Resource Letter RBAI-1: Research-Based Assessment Instruments in Physics and Astronomy

Adrian Madsen and Sarah B. McKaganEleanor C. Sayre

Citation: American Journal of Physics **85**, 245 (2017); doi: 10.1119/1.4977416
View online: http://dx.doi.org/10.1119/1.4977416
View Table of Contents: http://aapt.scitation.org/toc/ajp/85/4
Published by the American Association of Physics Teachers

RESOURCE LETTER

Resource Letters are guides for college and university physicists, astronomers, and other scientists to literature, websites, and other teaching aids. Each Resource Letter focuses on a particular topic and is intended to help teachers improve course content in a specific field of physics or to introduce nonspecialists to this field. The Resource Letters Editorial Board meets annually to choose topics for which Resource Letters will be commissioned during the ensuing year. Items in the Resource Letter below are labeled with the letter E to indicate elementary level or material of general interest to persons seeking to become informed in the field, the letter I to indicate intermediate level or somewhat specialized material, or the letter A to indicate advanced or specialized material. No Resource Letter is meant to be exhaustive and complete; in time there may be more than one Resource Letter on a given subject. A complete list by field of all Resource Letters published to date is at the website <http://ajp.dickinson.edu/Readers/resLetters.html>. Suggestions for future Resource Letters, including those of high pedagogical value, are welcome and should be sent to Professor Mario Belloni, Editor, AJP Resource Letters, Davidson College, Department of Physics, Box 6910, Davidson, NC 28035; e-mail: mabelloni@davidson.edu.

Resource Letter RBAI-1: Research-Based Assessment Instruments in Physics and Astronomy

Adrian Madsen and Sarah B. McKagan
American Association of Physics Teachers, College Park, Maryland 20740

Eleanor C. Sayre
Department of Physics, Kansas State University, Manhattan, Kansas 66506

(Received 1 May 2016; accepted 30 January 2017)

This resource letter provides a guide to Research-Based Assessment Instruments (RBAIs) of physics and astronomy content. These are standardized assessments that were rigorously developed and revised using student ideas and interviews, expert input, and statistical analyses. RBAIs have had a major impact on physics and astronomy education reform by providing a universal and convincing measure of student understanding that instructors can use to assess and improve the effectiveness of their teaching. In this resource letter, we present an overview of all content RBAIs in physics and astronomy by topic, research validation, instructional level, format, and themes, to help faculty find the best assessment for their course. More details about each RBAI available in physics and astronomy are available at PhysPort: physport.org/assessments.
[http://dx.doi.org/10.1119/1.4977416]

I. INTRODUCTION

The physics and astronomy education research communities have produced 40+ Research-Based Assessment Instruments (RBAIs) of physics and astronomy content, which evaluate the effectiveness of different teaching methods. We define a research-based assessment as an assessment that is developed based on research into student thinking for use by the wider physics and astronomy education community to provide a standardized assessment of teaching and learning. Conceptual RBAIs have had a major impact on physics education reform by providing a universal and convincing measure of student understanding that instructors can use to assess and improve the effectiveness of their teaching. Studies using these instruments consistently show that research-based teaching methods lead to dramatic improvements in students' conceptual understanding of physics.[2,3] These instruments are already being used on a very large scale: The Force Concept Inventory[4] (FCI), a test of basic concepts of forces and acceleration, has been given to thousands of students throughout the world; the use of similar instruments in nearly every subject area of physics is becoming increasingly widespread. These kinds of conceptual assessments are especially powerful because, especially with the FCI, physics instructors' first impression is often that it is too trivial, but then they are surprised when their students score poorly, although students often think that they did well. According to a recent survey of faculty who are about to participate in the Workshop for New Faculty in Physics and Astronomy, nearly half have heard of the FCI, and nearly a quarter have used it in their classrooms.[5] The use of these instruments has the potential to transform teaching practice by informing instructors about their teaching efficacy so that they can improve it. For further discussion of the affordances and constraints of using RBAIs, see our article about "best practices for using concept inventories".[6]

Our previous research shows that many physics faculty are aware of the existence of RBAIs for introductory physics, but want to know more about RBAIs for a wider range of topics, including upper-level physics, and which assessments are available and how to use them.[7] This resource letter addresses these needs of physics faculty by presenting an overview of content RBAIs by topic, research validation, instructional level, format, and themes, to help faculty find the best assessment for their course. A second resource letter will discuss the large number of RBAIs that cover non-content topics such as attitudes and beliefs about physics, epistemologies and expectations, the nature of physics, problem solving, self-efficacy, math skills, reasoning skills, and lab skills.

