Chapter 17
Assessment of Evidentiary Reasoning in Undergraduate Biology: A Lit Review and Application of the Conceptual Analysis of Disciplinary Evidence (CADE) Framework

Shiyao Liu, Chao Cai, Chaonan Liu, Ala Samarapungavan, Stephanie M. Gardner, Kari L. Clase, and Nancy J. Pelaez

17.1 Introduction

The knowledge of concepts and development of the competence to use that knowledge are foundational for students learning biology as a disciplinary practice. In accordance with the ACE-Bio Competencies framework (Pelaez et al., 2017; Chap. 1 in this volume), the AAAS (2011) *Vision and Change in Undergraduate Biology Education* report emphasizes that all post-secondary biology students need to develop core competencies applied to biology research practice. To understand how the design of scientific processes reveals what is known about living systems, competent students must demonstrate observational strategies, hypothesis testing, experimental design, evaluation of experimental evidence, and problem-solving strategies (AAAS, 2011). However, in this description of teaching and learning biology as an evidence-based discipline, the notion of evidence remains obscure. The

S. Liu · C. Liu · S. M. Gardner · N. J. Pelaez (✉)
Department of Biological Sciences, Purdue University, West Lafayette, IN, USA
e-mail: liu2313@purdue.edu; sgardne@purdue.edu; npelaez@purdue.edu

C. Cai
Libraries and School of Information Studies, Purdue University, West Lafayette, IN, USA
e-mail: caic@purdue.edu

A. Samarapungavan
Department of Educational Studies, Purdue University, West Lafayette, IN, USA
e-mail: alasam@purdue.edu

K. L. Clase
Department of Agricultural and Biological Engineering, Biotechnology Innovation and Regulatory Science (BIRS) Center, Purdue University, West Lafayette, IN, USA
e-mail: klclase@purdue.edu

351

monitoring of students' developing competence for reasoning with and about evidence in the context of biology disciplinary knowledge of relevance to an investigation is another challenge in teaching biology students to understand and do research. With the aim to facilitate appropriate choice of assessment tools and to identify gaps for the development of new assessments that reveal evidentiary reasoning difficulties in post-secondary biology laboratory classrooms, here we present a comprehensive literature review with existing assessments categorized using the Conceptual Analysis of Disciplinary Evidence (CADE) framework, which links biological knowledge with epistemic considerations, in addition to the Basic Competencies of Biological Experimentation (ACE-Bio) framework (Pelaez et al., 2017). The ACE-Bio Competencies and CADE frameworks partially overlap. Findings with the CADE show that some assessments fail to link disciplinary knowledge with epistemic reasoning processes while assessing students' evidentiary reasoning. To address the gaps revealed by the literature review and to extend our study of evidentiary reasoning beyond experimentation, two assessments were designed to identify difficulties that students have in reasoning about evidence in the context of research that involves evolutionary tree-thinking.

The performance expectations of the ACE-Bio Competencies, the CADE framework, or the *Vision and Change* (AAAS, 2011) report serve as a foundation for the development of assessments. However, such reform documents do not give enough detail to construct or use an assessment. As background for constructing new assessments, we first review literature on aspects to consider: (1) how to prompt or elicit a performance that reveals evidence of students' abilities; (2) what format is ideal for eliciting students' thoughts in a way that is feasible for the intended use of that information; (3) the need to situate the assessment tool or task in a relevant disciplinary context; and (4) what difficulties or competent performances are expected to be observed in the students (National Research Council, 2014).

17.1.1 Assessment Triangle

As a framework, we considered the assessment triangle as a process of reasoning from evidence about what students know and can do with their knowledge. The assessment triangle is defined as a "theory or set of beliefs about how students represent knowledge and develop competence in a subject domain" (National Research Council, 2001, p. 44). The triangle has three main points: *cognition* is the foundation which refers to a set of knowledge and abilities to use that knowledge that are important for a competent student; *observation* is the process of using an assessment instrument or a specific assignment to elicit a performance that reveals a student's cognition abilities; *interpretation* is the process of comparing that performance to a standard that would be expected for a competent student, which also involves

identifying students' difficulties that can then be addressed. To monitor a student's developing competence for reasoning from evidence, our study employed the assessment triangle according to the National Research Council (2001) and Pellegrino (2012).

Cognition serves as the foundation of the triangle, meaning that as a starting point, there is a need for a clearly defined set of knowledge and skills that are important for a competent student. A clear cognitive model like the ACE-Bio Competencies framework (Pelaez et al., 2017; Chap. 1 in this volume) provides understanding of how a student typically demonstrates domain expertise.

Observation stands on the second corner of the assessment triangle. Observation involves the process of eliciting a performance such as a writing assignment, research poster presentation, or a test item response designed to reveal a student's ability in the context of specific tasks. In particular, the assessment must have a precisely defined target for cognitive competence. For example, the assessment examples provided later in this chapter aim to gather information about how students use or apply the notion of evidence.

Interpretation is the last corner of the assessment triangle. Interpretation has been defined as the "methods and tools used to reason from fallible observations" (National Research Council, 2001, p. 48). Also, consider the audience who will engage in the interpretation. One important audience is the instructor who might modify instruction to help students address their difficulties according to what the instructor has observed in student responses to various assessment tasks. Instructors and administrators might use sample student responses to assessment tasks to track progress on anticipated learning outcomes, to identify the quality and range of student performances, and to determine if what was anticipated can be verified as a learning outcome resulting from a particular course or learning experience. This sort of summative assessment refers to the use of assessment data to evaluate students' knowledge upon completion of a learning sequence (Phelps, 2011).

Perhaps the most important interpretation is done by the student, who gets feedback and, as a result, may increase their own effort or they may abandon their goals or settle for lower personal expectations. Assessment used for individualized feedback to help students address their difficulties, also referred to as formative assessment, can also target motivating and helping students to develop their own improvement strategies. Student motivation involves changing their beliefs about themselves so that they can appropriately respond in ways that will advance their competence (Black & Wiliam, 1998). In this study we are interested in both formative and summative assessment. By formative assessment, we refer to the use of methods to encourage students to express what they are thinking so that they can adapt to the teaching flow and adopt strategies to achieve the anticipated learning outcomes (Black et al., 2003).

17.1.2   The CADE Framework

In this study our focus was on assessment tools, to identify gaps in existing assessments and to develop and implement new assessments that reveal evidentiary reasoning difficulties in post-secondary biology laboratory classrooms. But what are scientific evidence and evidentiary reasoning and why are these topics so important for students? First, various consensus reports have shifted instructional emphasis toward these notions in biology. According to the AAAS (2011) *Vision and Change* report mentioned earlier, undergraduate students should learn biology by applying the process of science, which involves getting data and evaluating it as experimental evidence. The ACE-Bio Competencies framework is not explicit about how data is used as evidence, but this can be inferred, for example, when the *Plan* competency item C. mentions Variables, which points out that a competent scientist will identify relevant, measurable variables for testing the hypothesis (Pelaez et al., 2017; Chap. 1 in this volume). According to Sandoval et al. (2004) and the *Next Generation Science Standards* (NGSS Lead States, 2013), scientific evidence is defined as data for addressing a question or supporting a claim. Thus, here we refer to evidentiary reasoning as reasoning with and about evidence throughout the entire research process. More specifically, this means applying evidence generated from a set of theoretical and methodological frameworks to assess the consistency or fit between potential theories and the reality (Giere, 2010). It is important for students to get a better understanding of evidence to help them make better decisions in their future when faced with issues like vaccines and climate change. Furthermore, an interest in biological evidence may even encourage some students to choose a biology career.

Instructors generally recognize that students struggle with understanding, using, and evaluating the evidence underpinning scientific knowledge, but the nature of those problems is not entirely clear. When Sandoval and Millwood (2005) examined the quality of secondary school students' use of evidence in written scientific explanations of natural selection, they found that students often failed to cite sufficient evidence. Despite a significant body of literature in science education focusing on issues like students' use of evidence, epistemic understandings about the nature of science, and development of scientific knowledge, educators consistently find that both K-12 and undergraduate students struggle with understanding the evidence to support their advanced science knowledge, as well as applying and evaluating this evidence by using scientific practices (Abd-El-Khalick et al., 2004; Duncan et al., 2018; Duschl, 2008; Furtak et al., 2010; Manz et al., 2020; McNeill & Berland, 2017; Tytler & Peterson, 2005). Since these problems may also relate to the complex nature of evidence (Samarapungavan, 2018), it is useful to consider what students should be doing when they are testing hypotheses and generating evidence to draw conclusions in terms of the processes that professional scientists engage in when they discover new knowledge. Scientific research practices involve decisions about what is worth investigating in addition to considering the limitations, uncertainties, and strength of any conclusions. Thus, by evidentiary reasoning, we include the evaluation of theories or models based on evidence, referred to as

evidence-based reasoning in recent studies built upon Toulmin's (1958) *The Uses of Argument*, according to Erduran et al. (2015) and Furtak et al. (2010) who have focused on students' reasoning about science phenomena and their use of evidence for backing their claims. In our approach to evidentiary reasoning, we also include scientists' theoretical and disciplinary knowledge applied to designing, executing, and analyzing investigations based on norms and procedures that are shared by members of their discipline, and considering the nature, scope, quality and sufficiency of the data, approaches, and theories according to what is already known of relevance to the research evidence.

The CADE framework aims to promote evidentiary reasoning by unpacking the notions of evidence described above into component parts (Samarapungavan, 2018). It is a wholistic framework that explicitly examines both the disciplinary knowledge as well as epistemological considerations of relevance to students' use of evidence at all stages of the research process by deconstructing evidence into four research practice component relationships: (1) *Theory->Evidence (T->E)* relationships are of relevance to formulating testable models; (2) *Evidence<=>Data (E<=>D)* relationships relate to the design, execution, and analysis of investigation findings; (3) *Evidence->Theory (E->T)* relationships refer to evaluation of evidence to draw and justify conclusions; (4) *Social Dimension* relationships refer to communicating with and about evidence to the public. By linking disciplinary and epistemic knowledge, the CADE draws attention to the knowledge and practices of the discipline as well as the scientific skepticism for justifying the nature, scope, and quality of the data, approaches, theories, and claims that underpin the evidence. With the CADE as a comprehensive and practical framework, it may be feasible to address students' difficulties in evidentiary reasoning among students who have struggled with understanding, using, and evaluating the evidence underpinning scientific knowledge.

In the first part of our study, assessments were examined using the CADE framework as a lens to monitor both domain-general and discipline-specific aspects of evidentiary reasoning that is a target of assessment. When the CADE framework components were mapped to current established assessments and rubrics, it served as the cognitive model to provide an explicit target for how students and experts represent the notion of evidence when they conduct evidentiary reasoning. In the second part of our study, the CADE framework was further used to identify students' difficulties with evidentiary reasoning in the context of biological science research practices.

17.1.3   Research Goals

The overarching goal of this study was to gain an understanding of established assessments that are being used to evaluate students evidentiary reasoning, to identify gaps, and then to address these gaps using the CADE to inform the design of new assessment items for revealing students' difficulties in evidentiary reasoning in the context of evolutionary tree-thinking as an example.

(1) What assessments are being used to reveal evidentiary reasoning difficulties among students in post-secondary biology laboratory classrooms where students conduct practical research? (2) What assessment gaps remain for new development of useful assessments? and (3) How did two CADE informed assessments of evolutionary tree-thinking used as undergraduate biology lab class test items reveal students' difficulties with evidentiary reasoning and address the gaps from the literature review?

17.2 Published Assessments Target Reasoning About Evidence

Our first study was a comprehensive literature review to identify a range of assessments used to monitor students' progress in understanding and using evidence as they learn to conduct biological research. Mapping of existing assessments to the CADE (Samarapungavan, 2018) and ACE-Bio Competencies (Pelaez et al., 2017) frameworks made it possible to identify gaps that remain in the assessments that are being used to reveal students' difficulties in evidentiary reasoning.

17.2.1 Literature Review

To find out what assessments are being used to reveal evidentiary reasoning difficulties among students in post-secondary biology laboratory classrooms where students conduct practical research and what assessment gaps remain for the new development of useful assessments, we first conducted a comprehensive literature review. We searched for assessments of students' evidence reasoning that have been used or adapted in the context of experimental/practical work in undergraduate biology laboratory classrooms. According to the National Research Council (2005), *America's Lab Report*, practical work includes experiences where learners interact with data about the natural world gathered by the learners themselves and with data about the natural world provided to them. Again, evidentiary reasoning in this chapter refers to the use of shared disciplinary norms to generate and evaluate evidence to reach scientific consensus (Giere, 2010; Manz et al., 2020; Samarapungavan, 2018). With the literature review, we were interested in both formative and summative assessments. We include formative assessments such as a coding rubric to understand students' classroom discussions. Summative assessments include pre- and post-test assessment items, proposals, and surveys measuring students' difficulties with evidentiary reasoning.

