more sensitive to the judges' "absolute agreement."

Based on the review of the results from the blind scoring by members of the project team, and after some additional minor refinements to both the definition and rubric made by team consensus, the rubric was then examined further for interrater reliability using scores from five colleagues. All five were instructors at the three participating institutions and had not been a part of and were not familiar with the ongoing project discussions. These colleagues scored a new set of twenty-one selected student responses chosen from the eighty-six student responses described earlier. For these new scores from the five new judges on these responses, the interrater reliability of the assessment was again calculated using the percent agreement and ICC procedures. The percent agreement for this new set of scores was calculated to be 89.7% and had an ICC score of 0.620, representing adequate reliability for a short rubric-based assessment instrument.

Given the reliability testing results, the team concluded that the rubric had the reliability to be recommended for further use and refinement by other researchers. In particular, we believed it to be appropriate to share the instructional definition and related rubric with other researchers and instructors who might be interested in using it in their coursework or for potential educational research studies.

Discussion

For this project we developed an instructional definition for the field of bioinformatics along with a scoring rubric to assess the level of student understanding of this definition. Although two formal definitions of bioinformatics exist in the literature, both are dated, and the project team felt they did not adequately reflect the current state of the field especially as it is now applied in biology. As noted above, the instructional aspect of a definition is important as student learning contexts usually require a general definition. We have implemented bioinformatics in the undergraduate life science curricula at three Nebraska institutions: Nebraska Wesleyan University, University of Nebraska at Kearney, and University of Nebraska at Omaha. Specifically, we developed the rubric to assess multiple levels of understanding— from a basic definition to a more complex understanding of the interdisciplinary nature of bioinformatics—following the initial integration of bioinformatics into life science curricula.

During the first phase of rubric development, we created evaluation categories to assess: 1) minimal to no knowledge of bioinformatics, 2) general understanding of bioinformatics, and 3) a complete understanding of the applications and interdisciplinary nature of bioinformatics. These categories were defined to assess knowledge of undergraduate life science students across multiple stages of their scientific training, where we assumed students would gain a better understanding of bioinformatics after completing multiple inquiry-based learning exercises throughout their undergraduate career. Following assessment of our bioinformatics rubric, we found that STEM students and program majors do not have a good understanding of bioinformatics, and most have little knowledge of the discipline, even after completing inquiry-based bioinformatics exercises. Thus, we need to provide experiences to specifically point out the broad application and multidisciplinary nature of bioinformatics, to potentially allow students to form a mature understanding of bioinformatics as a discipline.

The rubric was found to be a reliable tool to evaluate student understanding of bioinformatics concepts incorporated throughout multiple levels of undergraduate life science curricula and can be effectively used with instructors of varying skill levels. To improve student learning, it is important that students have relevant experiences throughout the undergraduate curriculum. They also need to be repeatedly reminded of the "big picture" and to have the chance to apply bioinformatics concepts in classroom activities (e.g., laboratory exercises). We envision this instrument being useful in assessing how effective different learning strategies are applied across the curriculum. Ideally, students would be assessed multiple times throughout their undergraduate education to better access the impact of different learning experiences on their understanding (e.g., first exposure vs. multiple experiences, lecture-based vs. application-based experiences). It would also be helpful in understanding the overall impact of a revised curriculum if the trends in differences in understanding between beginning students and more experienced students could be more systematically examined. Finally, the existence of rigorous tools for assessing student learning may encourage the integration of bioinformatics into curricula. We hope that as other undergraduate institutions incorporate bioinformatics into their curricula, this rubric will be a useful tool to allow instructors to assess student learning about the emerging discipline of bioinformatics.