We begin with a general discussion of the process of development and validation of RBAIs (Sec. II), and then discuss specific RBAIs in each of the major content areas in physics and astronomy. These RBAIs cover a diverse set of topics including mechanics (Sec. III), electricity and magnetism (Sec. IV), quantum mechanics and modern physics (Sec. V), thermodynamics (Sec. VI), waves and optics (Sec. VII), and astronomy (Sec. VIII), at a range of levels from high school to graduate school. The only major physics content area where we are unaware of any RBAI is statistical mechanics.

Most RBAIs are multiple-choice tests with five answer choices and are based on research into students' ideas about a narrow range of introductory-level topics. Some tests have more than five answer choices in order to span the space of student ideas. There are also some assessments of upper-level topics, which are often free-response format, and are based on experts' ideas about a topic, since students' have fewer ideas about these topics coming into the course. There are also a few assessments that use a multiple-response format to capture more of students thinking about the topic, but are still easy to score. There are also RBAIs which cover a wide range of topics, with fewer questions about each. These can give instructors a better sense of what their students learned about many topics, though, since each topic is not probed in depth, there is more uncertainty in the results for each topic.

More details about each RBAI available in physics and astronomy are available at PhysPort:[1] physport.org/assessments, where verified educators can download most RBAIs. You can also see example problems from all of the RBAIs discussed here, recommended time to take each, whether they should be given as pre and post, or post only, as well as information on the research behind the assessment, typical results, translations available, and other resources. Wilcox *et al.*[8] have a more detailed discussion of upper-division RBAIs.

There are specific guidelines for using RBAIs in your class, e.g., maintaining test security, ensuring validity, encouraging student participation, etc. We have written an article, Best practices for administering concept inventories,[6] currently available on the arXiv. We encourage you to read this article if you are new to using RBAIs in your course.

1. PhysPort, Browse Assessments, www.physport.org/assessments. PhysPort is a free website developed by the American Association of Physics Teachers in collaboration with Kansas State University and supported by the National Science Foundation. It was previously called "The PER User's Guide." At PhysPort verified educators can learn about download 80+ research-based assessments in physics and related fields, covering content as well as non-content topics such as attitudes, beliefs and scientific reasoning, for various courses from high school to graduate levels. (E)

2. "Secondary analysis of teaching methods in introductory physics: A 50 k-student study," J. Von Korff, B. Archibeque, A. Gomez, S. B. McKagan, E. C. Sayre, E. W. Schenk, C. Shepherd, and L. Sorell, Am. J. Phys. **84**(12), 969–974 (2016). (E)

3. "Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses," R. R. Hake, Am. J. Phys. **66**(1), 64–74 (1998). (E)

4. "Force concept inventory," D. Hestenes, M. M. Wells, and G. Swackhamer, Phys. Teach. **30**(3), 141–166 (1992). (E)

5. "Promoting instructional change in new faculty: An evaluation of the physics and astronomy new faculty workshop," C. Henderson, Am. J. Phys. **76**(2), 179–187 (2008). (E)

6. "Best practices for administering concept inventories," A. Madsen, S. B. McKagan, and E. C. Sayre, Phys. Teach. (submitted). (E)

7. "Research-based assessment affordances and constraints: Perceptions of physics faculty," A. Madsen, S. B. McKagan, M. S. Martinuk, A. Bell, and E. C. Sayre, Phys. Rev. Phys. Educ. Res. **12**, 010115 (2016). (E)

8. "Development and uses of upper-division conceptual assessments," B. R. Wilcox, M. D. Caballero, C. Baily, S. V. Chasteen, Q. X. Ryan, and S. J. Pollock, Phys. Rev. ST Phys. Educ. Res. **11**, 020115 (2015). (I)

II. DEVELOPMENT AND VALIDATION OF RESEARCH-BASED ASSESSMENTS

Good research-based assessment instruments are different from typical exams in that their creation involves extensive research and development by experts in Physics Education Research (PER) and/or Astronomy Education Research (AER) to ensure that the questions represent concepts that faculty think are important, where the possible responses represent real student thinking and make sense to students, and that students' scores reliably tell us something about their understanding. The typical process of developing a research-based assessment includes the following steps:[9,10]

(1) Gathering students' ideas about a given topic, usually with interviews or open-ended written questions.