17.2.1.1   Search Procedure

We included peer-reviewed journal articles, proceedings, books, and dissertations to gain a comprehensive understanding of the assessments that are being used to reveal students' difficulties in evidentiary reasoning. We first identified 19 articles we thought must be included based on our experience with assessment of student learning about biology research. We expanded and refined the searching key words by reading through the 19 articles. Comprehensive literature searches were conducted in six databases by our second author, a librarian and a biological sciences specialist. These databases include education research related databases: ERIC, Education Sources, Education Full Text, and APA PsycINFO, which were searched in the EBSCO interface; a general database: Web of Science Core Collection, which was searched in the Web of Science interface; a dissertation database: ProQuest Dissertations and Theses in the ProQuest interface. The search was performed in EBSCO using the search string: (assess* OR evaluat* OR measur* OR test* OR effective* OR rubric*) AND (reasoning OR "critical thinking" OR "scientific writing*" OR "scientific literac*" OR "research concept*" OR "biolog* concept*" OR "experimental design*" OR "hypothesis testing" OR "test* hypothesis" OR "variability" OR "variation") AND (lab* OR experiment* OR "practical work*" OR "investigation*" OR "research experience*" OR "scientific practi*") AND (bio*) AND (undergrad* OR post-secondary). The same search string was adapted to fit the syntax for searches in the Web of Science and ProQuest. Additional articles were also obtained using hand searching in Google Scholar. The search was performed on September 20, 2021, and was limited to articles published after January 1, 2001. We selected 2001 as the beginning date range in order to include a decade before the report on *Vision and Change in Undergraduate Biology Education* (AAAS, 2011), which emphasizes the essential role of thinking with and about evidence in undergraduate biology education. Our method was designed to capture a comprehensive picture of assessments focusing on evidentiary reasoning.

Based on the search criteria described above, the number of results retrieved in the initial online searches was 719 articles. Among these 719 articles, only 10 out of 19 articles we thought must be included were found with the search strategy. This indicates the difficulty in conducting an educational literature review on this topic, as in disciplinary biology education, people tend to use different terms to describe one concept. Therefore, although we may not have found all publications of relevance to our study, the number of articles in our sample was sufficient plus we decided to include the 9 articles that we had already identified to make the literature review more comprehensive.

17.2.1.2   Screening the Search List

The second author uploaded the 719 articles in the search lists plus the 9 additional articles we had identified using Rayyan (https://www.rayyan.ai/), a collaborative platform. The first and last authors carried out a preliminary review of the articles in

the list together. Criteria used for including an article in our review were focused on the purpose of this literature review. To be more specific, as we defined inclusion criteria for the screening, we decided that the following categories were excluded: (1) Articles that do not contain an assessment, such as articles about curriculum designed for improving students evidentiary reasoning as a learning outcome without measuring the effectiveness of the curriculum design or with measuring the effectiveness by using student self-reports, which excluded 331 articles (e.g., Fry & Burr, 2011); (2) Articles that do not focus on measuring students' evidentiary reasoning, including articles that are trying to assess students' understanding of the nature of science (NOS) (e.g., Bautista et al., 2014), content knowledge understanding/retention (e.g., Gauthier et al., 2019), moral reasoning (e.g., Stransky et al., 2021), and self-efficacy (e.g., Beck & Blumer, 2021), which excluded 93 articles; (3) Articles not targeting the undergraduate level, which excluded 14 articles; (4) Articles not in the context of education, such as experiments about phycology, which excluded 44 articles; (5) Articles not in the context of biology or that would not ever be taught in a biological sciences department or in biology classrooms, such as studies about clinical reasoning for diseases diagnosis, analytical chemistry, and evidence reasoning in a domain-general context (e.g., Bhavana, 2009), which excluded 155 articles; (6) Articles that do not target students, such as studies of GTAs or instructor groups (e.g., Gardner & Jones, 2011), which excluded 3 articles; (7) Articles not in English, which excluded 3 articles; (8) Articles that could result in the same assessment being included twice because the authors used established assessments or adapted established assessments without much change (e.g., Auerbach & Schussler, 2017), which excluded 20 articles; (9) Articles with an assessment without a rubric or scoring structure with the assessment (e.g., Bugarcic et al., 2012), which excluded 26 articles; (10) Scientific literacy reading skills without evidentiary reasoning (e.g., Krontiris-Litowitz, 2013), which excluded 1 article; (11) and 93 duplicate articles. Since an article may be excluded by multiple exclusion criteria, a total of 46 articles were included.

17.2.2   Coding

The 46 included articles were coded both into the four relationships of the CADE and the seven scientific practice competencies of ACE-Bio theoretical frameworks (Pelaez et al., 2017; Chap. 1 in this volume). First, the seven ACE-Bio competencies were mapped into the four relationships of the CADE framework (See columns 1 and 2 in Tables 17.1, 17.2, 17.3, and 17.4). Then, the coding scheme was further divided into both disciplinary knowledge and epistemic considerations. The subcategories of each relationship were divided into different scientific practice competencies, as shown in column two of Tables 17.1, 17.2, 17.3, and 17.4. Using the *Plan* competence as an example, we divided and mapped it from the ACE-Bio Competencies framework into two different relationships within the CADE framework, which are the *Theory->Evidence* and the *Evidence<=>Data* relationships.

Table 17.1 Theory to evidence relationships (Codes T -> E)

CADE Evidentiary Practices	ACE-Bio Competencies/CADE questions	Rubric/Worksheets for Presentations/ Proposals/Reports	Assessment that can be used on an Exam
T -> E **Knowledge** Identify, Question, Plan	*Identify*: the ability to identify gaps or limitations in current research knowledge through the review, filtering and synthesis of relevant literature. **T->E** What are the key domain phenomena? What are the important unsolved problems?	Blair, 2014[a, b] Fisher, 2018 (IWCR) Full, 2015 King, 2018 Klein, 2014 (Written Communication Rubric)[b] Kowalski, 2016 Lansverk, 2020[b] Martinez-Vaz, 2020[b] Resendes, 2015[b] Seixas Mello, 2021 (Scheme representing the epistemic levels) Simmons, 2014 Sorte, 2020 Spence, 2020 (Proposal, Poster) St. Onge, 2007 Timmerman, 2011 Ward, 2014 Weaver, 2016 (adapted from AAC&U VALUE rubrics) Younkin & Romano, 2018[b]	Deane, 2014 (BEDCI)[b] Killpack, 2018[b]
T -> E **Epistemology** Identify, Question, Plan	*Identify*: the ability to identify gaps or limitations in current research knowledge through the review, filtering and synthesis of relevant literature. **T->E** Are alternative models or theories considered?	Fisher, 2018 (IWCR) Full, 2015 King, 2018 Kowalski, 2016 Seixas Mello, 2021 (Scheme representing the epistemic levels) Simmons, 2014 Sorte, 2020 Spence, 2020 (Poster) St. Onge, 2007 Timmerman, 2011 Ward, 2014 Weaver, 2016 (adapted from AAC&U VALUE rubrics)	

Table 17.1 (continued)

CADE Evidentiary Practices	ACE-Bio Competencies/CADE questions	Rubric/Worksheets for Presentations/ Proposals/Reports	Assessment that can be used on an Exam
T -> E **Knowledge** Identify, Question, Plan	*Question*: the ability to generate research questions and formulate hypotheses. **T->E** What are the possible mechanisms, causal relationships, and processes?	Blair, 2014[b] Boomer, 2021 Full, 2015 King, 2018 Kowalski, 2016 Lansverk, 2020 Martinez-Vaz, 2020 Ott & Carson, 2014[b] Resendes, 2015[b] Simmons, 2014 Sorte, 2020 Spence, 2020 (Proposal, Poster) St. Onge, 2007 Timmerman, 2011 Ward, 2014 Younkin, 2018[b]	Anderson, 2011 (IPSA) Dasgupta, 2016 (neuron)[b] Dasgupta, 2014 (RED Shrimp, Drug, Bird) Deane, 2014 (BEDCI) Kiillpack, 2018 (TIED) Kowalski, 2016 (BIOCHEM, NEURO SRQ) Robertson, 2008[b]
T -> E **Epistemology** Identify, Question, Plan	*Question*: the ability to generate research questions and formulate hypotheses. **T->E** Have relationships between variables been clearly specified?	Boomer, 2021 Full, 2015 King, 2018 Kowalski, 2016 Lansverk, 2020 Martinez-Vaz, 2020 Simmons et al., 2014 Sorte, 2020 Spence, 2020 (Poster) St. Onge, 2007 Timmerman et al., 2011 Ward, 2014	Dasgupta, 2014 (RED Shrimp, Drug, Bird) Deane, 2014 (BEDCI) Kiillpack, 2018 (TIED) Kowalski, 2016 (BIOCHEM, NEURO SRQ)

Table 17.1 (continued)

CADE Evidentiary Practices	ACE-Bio Competencies/CADE questions	Rubric/Worksheets for Presentations/ Proposals/Reports	Assessment that can be used on an Exam
T -> E **Knowledge** Identify, Question, Plan	*Plan*: the ability to plan feasible and ethical experiments to answer research questions or test hypotheses. **T->E** What variables are relevant? Why did you decide to look at those variables?	Blair, 2014 Boomer, 2021[b] Full, 2015 Kowalski, 2016 Lansverk, 2020 Martinez-Vaz, 2020[b] Resendes, 2015[b] Reyynders, 2020 Simmons, 2014 Sorte, 2020 Spence, 2020 (Proposal, Poster) St. Onge, 2007 Timmerman, 2011 Ward, 2014 Weaver, 2016 (adapted from AAC&U VALUE rubrics) Younkin & Romano, 2018	Brownell, 2014[b] Dasgupta, 2016 (neuron) Dasgupta, 2014 (RED Shrimp, Drug, Bird) Deane, 2014 (BEDCI) Deane, 2016 (SRBCI) Irby, 2018 Kiillpack, 2018 (TIED) Kowalski, 2016 (CHEMBIO, BIOCHEM, NEURO SRQ) Robertson, 2008[b] Rybarczyk, 2014 (MBDAT) Shi, 2011 (Experimental Control Exercises)
T -> E **Epistemology** Identify, Question, Plan	*Plan*: the ability to plan feasible and ethical experiments to answer research questions or test hypotheses. **T->E** Is relevant evidence used to render the question, hypotheses, plausible? Is an articulated model complete, specific, and internally consistent?	Boomer, 2021 Full, 2015 Kowalski, 2016 Lansverk, 2020 Reynders, 2020 Simmons, 2014 Sorte, 2020 Spence, 2020 (Poster) St. Onge, 2007 Timmerman, 2011 Ward, 2014 Weaver, 2016 (adapted from AAC&U VALUE rubrics) Younkin, 2018	Dasgupta, 2016 (neuron) Dasgupta, 2014 (RED Shrimp, Drug, Bird) Deane, 2014 (BEDCI) Deane, 2016 (SRBCI) Irby, 2018 Kiillpack, 2018 (TIED) Kowalski, 2016 (CHEMBIO, BIOCHEM, NEURO SRQ) Rybarczyk, 2014 Shi, 2011 (Experimental Control Exercises) Sirum, 2011 (EDAT)[b]

[a]To save space only the first author was used in the table
[b]This assessment or rubric fails to link disciplinary knowledge with epistemic considerations in examining students' evidentiary reasoning ability

Table 17.2 Evidence ⇔ Data Relationships (Codes E ⇔ D)