References

1. Magana AJ, Taleyarkhan M, Rivera Alvarado D, Kane M, Springer J, et al. A survey of scholarly literature describing the field of bioinformatics education and bioinformatics educational research. CBE Life Sci Educ. 2014; 13: 607-623.
2. Mason J. Bringing definitions into high definition. The Math Teach. 2010; 218: 10-12.
3. Van Dormolen J, Zaslavsky O. The many facets of a definition: the case of periodicity. Math Behav. 2003; 22: 91-106.
4. Stahl G, Koschmann T, Suthers D. Computer-supported collaborative learning: an historical perspective. Cambridge handbook of the learning sciences. 2006.
5. Cooper S, Cunningham S. Teaching computer science in context. ACM Inroads Magazine. 2010; 1: 5-8.
6. Guzdial M. Plain talk on computing education. Commun ACM. 2015; 58: 10-11.
7. Pemberton D. Definitional problems for environmental education and geographic education. J Environ Educ. 1989; 21: 5-14.
8. Annas G. Intelligent judging—Evolution in the classroom and the courtroom. N Engl J Med. 2006; 354: 2277-2281.
9. Wolff-Michael R, Van Eijck M. Fullness of life as minimal unit: Science, technology, engineering and mathematics (STEM) learning across the life span. Sci Ed. 2010; 94: 1027-1048.
10. Zollman A. Learning for STEM literacy: STEM literacy for learning. Sch Sci Math. 2012; 112: 12-19.
11. Huerta M, Haseltine F, Liu Y, Downing G, Seto, B. NIH Working Definition of Bioinformatics and Computational Biology, The Biomedical Information Science and Technology Initiative Consortium (BISTIC) Definition Committee of National Institutes of Health (NIH). 2000.
12. Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? A proposed definition and overview of the field. Methods Inf Med. 2001; 40: 346-358.
13. Welch L, Lewitter F, Schwartz R, Brooksbank C, Radivojac P, et al. Bioinformatics curriculum guidelines: toward a definition of core competencies. PLoS Comput Biol. 2014; 10: 1003496.
14. Welch L, Brooksbank C, Schwartz R, Morgan SL, Gaeta B, et al. Applying, evaluating and refining bioinformatics core competencies (an update from the curriculum task force of ISCB's education committee). PLoS Comput Biol. 2016; 12: 1004943.

15. Vincent AT, Charette SJ. Who qualifies to be a bioinformatician? Front Genet. 2015; 6: 164.
16. Smith DR. Broadening the definition of a bioinformatician. Front Genet. 2015; 6: 258.
17. Dinsdale E, Elgin SCR, Grandgenett N, Morgan W, Rosenwald A, et al. NIBLSE: A network for integrating bioinformatics into life sciences education. CBE Life Sci Educ. 2015; 14: 3.
18. Network for Integrating Bioinformatics into Life Sciences Education (NIBLSE) (2014). NIBLSE home page. http://niblse.unomaha.edu (retrieved 27 October 2016).
19. Bransford J, Brown A, Cocking R. How people learn: brain, mind, and experience and school. Washington, DC: National Academy Press. 2000.
20. Lajoie S. Extending the scaffolding metaphor. Instr Sci. 2005; 33: 541-557.
21. Lai M, Law N. Peer scaffolding of knowledge building through collaborative groups with differential learning experiences. J Educ Comput Res. 2006; 35: 123-144.
22. Arter J, McTighe J. Scoring rubrics in the classroom. Newbury Park, CA: Corwin Press. 2001.
23. Mishra P, Koehler MJ. Technological pedagogical content knowledge: A new framework for teacher knowledge. Teach Coll Rec. 2006; 108: 1017-1054.
24. Koehler MJ, Mishra P, Yahya K. Tracing the development of teacher knowledge in a design seminar: Integrating content, pedagogy, and technology. Comp Educ. 2007; 49: 740-762.
25. Niess M L. Preparing teachers to teach science and mathematics with technology: Developing a technology pedagogical content knowledge. Teaching and Teacher Education. 2005; 21: 509-523.
26. Gronlund NE. Measurement and evaluation in teaching (5th ed.), New York: McMillian. 1985.
27. Moskal BM, Leydens JA . Scoring rubric development: validity and reliability. PARE. 2000; 7: 71-81.
28. Litwin MS. How to assess and interpret survey psychometrics. The Survey Kit Series. Thousand Oaks, CA: Sage Publications. 2002; 8.
29. Allen MJ, Yen WM. Introduction to measurement theory. Long Grove, IL: Waveland Press. 2002.
30. Arter J, McTighe J. Scoring rubrics in the classroom. Thousand Oaks: Corwin Press, Inc. 2001.
31. Field AP. Intraclass correlation. In: Encyclopedia of Statistics in the Behavioral Sciences, ed. B Everitt and D Howell, Chichester. England: Wiley. 2005.
32. Griffin D, Gonzalez R. Correlational analysis of dyad-level data in the exchangeable case. Psychol Bull. 1995; 118: 430-439.
33. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. Psychol Methods. 1996; 1: 30-46.