CADE Evidentiary Practices	ACE-Bio Competencies/ CADE questions	Rubric/Worksheets for Presentations/ Proposals/Reports	Assessment that can be used on an Exam
E ⇔ D **Knowledge** Analyze, Conduct, Plan	_Analyze_: the ability to apply analytical reasoning to data processing. **E ⇔ D** What data models are used to organize/analyze data (e.g., graphs, statistical models)? What are known sources of error and how will they be accounted for?	Angra, 2017 (graph rubric)[b] Boomer, 2021[a, b] Brunnauer, 2016 (Summary Report)[b] Fisher, 2018 (IWCR) Full, 2015 Klein, 2014 (Written Communication Rubric)[b] Kowalski, 2016 Lansverk, 2020 Martinez-Vaz, 2020 Ott, 2014 Reynders, 2020 Seixas Mello, 2021 (Scheme representing the epistemic levels) Simmons, 2014 Sorte, 2020 Spence, 2020 (Proposal, Poster) St. Onge, 2007 Timmerman, 2011 Volz, 2009 Ward, 2014 Weaver, 2016	Anderson, 2011 (IPSA)[b] Dasgupta, 2016 (neuron)[b] Dasgupta, 2014 (RED Shrimp, Drug, Bird) Deane, 2014 (BEDCI) Deane, 2016 (SRBCI) Fisher, 2018 (QLR) Gormally, 2012 (TOSLS) Hester, 2014 (adapt IMCA)[b] Irby, 2018[b] Rybarczyk, 2014 (MBDAT)[b]

Table 17.2 (continued)

CADE Evidentiary Practices	ACE-Bio Competencies/ CADE questions	Rubric/Worksheets for Presentations/ Proposals/Reports	Assessment that can be used on an Exam
E ⇔ D **Epistemology** Analyze, Conduct, Plan	*Analyze*: the ability to apply analytical reasoning to data processing. **E ⇔ D** Are the models used appropriate? Have potential sources of error and confounding factors been evaluated?	Angra, 2017 (graph rubric)[b] Fisher, 2018 (IWCR) Full, 2015 Kowalski, 2016 Lansverk, 2020 Martinez-Vaz, 2020 Ott, 2014 Reynders, 2020 Seixas Mello, 2021 (Scheme representing the epistemic levels) Simmons, 2014 Sorte, 2020 Spence, 2020 (Proposal) St. Onge, 2007 Timmerman, 2011 Ward, 2014 Weaver, 2016 Younkin, 2018	Brunnauer, 2016 [b] Dasgupta, 2014 (RED Shrimp, Drug, Bird) Deane, 2014 (BEDCI) Deane, 2016 (SRBCI) Fisher, 2018 (QLR) Gormally, 2012 (TOSLS) Hicks, 2020 (BioVEDA)[b]
E ⇔ D **Knowledge** Analyze, Conduct, Plan	*Conduct*: the ability to conduct an investigation to achieve research goals. **E ⇔ D** Have relevant investigations been conducted? Are diverse relevant data types collected? Has an investigation been replicating with enough trials?	Brunnauer, 2016 (Summary Report)[b] Full, 2015 St. Onge, 2007 Ott, 2014[b] Simmons, 2014 Sorte, 2020[b] Spence, 2020[b] Volz, 2009[b] Ward, 2014[b] Younkin, 2018[b]	Brownell, 2014 Irby, 2018[b] Robertson, 2008[b]
E ⇔ D **Epistemology** Analyze, Conduct, Plan	*Conduct*: the ability to conduct an investigation to achieve research goals. **E ⇔ D** Are different types of data collected from diverse measures to provide support? Are sufficient trials conducted to identify data variability?	Full, 2015 Simmons, 2014 St. Onge, 2007	Brownell, 2014 Dasgupta, 2014 (RED)

Table 17.2 (continued)

CADE Evidentiary Practices	ACE-Bio Competencies/ CADE questions	Rubric/Worksheets for Presentations/ Proposals/Reports	Assessment that can be used on an Exam
E ⇔ D **Knowledge** Analyze, Conduct, Plan	_Plan_: the ability to plan feasible and ethical experiments to answer research questions or test hypotheses. **E ⇔ D** Deciding what to observe or measure: ·How are variables defined? ·Continuous or categorical, ·Independent, dependent, controlled etc. ·Intervals, range sampled What instruments, techniques, apparatus will be used to collect/record data and why are these appropriate? What sampling procedures are used?	Angra, 2017 (graph rubric)[b] Blair, 2014[b] Fisher, 2018 (IWCR) Full, 2015[b] Klein, 2014 (Written Communication Rubric)[b] Kowalski, 2016 Lansverk, 2020 Martinez-Vaz, 2020 Ott, 2014 Simmons, 2014 Sorte, 2020 Spence, 2020[b] St. Onge, 2007 Timmerman, 2011 Volz, 2009[b] Weaver, 2016	Brownell, 2014 Dasgupta, 2014 (RED Shrimp, Drug, Bird) Deane, 2014 (BEDCI) Deane, 2016 (SRBCI) Gormally, 2012 (TOSLS) Hicks, 2020 (BioVEDA) Kiillpack, 2018 (TIED) Kowalski, 2016 (CHEMBIO, BIOCHEM, NEURO SRQ) Rybarczyk, 2014 (MBDAT)[b] Shi, 2011
E ⇔ D **Epistemology** Analyze, Conduct, Plan	_Plan_: the ability to plan feasible and ethical experiments to answer research questions or test hypotheses. **E ⇔ D** Are variables clearly defined? Are variables defined in a way that is consistent with what is known (similarity versus differences)? Is technical precision, power, sensitivity, reliability, of data collection procedures adequate?	Angra, 2017 (graph rubric)[b] Fisher, 2018 (IWCR) Full, 2015 Kowalski, 2016 Lansverk, 2020 Martinez-Vaz, 2020 Simmons, 2014 Sorte, 2020 St. Onge, 2007 Timmerman, 2011 Weaver et al., 2016 (adapted from AAC&U VALUE rubrics)	Brownell, 2014 Dasgupta, 2014 (RED)[b] Dasgupta, 2014 (RED Shrimp, Drug, Bird) Deane, 2014 (BEDCI) Deane, 2016 (SRBCI) Gormally, 2012 (TOSLS) Hicks, 2020 (BioVEDA) Kiillpack, 2018 (TIED) Kowalski, 2016 (CHEMBIO, BIOCHEM, NEURO SRQ) Shi, 2011 Sirum, 2011 (EDAT)[b]

[a]To save space only the first author was used in the table
[b]This assessment or rubric fails to link disciplinary knowledge with epistemic considerations in examining students' evidentiary reasoning ability

Table 17.3 Evidence -> Theory Relationships (Codes E -> T)

CADE Evidentiary Practices	ACE-Bio Competencies/CADE questions	Rubric/Worksheets for Presentations/ Proposals/Reports	Assessment that can be used on an Exam
Evidence -> Theory **Knowledge** Conclude	*Conclude*: the ability to draw conclusions about data that are limited to the scope inherent in the experimental design. **E->T** What data reporting standards apply? (e.g., attrition, error rates, outliers)? What has been learned from the evidence? Are findings/conclusions explained in terms of what is already known in biology? What other conclusions are compatible with the evidence?	Angra, 2017 (graph rubric)[b] Blair, 2014[b] Fisher, 2018 (IWCR) Full, 2015[a] Klein, 2014 (Written Communication Rubric) King, 2018 Kowalski, 2016 Lansverk, 2020 Martinez-Vaz, 2020 Ott, 2014[b] Reynders, 2020 Seixas Mello, 2021 (Scheme representing the epistemic levels) Simmons, 2014 Sorte, 2020 Spence, 2020 St. Onge, 2007 Timmerman, 2011 Volz, 2009 Ward, 2014 Weaver, 2016 Younkin, 2018[b]	Anderson, 2011 (IPSA) Brownell, 2014 Coleman, 2015 (correlation) Dasgupta, 2016 (neuron) Dasgupta, 2014 (RED Shrimp, Drug, Bird) Deane, 2016 (SRBCI) Gormally, 2012 (TOSLS) King, 2018 Rybarczyk, 2014 (MBDAT) Schen, 2007 Shi, 2011[b] Terry, 2007 (CEAT)

Table 17.3 (continued)

CADE Evidentiary Practices	ACE-Bio Competencies/CADE questions	Rubric/Worksheets for Presentations/ Proposals/Reports	Assessment that can be used on an Exam
Evidence -> Theory **Epistemology** Conclude	*Conclude*: the ability to draw conclusions about data that are limited to the scope inherent in the experimental design. E->T Are data reports fair/complete? Are the conclusions internally consistent? Are the conclusions aligned with what is known? How does the evidence distinguish between multiple interpretations or hypotheses? Have alternative conclusions been explored and rebutted? Are limitations and uncertainties explicitly acknowledged/ addressed?	Fisher, 2018 (IWCR) Full, 2015 Klein, 2014 (Written Communication Rubric) King, 2018 Kowalski, 2016 Lansverk, 2020 Martinez-Vaz, 2020 Reyynders, 2020 Seixas Mello, 2021 (Scheme representing the epistemic levels) Simmons, 2014 Sorte, 2020 Spence, 2020 (Proposal, Poster) St. Onge, 2007 Timmerman, 2011 Volz, 2009 Ward et al., 2014 Weaver, 2016	Brownell, 2014 Coleman, 2015 (correlation) Dasgupta, 2016 (neuron) Dasgupta, 2014 (RED) Dasgupta, 2014 (RED Shrimp, Drug, Bird) Deane, 2016 (SRBCI) Gormally, 2012 (TOSLS) King, 2018 Rybarczyk, 2014 Schen, 2007 Sirum, 2011 (EDAT)[b] Terry, 2007(CEAT)

[a]To save space only the first author was used in the table
[b]This assessment or rubric fails to link disciplinary knowledge with epistemic considerations in examining students' evidentiary reasoning ability

The *Plan* competence within the *Theory->Evidence* relationship considers the variables from the theory perspective, which includes the biological disciplinary knowledge of choosing relevant variables based on some established biological theories and epistemic justification of the variables and the model used to organize them. In contrast, the *Plan* competence within the *Evidence<=>Data* relationship considers the variables from the data perspective, which includes the biological disciplinary knowledge of how to define and measure the variables, what sampling procedures are used, and epistemic justification of the definition and techniques chosen.

17.2.2.1 Data Analysis Method

By reading the full text of the included articles in detail, we identified the assessments and rubrics linked to the assessments that aim to reveal students' evidentiary reasoning competence and difficulties. The mapped CADE and ACE-Bio theoretical frameworks were used as the coding scheme for data analysis. Each subcategory related to the notion of evidence that is measured by the assessment was coded as a unit. The first author inductively coded all the assessments for the first pass, and

Table 17.4 Social dimensions (Codes E -> T)

CADE Evidentiary Practices	ACE-Bio Competencies/CADE questions	Rubric/Worksheets for Presentations/ Proposals/Reports	Assessment that can be used on an Exam
Social Dimensions **Knowledge** Communicate	*Communicate*: the ability to communicate research work in professionally appropriate modes, including visual, written, and oral formats. **Social dimensions** Are the credentials/expertise described? Has work been peer-reviewed? Is research infrastructure adequate?	Angra, 2017 (graph rubric)[b] Blair, 2014[a, b] Boomer, 2021[b] Fisher, 2018 (IWCR) Full, 2015 Klein, 2014 (Written Communication Rubric)[b] King, 2018 Lansverk, 2020 Martinez-Vaz, 2020[b] Simmons, 2014 Sorte, 2020 Spence, 2020 (Poster)[b] St. Onge, 2007 Timmerman, 2011[b] Ward, 2014 Weaver, 2016	Brrunauer, 2016 (Graphical analyses) Fisher, 2018 (QLR)
Social Dimensions **Epistemology** Communicate	*Communicate*: the ability to communicate research work in professionally appropriate modes, including visual, written, and oral formats. **Social dimensions** Are researchers credible? Has there been expert critique? Was it feasible to do research well?	Angra, 2017 (graph rubric)[b] Fisher, 2018 (IWCR) Full, 2015 King, 2018 Lansverk, 2020 Simmons, 2014 Sorte, 2020 St. Onge, 2007 Ward et al., 2014 Weaver, 2016	Brrunauer, 2016 (Graphical analyses) Fisher, 2018 (QLR)

[a]To save space only the first author was used in the table
[b]This assessment or rubric fails to link disciplinary knowledge with epistemic considerations in examining students' evidentiary reasoning ability

"peer debriefing to enhance the accuracy of the account" strategy suggested by Creswell and Poth (2018) was used to enhance the accuracy of findings. The last author, an experienced researcher on assessment, consistently challenged the coding and played the role of peer debriefer. All disagreements raised during the second pass were discussed until reaching consensus. As both the CADE and the ACE-Bio frameworks unpack the complex notion of evidence and the meaning of scientific practice, instead of comparing interrater reliability, we chose to use peer consensus coding to discover complexities in the data (Richards & Hemphill, 2018).

17.2.2.2   Data Analysis Examples

As an example, in the neuron assessment where biological disciplinary knowledge related to mitochondria movement in neurons is provided as a scenario, Dasgupta et al. (2016) assess students' reasoning about visualization of experiments by letting students predict their expected key findings in diagrams and explain what improvement they could make in the data to become more certain of their diagrams. In comparing the components of this assessment as well as actual expert and student responses from the publication to the table of CADE categories and criteria at https://tinyurl.com/CADE2022, we established, for example, that the neuron assessment measures students' evidentiary reasoning regarding the *Evidence<=>Data* relationship with emphasis on the *Analyze* competence, which includes knowledge about understanding the data models that were used to organize the data and epistemic considerations about the model's appropriateness and limitations. Disciplinary knowledge of experimentation research design is called for even if this assessment provides the relevant cell biology disciplinary knowledge in the form of a narrative scenario in the assessment with three diagrams to illustrate the mechanisms for moving mitochondria that can be modified in cells exposed to various drugs. Another example is from the Scheme Representing the Epistemic Levels framework of Seixas Mello et al. (2021) who measure students' arguments based on the quality of the justifications for conclusion validity in the context of the complement system in seven epistemic levels. In comparing components of the CADE categories and criteria at https://tinyurl.com/CADE2022 to the seventh epistemic level "statements incompatible with scientific knowledge" of Seixas Mello et al. (2021), we established that the authors measured students' conclusion competency under *Evidence->Theory* relationships regarding their use of established knowledge and theories linked to their justification of external consistency, which is an epistemic consideration.

17.2.3   Findings from a Review of Published Assessments

For our first research goal about how established assessments are being used to evaluate students' evidentiary reasoning, we discuss here the findings in terms of how the included assessments/rubrics evaluate students' difficulties in evidentiary reasoning. We then identify gaps that remain to be addressed according to the current established assessments/rubrics.

17.2.3.1 What Assessments Are Being Used to Reveal Evidentiary Reasoning Difficulties Among Students?

Tables 17.1, 17.2, 17.3, and 17.4 show which competencies of scientific practice and which relationships of the notion of evidence are being measured by the established assessments/rubrics that have been used in tracking the progress of post-secondary biology students in laboratory classrooms where students conduct practical research on a variety topics.

17.2.3.2 What Assessment Gaps Remain for Development of New and Useful Assessments?

As the coding results shown, first there are both assessments and rubrics that fail to link the disciplinary knowledge with epistemic considerations while assessing students' evidentiary reasoning ability (indicated by the superscript[b] in the tables). Some of these assessments/rubrics pay close attention to the important role of knowledge in evidentiary reasoning but fail to provide students with opportunities to justify the validity of their claims and less attention is directed to examining students' epistemic considerations. For example, Killpack & Fulmer (2018) assess students' experimental design skills, where students have to conduct evidentiary reasoning by designing an experiment to explore the factors that cause the diversity of feeding behavior in guppies. By using questions like "what are the control group(s)?" and "what data will you collect, and "how will you collect it?", the assessment evaluates students' biological disciplinary knowledge related to the experimental design, while ignoring the importance of assessing students' epistemic considerations by having students justify their decisions. Others measure students' evidentiary reasoning in a general context without linking epistemic considerations with specific biological disciplinary knowledge (see, for example, the EDAT by Sirum & Humburg, 2011).

Secondly, few assessments examine students' competence to *Conduct* an experiment within the *Evidence<=>Data* relationship. While only two assessments measure students evidentiary reasoning regarding reasoning about variation with replication (Brownell et al., 2014; Dasgupta et al., 2014), there is no assessment to evaluate students' evidentiary reasoning about the necessity of using diverse evidence in drawing conclusions or the use of convergent evidence for conclusions.

Finally, only two of the assessments measure students' competence to *Identify* a research problem to address in the *Theory->Evidence* relationship, where students need to reason through their decisions about the evidence to be examined with disciplinary knowledge of relevance to the investigation, and whether alternative models or theories are considered.

17.3 Assessment Gaps Addressed with CADE-Informed Test Questions

Two assessments informed by the CADE framework were developed to specifically target assessment gaps identified in the literature review but in the context of a lab activity on evolutionary tree-thinking, thus expanding our focus from experimentation in biology to include another research approach. The assessments were implemented as part of a biology lab classroom test guided by the assessment triangle to address several assessment gaps in order to track undergraduate students' evidentiary reasoning progress in biology: the linking of disciplinary knowledge with epistemic reasoning, use of disciplinary knowledge to inform a hypothesis or research goal, considering alternative models to test, and evaluating claims in terms of convergent evidence that could support or raise questions about the strength of an inference.

17.3.1 Design of the Assessments

The assessments presented below were designed to reveal post-secondary biology students' difficulties in evidentiary reasoning. To do so, assessment design was informed by both the CADE (Samarapungavan, 2018) and the assessment triangle frameworks (National Research Council, 2001; Pellegrino, 2012). Each assessment prompted a response that would link disciplinary knowledge with epistemic considerations and to target evidentiary reasoning according to *Theory->Evidence*, *Evidence<=>Data*, and *Evidence->Theory* science research practice relationships by using three open-ended questions.

Based on the assessment triangle framework (National Research Council, 2001; Pellegrino, 2012), a cognitive model with a rich psychological perspective provides detailed information to inform the assessment design. Thus, we linked each epistemic consideration that has been identified in the CADE framework we would like to assess with the correlated specific biology disciplinary knowledge in the context of evolutionary tree-thinking identified by the NGSS (NGSS Lead States, 2013) and an authoritative undergraduate evolution website (Thanukos et al., 2010) to establish the cognitive foundation of our assessments. To be more specific, in the *Theory->Evidence* relationship, students must consider if relationships between variables been clearly specified. To do this, students need to link their evolutionary tree-thinking disciplinary knowledge such as using convergent evidence from diverse sources to infer the relatedness of taxa, which includes the similarity and differences of unique DNA nucleotide sequences, anatomical evidence, variable features of fossils such as comparing the shape or number of bones, physical, chemical, and geological evidence to establish the age of fossils, etc. to reason about this

knowledge in concert with epistemic considerations for justifications. For interpretation, this cognitive foundation also served as the rubric, "the methods and tools used to reason from fallible observations" (National Research Council, 2001, p. 48). In order to observe students' evidentiary reasoning competence and difficulties, the assessments provide students with rich conflict through open-ended scenarios where two scientists have different claims regarding the closest living relatives of whale/echidna according to their different evidence. These open-ended scenarios aim to invite students to reason with and about evidence without worrying about the correct answers, since there is no one correct answer. In order to interpret students' evidentiary reasoning, we inductively coded each student's answer into the rubric we established (roughly based on the CADE table at https://tinyurl.com/CADE2022 or contact the first author for the rubrics). If the specific disciplinary knowledge correlated to the epistemic consideration is hard to define by referring to standard reports such as *Vison and Change* (AAAS, 2011), the cognitive foundation can also be established using expert answers.

17.3.2 Participants

The assessments were implemented in an introductory biology lab course at a large midwestern university with high research activity. Expert answers were from a graduate student in the Ecology and Evolutionary Biology program and a professor who teaches an upper division Evolutionary Biology course for teachers. All assessment responses were collected according to a protocol that was reviewed and approved by the Institutional Review Board (IRB#17020187760251). The graduate student had served as a graduate teaching assistant for the target course during three semesters without any CADE or ACE-Bio Competencies training. Student responses were collected from pre- and post-tests at the beginning and end of the target lab course.

17.3.3 Addressing Assessment Gaps to Reveal Students' Difficulties with Evidentiary Reasoning About Evolutionary Trees

To address the gaps, two assessments (Boxes 17.1 and 17.2) were designed to reveal students' difficulties in evidentiary reasoning by using scenarios where two scientists from different biological disciplines using different sets of convergent evidence draw different claims. With the aim to evaluate students' evidentiary reasoning in a comprehensive matter, there are no right or wrong answers to the questions. The questions are open-ended, inviting the students reasoning through

the scenarios from *Theory->Evidence*, *Evidence<=>Data*, and *Evidence->Theory* relationships. All questions aimed to provide students with inquiries to link their biological disciplinary knowledge with epistemic considerations, by asking them to use reasoning and their biological disciplinary knowledge and to justify the answers in order to reveal their epistemic considerations. Below we provide the assessments, example expert answers, and then we discuss selected examples of students' answers that meet expectations and others that do not meet expectations to show how observation and interpretation of the variation in performance works with these assessments.

17.3.3.1 Assessment Items Informed by CADE

A question about whale evolution was used as a pre-test at the start of an undergraduate biology lab course and a question about echidna evolution was used as a post-test item on the final exam. In addition to a scenario, each assessment had three probing questions: "Why did the two scientists make different decisions about what types of evidence to gather" to assess students evidentiary reasoning under the *Theory ->Evidence* relationship; "Which scientist provides the strongest evidence for their claims" under the *Evidence->Theory* relationship; and "What additional kinds of evidence to consider and why" under the *Evidence<=>Data* relationship in terms of the CADE practices of reasoning with and about the evidence.

Box 17.1: Whale Assessment

It has been long established that whales are mammals, but scientists are not yet certain of their exact ancestry and which current species are their closest living relatives. Two scientists told our local news reporter their ideas about whale evolution:

Scientist Sandra Wells says:	Scientist Rosendo Pascual says:
The manatee is the closest living relative of the whale because manatees have flippers and tail structures more like whales and can spend long periods of time under the water like whales. We also found dozens of DNA sequences shared by whales and manatees.	The hippopotamus is the closest living relative of the whale. We found a fossil of the hippo's ancestor with a complete ancient skeletal remain like the backbone of a whale. It also had limb and teeth structures found in the modern hippopotamus. We even found one DNA sequence that whales and hippopotami share.

1. Why did these two scientists make different decisions about what types of evidence to gather and how did their assumptions influence the quality and the accuracy of their claims?

2. Which scientist (Dr. Wells or Dr. Smith) do you believe provides the strongest evidence for their claims about the closest living relative of the whale? Explain the reasons for your answer.

3. What additional kinds of evidence should the two scientists consider? Why you think these additional kinds of evidence might be useful to test their ideas?

Box 17.2: Echidna Assessment

It has been long established that there are many different types of mammals, but scientists are not yet certain of their exact ancestry and which groups are more closely related. Two scientists told our local news reporter their ideas about echidna evolution:

Conservationist Mandy Watson says:	Scientist Rosendo Pascual says:
The bandicoot is the closest living relative of the echidna. Both have long slender snouts that function as both mouth and nose and both feed primarily on earthworms. Both are found in Australia near a water supply. Throughout Australia we found four kinds of fossil bandicoots and also fossils of the echidna, both with short, strong limbs and claws for powerful digging. A collaborator found many DNA sequences shared by modern echidnas and bandicoots.	The duck-billed platypus is the closest living relative of the echidna. In both animals the upper appendage bones are held roughly parallel to the ground when the animal walks, more like most modern reptiles. The platypus has a cloaca through which eggs are laid and both liquid and solid waste is eliminated. The echidna also has one body cavity for the external openings of the urinary, digestive, and reproductive organs. We even found one DNA sequence that modern platypus and echidna share.

Animal Names	Eastern Barred Bandicoot	Long-beaked Echidna	Duck-billed Platypus
Average mass	640–766 g	11 kg	1.52 kg
Average basal metabolic rate	1.902 W	6.493 W	1.931 W

1. Why did these two scientists make different decisions about what types of evidence to gather and how did their assumptions influence the quality and the accuracy of their claims?

2. Which scientist (Dr. Watson or Dr. Pascual) do you believe provides the strongest evidence for their claims about the closest living relative of the echidna? Explain the reasons for your answer.

3. What additional kinds of evidence should the two investigators examine? Explain why they should consider that evidence and why you think this additional evidence is reasonable to consider.

17.3.3.2   Expert Answers for Whale and Echidna Questions

Expert answers to both the whale and echidna questions provide examples of how the three questions in each assessment were able to invite reasoning through the different subcategories within the research practice categories of the CADE, linking disciplinary knowledge with epistemic considerations. For example, regarding T -> E, the model articulation component of evidentiary reasoning is evident when decisions about the evidence to be examined are informed by disciplinary knowledge of relevance to the investigation. **Bold font** in the answers indicate that **alternative models or theories are considered**, which is an epistemological consideration for T -> E. Regarding E <=> D, expert responses to both assessments consider comparison of DNA sequences as a research method but they also bring up several ideas about different types of data to collect from diverse measures to provide additional

support, which is an epistemological consideration for E <=> D. Regarding E -> T, epistemological considerations about using *convergent evidence to draw conclusions* are indicated with an *italics font,* and biological disciplinary knowledge of relevance to decisions about the evidence or conclusions is underlined.

WHALE ASSESSMENT EXPERT ANSWER

Whale Q1: **Dr. Wells is assuming that the morphological, behavioral, and some genetic similarities between manatees and whales are an indication of phylogenetic similarity.** These assumptions do not necessarily account for which DNA sequences are shared. Mammals generally have many homologous sequences within their genome, even taxa that are not closely related (e.g. there is genetic similarity between dogs and wallabies). Morphological and behavioral similarity could be a sign of convergent evolution rather than phylogenetic similarity (e.g. sugar gliders and flying squirrels share many morphological and behavioral traits even though they are not closely related). These assumptions could skew the interpretation of evidence by Dr. Wells whose conclusions may not be accurate.

Whale Q2: More data is needed to understand which claim is better supported by evidence. While shared physical and DNA traits could be an indicator of phylogenetic similarity, this evidence alone is not enough. Different animal lineages may share physical and behavioral traits while not being closely related. Similarly, common ancestors in the fossil record may give evidence for relatedness, but all mammals have a common ancestor if one looks back far enough.

Whale Q3: *Knowing which DNA sequences are shared between whales and manatees, the timing of the common ancestors (between the hippo and the whale) when whales and hippos diverged, and when whales and manatees diverged (evidence supporting/refuting convergent evolution) are some of the information that would help to fully understand which claim (if either) is more accurate. If the shared DNA sequences are unique to aquatic mammals (i.e. dealing with flipper formation), then this could be strong evidence for relatedness. However, if the DNA sequences are common among all mammals (i.e. general vertebrae formation), then this evidence may not be very robust. The time periods when the common ancestor existed and when whales and hippos diverged would be useful to know because if the ancestor is significantly more ancient than when these animals diverged, this common ancestor may not be a strong indicator of relatedness. Knowledge of when whales and manatees diverged could indicate whether similarities or differences suggest coevolution or speciation. If whales and manatees diverged very far back in time, then the common morphological and behavioral traits are likely due to coevolution. However, if they diverged recently, then the commonalities between whales and manatees may be strong evidence for phylogenetic similarity.* [Expert answer from graduate student in Ecology and Evolutionary Biology.]

ECHIDNA ASSESSMENT EXPERT ANSWER

An expert response to the echidna assessment shows disciplinary knowledge of relevance to the investigation including the depiction of two alternative cladogram models in this figure that was applied to decisions about evidence to be examined (Fig. 17.1).

The following expert response to the echidna assessment considers **alternative models or theories**, indicated as **bold font**, the use of *convergent evidence* to draw conclusions is indicated in *italics*, and underlined text indicates reasoning about evidence that is informed by biological disciplinary knowledge.

Echidna Q1. Mandy Watson is a Conservationist, so she studies where organisms live and what they eat in order to conserve living organisms along with their environment. She observed that both bandicoot and the echidna have long slender snouts as well as claws for powerful digging, and that both feed on earthworms and live near water. Historically, these types of observations were used to classify animals into groups when they were identified and named. As a conservationist, Mandy may not know much about evidence from a collaborator who found many DNA sequences shared by modern echidnas and bandicoots. Was she **wondering if the platypus shares those same sequences? Perhaps she has not considered that different animals generally have some homologous sequences within their DNA. More evidence is needed to determine if the structures she described represent homology (inherited from a shared ancestor) or homoplasy, which refers to structural similarity such as from convergent evolution rather than from a recent shared ancestor. Many examples of convergent evolution are found in animal morphology or the fossil record, where we find evidence that environments shape organisms. When some individuals from distantly related taxa are more likely to survive and reproduce, they become more similar because both are fit to eat similar food in similar environments.** Instead, an evolutionary biologist would use evidence to establish the chronology of evolution and common ancestry, not just with biogeographical evidence of fossils, but instead by using the fossils and other data to identify derived traits that distinguish organisms on one branch of their family tree from those on the other branches. An example in this case is that most mammals walk on four legs holding their body upright, unlike the duck-billed platypus and the echidna that walk like most modern reptiles, which is with their upper appendage bones being held roughly parallel to the ground, according to the Scientist, Rosendo Pascual. Rosendo may have been considering the chronology of their ancestry in noticing that the bandicoots have a trait like most modern mammals, unlike the duck-billed platypus and the echidna.

Echidna Q2. To decide who provides the strongest evidence, **consider a diagram of the different models being suggested (Fig. 17.1)**. Mandy's use of anatomical (snout and claws), food source, and biogeographical evidence leads her to believe that the bandicoot and echidna are sister taxa. In contrast, Rosenda places a more recent ancestor as one that
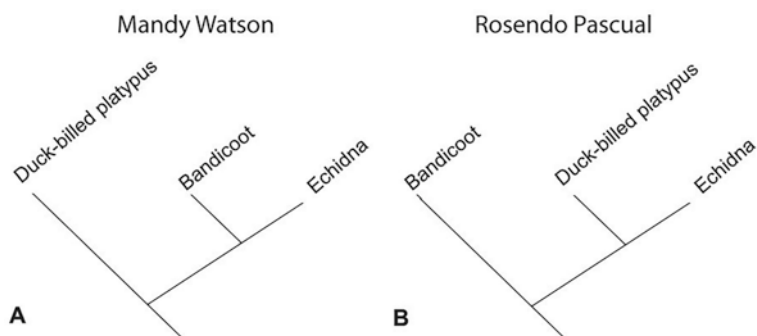

Fig. 17.1 According to an expert's answer, these cladograms depict alternative models for the chronology of ancestors shared among these three types of animals. The branch with bandicoot and echidna as sister taxa in panel A illustrates Mandy Watson's idea that bandicoot and echidna are more closely related whereas the branch with platypus and echidna connected by a more recent ancestor in Panel B illustrates Rosenda Pascual's claim. In both diagrams, all three share an ancient ancestor indicated by the branching point at the bottom of the tree, but the sister taxa share a branch through a connection to a more recent ancestor that the third group or outgroup does not share

is shared by the platypus and echidna. I agree with Rosendo because I know that the duck-billed platypus and the echidna are both egg-laying mammals. If I accept her description of *their upper appendage bone structures and walking behavior and with both having a clo-aca for the urinary, digestive, and reproductive organ external openings, this is strong evi-dence for* the sister taxa branch *model in panel B. In fact,* despite it's larger size, according to the data table provided, the higher metabolic rate of the bandicoot compared with the platypus and echidna provides additional *evidence* that it may be more closely related to other mammals like kangaroos or carnivorous mammals that are also quite warm-blooded. *However**, this evidence does not rule out the possibility that echidna and platypus are paraphyletic, with other mammals being a monophyletic subgroup **in which case Mandy's model could be right if data from fossil bones and DNA sequence homology both suggest that the platypus is** an outgroup for a clade **that includes all other mammals and the echidna according to the sister taxa branch model in panel A.*

Echidna Q3. As additional evidence, both could collaborate to more carefully examine the biogeographical and chronological history of the four kinds of fossil bandicoots and the fossil echidna. I would expect to find the upper appendage bones for the short, strong limbs and claws in the fossil echidna to be held roughly parallel to the ground, but do any of the fossil bandicoots have that anatomical feature? I would like to know whether features and the chronology of the fossil data suggests echidna and bandicoots share a more recent ancestor as in panel A and if additional data - not just one DNA sequence that modern platypus and echidna share, but instead a thorough comparison of homologous DNA sequences among all three animals - suggest that platypus and echidna share more recently derived traits. Neither Mandy nor Rosenda have stated whether DNA sequences shared by the two they group together are found to be more different in the third group. In summary, to rule out either model A or B, there is a need for evidence_of a more recent ancestor link-ing the two that are closest living relatives, leaving the third as an outgroup. *Evidence of homologous traits shared among the two but that is missing from the third and from other living mammals, and ruling out the possibility that the trait could result from convergent evolution, are two uses of evidence that could converge in strengthening claims about which two could be closer living relatives.* [Expert answer from a professor who teaches an Evolutionary Biology course for science teachers.]

17.3.4   Findings from Typical Examples of Students' Answers to the Whale and Echidna Questions

For our second research goal about how well the CADE-informed assessments reveal students' difficulties in evidentiary reasoning, we provide some typical stu-dents answers as examples. The following student answer examples were selected to illustrate how the gaps that were found in our review of published assessments have been addressed by the whale and echidna questions used as pre- and post-test in an introductory level undergraduate biological lab course. The examples range from good answers that cover many aspects of the notion of evidence in the reason-ing process of research practices as well as answers that fail to consider some aspect of the notion of evidence during reasoning. Again, the biological disciplinary knowledge within the reasoning is underlined.

17.3.4.1 The Assessments Probed Evidentiary Reasoning with Disciplinary Knowledge Linked to Epistemic Considerations

Student answer examples illustrate how disciplinary knowledge was linked with epistemic considerations in evidentiary reasoning.

> Whale Q3: These additional forms of evidence can reveal more detailed information about how closely the species are related. Comparing <u>mitochondrial DNA</u> is important because it is <u>present in most cells in an organism</u>, it <u>evolves quickly but at a known rate</u>, and <u>it is passed down through the maternal line</u>.

Above, the student applied specific disciplinary knowledge about mitochondrial DNA when using general epistemic considerations to justify the value of including mitochondrial DNA as additional evidence in determining how closely the species are related.

> Echidna Q3: The scientists should always examine DNA evidence further to establish an even stronger connection. <u>Similarities in the genome of two species</u> can greatly bolster any potential relationship and can <u>point to the closeness of the two species</u>. They should also delve deeper into the <u>molecular use of proteins, enzymes, metabolic pathways,</u> etc. to show that the two species utilize such compounds similarly. Also, examining the <u>species habitats and niches can point to commonalities</u>. If the two species live in similar environments and occupy the same niche and carry out the same functions in their ecosystems, that may be because they are closely related. Having the same <u>homologous and vestigial structures</u> can also be indicators of common ancestry.

Above, the student linked specific disciplinary knowledge about similarities in genome, molecular use of proteins, enzymes, metabolic pathways, species habitats and niches with the epistemic considerations to justify the value of those data as additional evidence in determining the closest living relatives.

17.3.4.2 Some Responses Described Disciplinary Knowledge But Failed to Link to Epistemic Reasoning About the Relevance or Quality of Evidence

Next we provide several examples of student answers that failed to link disciplinary knowledge with general epistemic considerations. Using guiding questions in the CADE table at https://tinyurl.com/CADE2022 for scoring, they failed to get full credit in all three Science Research Practices for reflective evaluation or critique of the evidence (epistemic considerations) in ways that link to the relevant disciplinary knowledge of biology.

> Whale Q2: I believe Dr. Smith provides the strongest evidence for her claims about the closest living relative of the whale because she compares <u>skeletal structure</u>.

This student correctly identified skeletal structure as relevant biological knowledge for determining evolutionary relationships. However, they failed to note specific properties of skeletal structures to examine or how the skeletal structure data might

be used as evidence. They could have detailed the quality and scope of structural data to use as evidence for a better reasoning process (epistemic considerations).

Without considering the relevance or quality of the evidence, the next examples of student responses imply that having MORE data is better without justification, which was common among undergrad student responses that failed to link to epistemic reasoning to justify the relevance, scope, or quality of evidence.

> Whale Q1: Scientist Sandra Wells based her conclusion on external morphological observations between manatees and whales. Ashley Smith based her conclusions on internal skeletal observations, comparing bone structures of the whale and hippo. Both Scientists found DNA segments that link both the hippo and the manatee to the whale. They both have evidence that can suggest logical assumptions about whale evolution and have genetic evidence to support their claims. The biggest difference among their findings, however, is Wells discovered DOZENS of DNA sequences and Smith found A DNA sequence.
>
> Whale Q2: I believe Dr. Wells' findings are more persuasive due to the dozens of DNA sequences found. If she just based her evidence on morphology, I wouldn't have found her evidence strong. But, since she presented DNA evidence to suggest linkage, her evidence is stronger.
>
> Whale Q3: Some other types of evidence that could be useful is dietary evidence. Do they have similar diets? Other possible evidence is behavioral aspects, life histories, mating strategies, geographical distributions, derived traits, and ancestral traits. Use the evidence gathered to start developing theories of how that animal evolved. After more information is gathered, then start placing the animal in a phylogeny and determining plausible common ancestors. This evidence can start to lead to answers and develop more theories that explains whale evolution better.

The above response thoroughly describes some useful evidence by repeating the information provided in the assessment in their response and then also listing a few additional types of biological evidence. Although there is mention of the need to "start developing theories of how that animal evolved," it does not explain how evidence might be used to establish the chronology in a phylogeny.

The next example illustrates this same problem of not considering the nature, scope, and quality of data for use as evidence, and the response again illustrates the "MORE data is better" superficial reasoning about evidence that we often find among undergrads.

> Echidna Q3: Other sorts of evidence that the scientists could consider are diets, lifespan, behavior (whether they live in groups, how long/if the offspring stay with their mother, how aggressive they are towards members of their own species or predators, level of activity etc.), body mass, migration patterns or whether or not they live in a singular area, feeding habits, and many more. Additional pieces of evidence are always very useful ... More data that can be collected can help refine your conclusion and it can help support the claims that you have already made.

Above, the student lists quite a bit of biological disciplinary knowledge, such as diets, lifespan, behavior as relevant to determining closest relatives. However, the student failed to link the disciplinary knowledge with general epistemic considerations in justifying reasons for including the data as evidence.

17.3.4.3 Student Answer Examples Discuss Convergent Evidence That Could Support or Raise Questions About the Strength of an Inference

Many student answers illustrate how convergent evidence (italicized words) are used in drawing conclusions, which is an epistemic consideration worth mentioning in justifying ideas about using more data in evidentiary reasoning.

> Whale Q1: Scientist Sandra Wells decided to gather more observational evidence (noting the physical appearance and characteristics of the modern whales and manatees), whereas Scientist Ashley Smith decided to gather more concrete evidence from past skeletal remains. This highlights the difference between their two approaches: Scientist Sandra Wells wanted to look more at current day observations, but Scientist Ashley Smith wanted to look at the past (hence the archaeological evidence). Scientist Sandra Wells' approach of just focusing on present day observations means that the quality and accuracy of her claims are not strong because she only has one part of the picture. You cannot just look at observations and make the assumption that it implies to a different concept, which <u>in this case was ancestry</u>. Many animals look closely related, but that is not enough to make such a broad claim (she needs more concrete evidence). Although she refers to <u>similar DNA sequences, they could just be extremely common sequences found between multiple organisms</u>. Scientist Ashley Smith's approach leads to the higher quality and accuracy of her claims because she looked at similarities in fossils, which is more concrete evidence. In order to make such a broad claim, however, *various types of approaches should be used*.
>
> Whale Q2: I believe that Dr. Smith provides the strongest evidence for her claim that the hippopotamus is the closest living relative of the whale. Her evidence, mainly the fossil of the hippo's ancestor, is more concrete than Dr. Wells' observational evidence on characteristics. Although both methods are important in science, <u>Dr. Smith's approach makes the most sense given that they are looking at ancestry</u>. Additionally, both scientists mention <u>similarities in DNA sequences, which is also good evidence</u>.
>
> Whale Q3: Additional kinds of evidence that <u>*the two scientists should consider is similarities in sequences of RNA, comparing organ systems, or looking more in depth at previous fossils and comparing their similarities (or the presence of homologous parts).*</u> I think that *additional kinds of evidence are useful to test their ideas because it offers a different perspective that will provide additional insight on each of their claims*. The more evidence that supports a claim, the more valid it becomes. On the other hand, *if other types of evidence don't support a claim, it's a good indication that the claim needs to be reevaluated*.

As with the expert responses, disciplinary knowledge has been underlined and the italicized words highlight ideas about convergent evidence. Both the previous example and the next connect epistemological considerations about the scope and relevance of the evidence with their disciplinary knowledge to highlight the chronology of biological evolution.

> Echidna Q1: The conservationist and scientist relied on their different scientific educations and exposure to distinctive scientific literature within their specialties to identify and evaluate evidence. The conservationist is trained to analyze the ecological interactions of the two species under the umbrella of environmental sciences. On the other hand, the scientist is trained to recognize their similar anatomical and physiological features. Their assumptions are expected to introduce some bias into the process of evidence selection as they could, unintentionally or intentionally, gravitate towards familiar explanations in regards to their different backgrounds. This is likely to have an effect on the final claim and reduces its accuracy and quality.

Echidna Q2: In my opinion, the scientist provides the strongest evidence for their claims because of their precise comparisons between the echidna and platypus. Anatomical similarities are discussed and encompass the entire organism including musculoskeletal system, digestive system and reproductive system. These are precise observations rooted in biology that strongly suggest that both species evolved from a recent, shared common ancestor. Furthermore, the investigators themselves identified a shared DNA sequence. This is in contrast to the conservationist who relied on a "collaborator" to provide DNA testing results. This shows that, in the case of the scientist, the process of data collection was controlled and accountable. The conservationist's claims are too broad to establish a close evolutionary relationship between the bandicoot and echidna. The statement that "both are found in Australia near a water supply" does not indicate a relationship because a vast majority of animals are expected to live near water supplies. The conservationist also notes that both bandicoot and echidna fossils were found in Australia. However, considering the geographical size of the country, this is insufficient because a variety of fossils are also found in Australia that do not necessarily imply that they all came from the same lineage. Similarly, the claim that both species feed on earthworms is also too broad because diet similarities is a tenuous relationship especially compared to the anatomical similarities discussed by the scientist.

Echidna Q3: Both investigators should consider the evolutionary history of echidna and trace its lineage back to a shared common ancestor with either the platypus or the bandicoot. This additional evidence is reasonable because they are attempting to determine the closest living relative of the echidna therefore by drawing its phylogenetic tree, _the investigators should be able to identify similar morphology and molecular characteristics_. _Fossil records will also be useful in this case as they will aid in pinpointing the moment of divergence_ of the bandicoot and platypus from the echidna. The closest living relative will therefore be the species that diverged most recently.

Here is another example from a student who was able to link their disciplinary knowledge about fossil structure and the type of creature with epistemic considerations about the use and quality of the evidence to test a model and draw conclusion.

Whale Q2: Dr. Smith, on the other hand, mainly looked at ancestral skeletal structures and these can easily evolve over time. Why yes, Smith found that the hippopotamus and whale may have been related quite a long time ago, but the two species have evolved greatly. The hippopotamus is both a terrestrial and aquatic creature, while a whale is only an aquatic creature. Also, Dr. Smith could only find one common DNA sequence between the hippo and the whale, which also shows that Wells' findings are more accurate.

In this case above, the student argued that considering the huge difference between the hippopotamus and whale nowadays, without knowing the age of the fossils, and using homologous structure of fossils and only one common DNA sequence alone would be insufficient evidence for testing the model.

Echidna Q2: Both scientists decided to take DNA into account. This is reasonable because shared sequences maybe link to ancestry. Both also decided to note the anatomy of each creature, citing similar features that could be a sign of homology. Watson looked at geographical location and fossil history which may point to a shared point of evolution considering conditions for the time period.

Above, the student used disciplinary knowledge related to DNA, anatomy, homology, geographical location and fossil history linked with their justification like similar features may indicate homology in evaluating the conclusion about closest relatives. The way the student organized different kinds of evidence illustrates how

convergent evidence was used to draw conclusions. Specifically, they pointed out that similarity in fossil structure indicates homology and that geographical location and fossil history can further help the scientist to identify the time period for speciation.

Of course, there were many examples of student answers that failed to use convergent evidence in drawing conclusions.

> Whale Q2: I believe that she is correct due to the fact she found multiple DNA sequences shared between the two of them.

In this example, the student relied on one single kind of evidence, the similarity in DNA sequences, in drawing conclusions. In fact, quite a few students claimed that the best and most sufficient evidence is based on multiple DNA sequences. By doing so, the student ignored other kinds of evidence, such as homology in fossil structures and chronology of the fossil evidence in reasoning about the models being tested.

> Echidna Q2: Neither scientist provides the strongest answer since both only used one criteria and method to support their assumptions.

Above, the student claimed that both scientists use only one criterion to support their assumptions. They failed to identify that each scientist used multiple kinds of evidence in drawing conclusions.

17.3.4.4 Some Responses Failed to Use Appropriate Disciplinary Knowledge to Inform a Hypothesis or Research Goal

There were many good answers such as the following which illustrates how students reasoned from theory to inform their hypothesis.

> Whale Q1: These two scientists made different decisions about what types of evidence to gather because they both felt that there were different characteristics that determined whether or not a certain animal evolved from another certain animal. One scientist, Dr. Wells, thought that you could determine if two animals were related based on the structures of the outside parts of their bodies (flippers, tails, can breathe in water, etc.) and their DNA sequences. The other scientist, Dr. Smith, thought you could determine if two animals were related based on the structure of the inner parts of their bodies (bones/teeth) and their DNA sequences."

Above, the student successfully identified reasons for different scientists to gather different evidence informed by a foundation of different knowledge and theories. One scientist based their hypothesis upon similarities of outside structure and DNA, while the other considered similarities of inner structure and DNA.

> Echidna Q1: The conservationist and scientist relied on their different scientific educations and exposure to distinctive scientific literature within their specialties to identify and evaluate evidence. The conservationist is trained to analyze the ecological interactions of the two species under the umbrella of environmental sciences. On the other hand, the scientist is trained to recognize their similar anatomical and physiological features.

Above, the student identified that conservationists inform their decisions based on their training to analyze ecological interactions in the context of environmental sciences, whereas the scientist was trained to recognize similar anatomical and physiological features.

When limitations to the use of modern DNA evidence for establishing a phylogeny were discussed in class, some students did not realize that we were discussing how evidence might be used in combination with other evidence of shared derived traits that might suggest a recent ancestor shared by two sister taxa but that is not shared with an outgroup clade. For example, the next response shows that potential use of DNA sequence data was not understood by a student who complained that "both scientists use DNA sequencing to prove their point (bad science)." With both questions, many responses showed difficulty understanding relevant disciplinary knowledge or, as in this case, they revealed wrong ideas about how to apply disciplinary knowledge when deciding how data might be used as evidence.

> Echidna Q1: Those two scientists made different decisions because they most likely work in different fields, so depending on their specialty this would have lead them to their particular claims. Both claims could be accurate; however, the finding of shared DNA sequences is irrelevant when trying to connect the species.

> Echidna Q2: Disregarding that both scientists use DNA sequencing to prove their point (bad science), I believe Dr. Pascual provides the strongest evidence because he relates the anatomy and physiology of both species. Ultimately, no scientist is more right than the other one because they both use valid arguments and data.

> Echidna Q3: The scientist can look into where all three species originated. This information can provide crucial geographical information as to if the species is native to where it's living - whether it naturally has always eaten, reproduced, and looked that way - or if they are an invasive species or migrated over there and have had to adapt these characteristics.

We also found many examples where students failed to recognize that different hypotheses were informed by different disciplinary approaches to research.

> Whale Q1: They made different decisions because they both looked at different types of data and interpreted it in different ways.

> Echidna Q1: These two scientists made different assumptions based on the specific data they each had.

In these examples, the students failed to consider any theory or disciplinary knowledge that informed each hypothesis.

17.3.4.5   The Assessments Probed Evidentiary Reasoning About Whether Alternative Model Had Been Considered

Many responses failed to consider alternative models and how data might be used as evidence to rule out one of the models, as illustrated with the following example response.

> Echidna Q1: These two scientists made different decisions because each scientist has a different way of thinking based on the information given to them. Their assumptions influenced the quality and the accuracy because both tested their hypothesis on what they believed to be true and both made observations that they can back up with evidence.

> Echidna Q2: Dr. Pascual provides the strongest evidence because he found evidence based on the animals' reproduction and bone structures plus the fact that they found evidence of sharing a DNA sequence.

> Echidna Q3: They should consider their genotypes because this will show a tree how each generations' alleles could have altered to provide evidence of evolution.

However, several examples of student responses correctly illustrate skepticism toward data and they raised alternative models to consider as they attempted to rule out some alternatives with evidentiary reasoning.

> Whale Q2: Although she refers to similar DNA sequences, they could just be extremely common sequences found between multiple organisms.

> Echidna Q2: Humans share many DNA sequences with bananas, and we aren't related to fruits, so I don't think that is sufficient evidence.

Other examples of student responses failed to identify the alternative models.

> Whale Q2: I believe Dr. Wells provides the strongest argument, because her evidence is based on the physical characteristics of how the manatee and whales look now. Their similar structures and DNA sequences can infer that they can be related through evolution.

In the above example, the student failed to identify that the similar structures shared between manatee and whales may be caused by selection pressures (homoplasy) instead of inherited from a common ancestor (homology).

> Echidna: Data such as structure, diet, habitat, and fossil record are all important. With the DNA sharing more sequences, I feel they are closer in relation as well.

In this example, the student simply listed data that was provided in the scenario of the question without reasoning about the evidence.

17.4   Summary and Discussion

The CADE provided the cognition foundation for evidentiary reasoning as a target for assessing how well students understand and would be able to do more authentic biological research. Our first study aimed to gain an understanding of established assessments that are being used to evaluate students evidentiary reasoning. We found that there is a need for assessments to track progress as students learn to reason through their decisions about evidence. Areas that still need to be more carefully examined include how well the students link epistemic considerations with their disciplinary knowledge to inform decisions about evidence, their use of disciplinary knowledge of relevance to the investigation to inform a hypothesis or research goal,

how diverse evidence is used to establish sufficiency of evidence such as by using convergent evidence to strengthen conclusions, and whether alternative models or theories are considered.

CADE-informed assessments helped to address these gaps by revealing competent reasoning about evidence and students' difficulties with evidentiary reasoning in a meaningful way in the context of evolutionary tree-thinking. Guiding questions in the CADE table at https://tinyurl.com/CADE2022 highlighted what to observe. The findings revealed difficulties that could be addressed by instructors and students.

According to the assessment triangle (National Research Council, 2001; Pellegrino, 2012), the foundation for assessment design is to identify a cognition model that shows how leaners typically represent information and build domain expertise. Here we adapted the CADE (Samarapungavan, 2018) to clarify how each research practice in biology involves disciplinary knowledge integrated with epistemic considerations as a cognitive foundation for understanding how students could demonstrate evidentiary reasoning throughout the process of biological research practice. A realistic research scenario as an assessment with three open-ended questions provided a useful test item that targeted three research practice relationships to provide students with opportunities to demonstrate comprehensive evidentiary reasoning with evolutionary tree-thinking as a biological context. In our implementation of the assessments, we were able to observe and compare biology experts' and students' evidentiary reasoning in terms of the targeted cognitive competencies. Examples of both expert and student answers show that assessments informed by the CADE framework were able to reveal aspects of the reasoning process of relevance to three types of science research practices: Theory to Evidence Relationships (T -> E) involve formulating testable models, hypotheses or explanations; Evidence <=> Data Relationships (E <=> D) relate to designing, executing, and analyzing data from investigations; and Evidence to Theory Relationships (E->T) relate to inferences and the sufficiency of conclusions. The E->T component reflects evidence-based reasoning in science (Erduran et al., 2015; Furtak et al., 2010; Toulmin, 1958; Tytler & Peterson, 2005) but the CADE more comprehensively highlights T->E and E <=> D relationships as well. Furthermore, the CADE framework provides a practical framework for interpreting student answers to reveal their difficulties with evidentiary reasoning. Example responses from undergraduate students show that some students have difficulty in using convergent evidence to draw conclusions. Instead of constructing a model to organize diverse evidence to draw conclusions, some students claim that more evidence is needed but without justifying why the evidence would be useful, some rely on one kind of evidence, or they emphasize the value of DNA sequence data while ignoring other evidence that would be useful for determining phylogenetic relationships among different animals.

The CADE-informed assessment examples should, however, be modified or new assessments developed for any other research context or subdiscipline in biology, such as experimental design, microbiology, or immunology. By altering the story about investigators with different disciplinary knowledge and by presenting a conflict that encourages the students to link their relevant knowledge of biology with

epistemic considerations, the CADE framework could be tested as a cognitive foundation model in other biology learning situations.

17.5   Conclusions

Use of the CADE framework as a lens made it possible to delve deeply into evidentiary reasoning in a comprehensive way that includes identifying a research problem and planning and conducting research in addition to evidence-based reasoning about scientific evidence for backing claims. CADE-informed assessments revealed students' evidentiary reasoning difficulties by emphasizing the fundamental role of epistemic cognition, and by further linking epistemic cognition with disciplinary knowledge. The CADE was compatible with ACE-Bio Competencies to provide meaningful insight into the meaning of evidentiary practices throughout the research process and it helped inform how to evaluate students' evidentiary reasoning.

   Several gaps were identified by the literature review: (1) Some assessments and rubrics failed to link disciplinary knowledge with epistemic considerations while assessing students' evidentiary reasoning; (2) Few assessments target student competence within the Evidence <=> Data relationship of relevance to *Conducting* a research study and no published assessment evaluated students evidentiary reasoning about the necessity of using diverse evidence in drawing conclusion, or how to use convergent evidence to draw conclusions; (3) Only two assessments in our study measure students competence to *Identify* a research problem with Theory->Evidence relationships where students must reason through the need to consider alternative models or theories and for making decisions about the evidence to be examined using disciplinary knowledge of relevance to the investigation. Our CADE-informed assessments addressed these gaps by revealing students' evidentiary reasoning competence and difficulties by emphasizing the fundamental role of epistemic cognition, and by further linking epistemic cognition with disciplinary knowledge.

   The CADE framework proved to be a useful guide for revealing students' difficulties with evidentiary reasoning. In terms of the assessment triangle (National Research Council, 2001; Pellegrino, 2012), it provides a cognition target detailing several types of reasoning about evidence that the assessment task should elicit. It also targets components to notice in expert responses to the assessment and it facilitates creating new assessments where the key components can be observed in student responses. When those components were missing from student responses, it helped with interpretation to target those areas of difficulty to address in the future (i.e. the linking of biological knowledge with epistemic considerations while considering both domain-general and discipline-specific aspects of evidence).

   As an implication, the findings of this research will benefit the teaching and assessment of learning about evolutionary trees by providing educators and students with a feasible way to deconstruct and unpack the notion of evidence for the various tree diagrams. The findings also provide insight into assessment instrument choices and the design of new assessment tools for use on tests to reveal students'

difficulties with evidentiary reasoning in biology as a discipline. The CADE table at https://tinyurl.com/CADE2022 is provided in a digital format to be easily modified for use in defining the target cognition or in developing scoring rubrics so that others can use CADE for their own purposes. Future studies should continue to modify the framework as it is applied to other contexts to help develop additional understanding for how evidentiary reasoning can be assessed in science students.

In summary, the CADE framework unpacks the notions of evidence into component parts (Samarapungavan, 2018) that might be applied to assess learning about evidentiary reasoning in other research contexts together with the assessment triangle (National Research Council, 2001; Pellegrino, 2012). Based on our work and synthesis of the literature, we recommend considering the following ideas and approaches when developing assessments to monitor students' competence for reasoning with and about evidence in the context of science investigations:

- According to the CADE framework, comprehensive knowledge of the quality and use of scientific evidence involves several inter-related research practices: theoretical knowledge informs what evidence is relevant, disciplinary practical knowledge guides collection of data for use as evidence, and interpretation of evidence to refute, confirm, or advance knowledge is informed by existing disciplinary knowledge.
- The discipline-specific components of scientific research practices of relevance to an investigation are linked to the domain-general epistemology of research.
- The assessment triangle is a useful guide with the CADE for implementing discipline-specific assessments to track students' evidentiary reasoning: first, use the CADE to detail a cognition target by unpacking the various types of reasoning about evidence that an assessment task should elicit, then implement the task to observe the range of student behaviors the task reveals, and finally interpret students' competence or their difficulties that should be addressed according to a CADE table from https://tinyurl.com/CADE2022 that can be modified for a particular research task.
- Prompt students to link disciplinary knowledge with epistemic reasoning as they consider how a hypothesis or research goal was informed, what alternative models to consider, or to evaluate sufficiency of claims in terms of relevant and convergent evidence.

References

Abd-El-Khalick, F., Boujaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., & Tuan, H. L. (2004). Inquiry in science education: International perspectives. *Science Education, 88*(3), 397–419.

American Association for the Advancement of Science (AAAS). (2011). *Vision and change in undergraduate biology education: A call to action* (C. Brewer & D. Smith (Eds.)). https://visionandchange.org/finalreport/. Accessed 18 Oct 2021.

Anderson, W. L., Sensibaugh, C. A., Osgood, M. P., & Mitchell, S. M. (2011). What really matters: Assessing individual problem-solving performance in the context of biological sciences. *International Journal for the Scholarship of Teaching and Learning, 5*(1), n1.

Angra, A., & Gardner, S. M. (2017). Reflecting on graphs: Attributes of graph choice and construction practices in biology. *CBE—Life Sciences Education, 16*(3), ar53.

Auerbach, A. J., & Schussler, E. E. (2017). Curriculum alignment with vision and change improves student scientific literacy. *CBE—Life Sciences Education, 16*(2), ar29.

Bautista, N. U., Schussler, E. E., & Rybczynski, S. M. (2014). Instructional experiences of graduate assistants implementing explicit and reflective introductory biology laboratories. *International Journal of Science Education, 36*(7), 1184–1209.

Beck, C. W., & Blumer, L. S. (2021). The relationship between perceptions of instructional practices and student self-efficacy in guided-inquiry laboratory courses. *CBE—Life Sciences Education, 20*(1), ar8.

Bhavana, A. R. (2009). Honing language skills using blogs. *Journal on School Educational Technology, 4*(3), 30–33.

Black, P., Harrison, C., Lee, C., Marshall, B., & Wiliam, D. (2003). *Assessment for learning: Putting it into practice*. Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice, 5*(1), 7–74.

Blair, A. C., Peters, B. J., & Bendixen, C. W. (2014). Promoting student inquiry using Zea mays (Corn) cultivars for hypothesis-driven experimentation in a majors introductory biology course. *The American Biology Teacher, 76*(5), 333–336.

Boomer, S. M., Kumar, S., & Dutton, B. E. (2021). A CURE for meat: Comparing bacterial contaminants on different ground beef sources emphasizes process of science and quantitative reasoning. *Journal of Microbiology & Biology Education, 22*(1), ev22i1-2237.

Brownell, S. E., Wenderoth, M. P., Theobald, R., Okoroafor, N., Koval, M., Freeman, S., Walcher-Chevillet, C. L., & Crowe, A. J. (2014). How students think about experimental design: Novel conceptions revealed by in-class activities. *BioScience, 64*(2), 125–137.

Brunauer, L. S. (2016). Purification and electrophoretic characterization of lactate dehydrogenase from mammalian blood: A different twist on a classic experiment. *Journal of Chemical Education, 93*(6), 1108–1114.

Bugarcic, A., Zimbardi, K., Macaranas, J., & Thorn, P. (2012). An inquiry-based practical for a large, foundation-level undergraduate laboratory that enhances student understanding of basic cellular concepts and scientific experimental design. *Biochemistry and Molecular Biology Education, 40*(3), 174–180.

Creswell, J. W., & Poth, C. N. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). SAGE.

Coleman, A. B., Lam, D. P., & Soowal, L. N. (2015). Correlation, necessity, and sufficiency: Common errors in the scientific reasoning of undergraduate students for interpreting experiments. *Biochemistry and Molecular Biology Education, 43*(5), 305–315.

Dasgupta, A. P., Anderson, T. R., & Pelaez, N. (2014). Development and validation of a rubric for diagnosing students' experimental design knowledge and difficulties. *CBE—Life Sciences Education, 13*(2), 265–284.

Dasgupta, A. P., Anderson, T. R., & Pelaez, N. J. (2016). Development of the neuron assessment for measuring biology students' use of experimental design concepts and representations. *CBE—Life Sciences Education, 15*(2), ar10.

Deane, T., Nomme, K., Jeffery, E., Pollock, C., & Birol, G. (2014). Development of the biological experimental design concept inventory (BEDCI). *CBE—Life Sciences Education, 13*(3), 540–551.

Deane, T., Nomme, K., Jeffery, E., Pollock, C., & Birol, G. (2016). Development of the statistical reasoning in biology concept inventory (SRBCI). *CBE—Life Sciences Education, 15*(1), ar5.

Duncan, R. G., Chinn, C. A., & Barzilai, S. (2018). Grasp of evidence: Problematizing and expanding the next generation science standards' conceptualization of evidence. *Journal of Research in Science Teaching, 55*(7), 907–937.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education, 32*(1), 268–291.

Erduran, S., Ozdem, Y., & Park, J. Y. (2015). Research trends on argumentation in science education: A journal content analysis from 1998–2014. *International Journal of STEM Education, 2*, 5. https://doi.org/10.1186/s40594-015-0020-1

Fisher, G. R., Olimpo, J. T., McCabe, T. M., & Pevey, R. S. (2018). The Tigriopus CURE – A course-based undergraduate research experience with concomitant supplemental instruction. *Journal of Microbiology & Biology Education, 19*(1), 19–11.

Fry, J. R., & Burr, S. A. (2011). A double-blind atropine trial for active learning of autonomic function. *Advances in Physiology Education, 35*(4), 438–444.

Full, R. J., Dudley, R., Koehl, M. A. R., Libby, T., & Schwab, C. (2015). Interdisciplinary laboratory course facilitating knowledge integration, mutualistic teaming, and original discovery. *Integrative and Comparative Biology, 55*(5), 912–925.

Furtak, E. M., Hardy, I., Beinbrech, C., Shavelson, R. J., & Shemwell, J. T. (2010). A framework for analyzing evidence-based reasoning in science classroom discourse. *Educational Assessment, 15*(3–4), 175–196. https://doi.org/10.1080/10627197.2010.530553

Gardner, G., & Jones, G. (2011). Perceptions and practices: Biology graduate teaching assistants' framing of a controversial socioscientific issue. *International Journal of Science Education, 33*(8), 1031–1054.

Gauthier, A., Jantzen, S., McGill, G., & Jenkinson, J. (2019). Molecular concepts adaptive assessment (MCAA) characterizes undergraduate misconceptions about molecular emergence. *CBE—Life Sciences Education, 18*(1), ar4.

Giere, R. N. (2010). *Scientific perspectivism*. University of Chicago Press.

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a test of scientific literacy skills (TOSLS): Measuring undergraduates' evaluation of scientific information and arguments. *CBE—Life Sciences Education, 11*(4), 364–377.

Hester, S., Buxner, S., Elfring, L., & Nagy, L. (2014). Integrating quantitative thinking into an introductory biology course improves students' mathematical reasoning in biological contexts. *CBE—Life Sciences Education, 13*(1), 54–64.

Hicks, J., Dewey, J., Brandvain, Y., & Schuchardt, A. (2020). Development of the biological variation in experimental design and analysis (BioVEDA) assessment. *PLoS One, 15*(7), e0236098.

Irby, S. M., Pelaez, N. J., & Anderson, T. R. (2018). Anticipated learning outcomes for a biochemistry course-based undergraduate research experience aimed at predicting protein function from structure: Implications for assessment design. *Biochemistry and Molecular Biology Education, 46*(5), 478–492.

Killpack, T. L., & Fulmer, S. M. (2018). Development of a tool to assess interrelated experimental design in introductory biology. *Journal of Microbiology & Biology Education, 19*(3), 19–13.

King, S. L. (2018). *Teaching assistant pedagogical development: Evolution, biological change over time, and concept based instruction*. Doctoral dissertation, The Pennsylvania State University (Order No. 28778767). Available from Dissertations & Theses @ CIC Institutions; ProQuest Dissertations & Theses Global. (2572521567). https://www.proquest.com/dissertations-theses/teaching-assistant-pedagogical-development/docview/2572521567/se-2?accountid=13360

Klein, G. C., & Carney, J. M. (2014). Comprehensive approach to the development of communication and critical thinking: Bookend courses for third-and fourth-year chemistry majors. *Journal of Chemical Education, 91*(10), 1649–1654.

Kowalski, J. R., Hoops, G. C., & Johnson, R. J. (2016). Implementation of a collaborative series of classroom-based undergraduate research experiences spanning chemical biology, biochemistry, and neurobiology. *CBE—Life Sciences Education, 15*(4), ar55.

Krontiris-Litowitz, J. (2013). Using primary literature to teach science literacy to introductory biology students. *Journal of Microbiology & Biology Education, 14*(1), 66–77.

Lansverk, A. L., Lichti, D. A., Blinka, K. W., & Callis-Duehl, K. L. (2020). Comparing the outcomes of "Pre-CURE" compared to inquiry-based introductory biology labs. *Bioscene: Journal of College Biology Teaching, 46*(1), 10–22.

Martinez-Vaz, B. M., & Mickelson, M. M. (2020). In silico phage hunting: Bioinformatics exercises to identify and explore bacteriophage genomes. *Frontiers in Microbiology, 11*, 2294.

Manz, E., Lehrer, R., & Schauble, L. (2020). Rethinking the classroom science investigation. *Journal of Research in Science Teaching, 57*(7), 1148–1174.

McNeill, K. L., & Berland, L. (2017). What is (or should be) scientific evidence use in K-12 classrooms? *Journal of Research in Science Teaching, 54*(5), 672–689.

National Research Council. (2001). *Knowing what students know: The science and design of educational assessment*. The National Academies Press. https://doi.org/10.17226/10019

National Research Council US. (2005). In S. R. Singer, M. L. Hilton, & H. A. Schweingruber (Eds.), *America's lab report: Investigations in high school science*. National Academies Press.

National Research Council. (2014). In J. W. Pellegrino, M. R. Wilson, J. A. Koenig, & A. S. Beatty (Eds.), *Developing assessments for the next generation science standards*. National Academies Press.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. The National Academies Press.

Ott, L. E., & Carson, S. (2014). Immunological tools: Engaging students in the use and analysis of flow cytometry and enzyme-linked immunosorbent assay (ELISA). *Biochemistry and Molecular Biology Education, 42*(5), 382–397.

Pelaez, N., Anderson, T. R., Gardner, S. M., Yin, Y., Abraham, J. K., Bartlett, E. L., Gormally, C., Hill, J. P., Hoover, M., Hurney, C. A., Long, T. M., Newman, D. L., Sirum, K., & Stevens, M. T. (2017). *The basic competencies of biological experimentation: Concept-skill statements* (PIBERG instructional innovation material paper 4). https://docs.lib.purdue.edu/pibergiim/4/. Accessed 8 Nov 2021.

Pellegrino, J. W. (2012). Assessment of science learning: Living in interesting times. *Journal of Research in Science Teaching, 49*, 831–841. https://doi.org/10.1002/tea.21032

Phelps, R. P. (2011). *Standards for educational & psychological testing*. American Psychological Association.

Resendes, K. K. (2015). Using HeLa cell stress response to introduce first year students to the scientific method, laboratory techniques, primary literature, and scientific writing. *Biochemistry and Molecular Biology Education, 43*(2), 110–120.

Reynders, G., Lantz, J., Ruder, S. M., Stanford, C. L., & Cole, R. S. (2020). Rubrics to assess critical thinking and information processing in undergraduate STEM courses. *International Journal of STEM Education, 7*(1), 1–15.

Richards, K. A. R., & Hemphill, M. A. (2018). A practical guide to collaborative qualitative data analysis. *Journal of Teaching in Physical Education, 37*(2), 225–231.

Robertson, A. L., & Phillips, A. R. (2008). Integrating PCR theory and bioinformatics into a research-oriented primer design exercise. *CBE—Life Sciences Education, 7*(1), 89–95.

Rybarczyk, B. J., Walton, K. L., & Grillo, W. H. (2014). The development and implementation of an instrument to assess students' data analysis skills in molecular biology. *Journal of Microbiology & Biology Education, 15*(2), 259–267.

Samarapungavan, A. (2018). Chapter 4: Construing scientific evidence: The role of disciplinary knowledge in reasoning with and about evidence in scientific practice. In F. Fischer,

K. Englemann, J. Osborne, & C. A. Chinn (Eds.), *Scientific reasoning and argumentation: The roles of domain-specific and domain-general knowledge*. Routledge Taylor & Francis Group. https://doi.org/10.4324/9780203731826

Sandoval, W. A., Brian, J., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education, 88*(3), 345–372.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction, 23*(1), 23–55.

Schen, M. S. (2007). *Scientific reasoning skills development in the introductory biology courses for undergraduates*. Doctoral dissertation, The Ohio State University.

Seixas Mello, P., Cotta Natale, C., Marzin-Janvier, P., Vieira, L. Q., & Manzoni-de-Almeida, D. (2021). Inquiry-based learning in immunology: Analysis of scientific argument construction by undergraduate students in biological science and health care classes. *Journal of Biological Education*, 1–15. https://doi.org/10.1080/00219266.2021.1877778

Shi, J., Power, J., & Klymkowsky, M. (2011). Revealing student thinking about experimental design and the roles of control experiments. *International Journal for the Scholarship of Teaching and Learning, 5*(2). https://digitalcommons.georgiasouthern.edu/ij-sotl/vol5/iss2/8/

Simmons, A. D., Larios-Sanz, M., Amin, S., & Rosell, R. C. (2014). Using mini-reports to teach scientific writing to biology students. *The American Biology Teacher, 76*(8), 551–555.

Sirum, K., & Humburg, J. (2011). The experimental design ability test (EDAT). *Bioscene: Journal of College Biology Teaching, 37*(1), 8–16.

Sorte, C. J., Aguilar-Roca, N. M., Henry, A. K., & Pratt, J. D. (2020). A hierarchical mentoring program increases confidence and effectiveness in data analysis and interpretation for undergraduate biology students. *CBE—Life Sciences Education, 19*(3), ar23.

Spence, P. L., Phillips, R. S., McAllister, A. R., White, S. L., & Hollowell, G. P. (2020). Student-scientist curriculum: Integrating inquiry-based research experiences and professional development activities into an introductory biology laboratory course. *Journal of Microbiology & Biology Education, 21*(3), 30.

St. Onge, S. (2007). *Long term urban wildlife research potential through course-based undergraduate research experiences*. Masters dissertation, University of Colorado at Denver (Order No. 10982264). Available from ProQuest Dissertations & Theses Global. (2150116451). https://www.proquest.com/dissertations-theses/long-term-urban-wildlife-research-potential/docview/2150116451/se-2?accountid=13360

Stransky, J., Bassett, L., Bodnar, C., Anastasio, D., Burkey, D., & Cooper, M. (2021). A retrospective analysis on the impacts of an immersive digital environment on chemical engineering students' moral reasoning. *Education for Chemical Engineers, 35*, 22–28.

Terry, D. R. (2007). *Using the case study teaching method to promote college students' critical thinking skills.* Doctoral dissertation, State University of New York at Buffalo (Order No. 3277766). Available from ProQuest Dissertations & Theses Global. (304778960). https://www.proquest.com/dissertations-theses/using-case-study-teaching-method-promote-college/docview/304778960/se-2?accountid=13360

Thanukos, A., Scotchmoor, J. G., Caldwell, R., & Lindberg, D. R. (2010). The real process of science: Understanding science at https://evolution.berkeley.edu/teach-evolution/conceptual-framework/ and http://www.understandingscience.org and Science 101: Building the foundations for real understanding. *Science*, *330*(6012), 1764–1765. https://doi.org/10.1126/science.118699

Timmerman, B. E. C., Strickland, D. C., Johnson, R. L., & Payne, J. R. (2011). Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education, 36*(5), 509–547.

Toulmin, S. E. (1958). *The uses of argument*. Cambridge University Press.

Tytler, R., & Peterson, S. (2005). A longitudinal study of children's developing knowledge and reasoning in science. *Research in Science Education, 35*, 63–98.

Volz, T., & Saterbak, A. (2009). Students' strengths and weaknesses in evaluating technical arguments as revealed through implementing Calibrated Peer Review™ in a bioengineering laboratory. *Across the Disciplines, 6*(1).

Ward, J. R., Clarke, H. D., & Horton, J. L. (2014). Effects of a research-infused botanical curriculum on undergraduates' content knowledge, STEM competencies, and attitudes toward plant sciences. *CBE—Life Sciences Education, 13*(3), 387–396.

Weaver, K. F., Morales, V., Nelson, M., Weaver, P. F., Toledo, A., & Godde, K. (2016). The benefits of peer review and a multi-semester capstone writing series on inquiry and analysis skills in an undergraduate thesis. *CBE—Life Sciences Education, 15*(4), ar51.

Younkin, K., & Romano, C. (2018). Student-centered microbioassay laboratory activity utilizing bioluminescent bacteria. *Journal of Microbiology & Biology Education, 19*(2), 19–12.