(2) Using students' ideas to write multiple-choice conceptual questions where the incorrect responses cover the range of students' most common incorrect ideas using the students' actual wording.

(3) Testing these questions with another group of students. Usually, researchers use interviews where students talk about their thinking for each question.

(4) Testing these questions with experts in the discipline to ensure that they agree on the importance of the questions and the correctness of the answers.

(5) Revising questions based on feedback from students and experts.

(6) Administering assessment to large numbers of students. Checking the reproducibility of results across courses and institutions. Checking the distributions of answers. Using various statistical methods to ensure the reliability of the assessment.

(7) Revising again.

Beichner[11] described a similar process for developing RBAIs that might also be of interest to a new RBAI developer. This rigorous development process produces valid and reliable assessments that can be used to compare instruction across classes and institutions. Based on the steps to developing a good research-based assessment, we have created list of seven categories of research validation (Table I).

Table I. Research validation categories.

Questions based on research into student thinking
Studied with student interviews
Studied with expert review
Appropriate use of statistical analysis
Administered at multiple institutions
Research published by someone other than developers
At least one peer-reviewed publication

Table II. Determination of the level of research validation for an assessment.

# Categories	Research validation level
All 7	Gold
5–6	Silver
3–4	Bronze
1–2	Research-based

Each of these categories says different things about the research validation behind the instrument. "Studied with student interviews" and "questions based on research into student thinking" are two different ways of connecting test questions with students' ideas. "Studied with expert review" ensures that the questions are relevant to physics educators. "Appropriate use of statistical analysis" compares students' performance on the questions in a robust way. "Administered at multiple institutions" ensures that the RBAI is applicable to more than one institution. "Research published by someone other than developers" and "at least one peer-reviewed publication" are two different ways of measuring community buy-in about the research behind the RBAI. Different members of the research community value these different methods in different ways. Several articles discuss the affordances and constraints of these categories in more depth.[9,10,12]

We determine the level of research validation for an assessment based on how many of the research validation categories apply to the RBAI (Table II). RBAIs will have a gold level validation when they have been rigorously developed and recognized by a wider research community. Silver-level RBAIs are also well-validated, but are missing 1 and 2 levels of research validation. In many cases, silver RBAIs have been validated by the developers but not the larger community, often because these assessments are new. Bronze-level assessments are those where developers have done some validation but are missing pieces. Finally, research-based validation means that an assessment is likely still in the early stages. While the research validation category given for each assessment in this paper is informative, you may be interested in knowing exactly what levels of research validation were completed for a particular assessment. To do this, go to the research tab on the PhysPort assessment[1] you are interested in. There you will find a list of the validation categories indicating which have been completed, and a short description of the research done for that assessment (Fig. 1).

9. "Development and validation of instruments to measure learning of expert-like thinking," W. K. Adams and C. E. Wieman, Int. J. Sci. Educ. **33**(9), 1289–1312 (2011). (E)

10. "An introduction to classical test theory as applied to conceptual multiple-choice tests," P. V. Engelhardt, Getting Started PER **2**(1), 1–40 (2009), <http://www.compadre.org/per/items/detail.cfm?ID=8807>. (E)

11. "Testing student interpretation of kinematics graphs," R. J. Beichner, Am. J. Phys. **62**(8), 750–762 (1994). (E)

12. "Are they all created equal? A comparison of different concept inventory development methodologies," R. S. Lindell, E. Peak, and T. M. Foster, AIP Conf. Proc. **883**, 14–17 (2007). (E)

III. MECHANICS ASSESSMENTS

The topic of mechanics has the largest number of RBAIs, because so many students take introductory mechanics courses at the university level and the content is very standardized. These mechanics RBAIs cover kinematics and forces, energy, rotation, and density (Table III). Because of the wide variety of topics taught in introductory mechanics courses, there is no assessment where all course content is covered. Instead these assessments have a more narrow range of topics, so that you can probe your students' understanding of each sub-topic in mechanics more thoroughly. There is also one mechanics RBAI for intermediate and upper-division mechanics courses (Table IV).

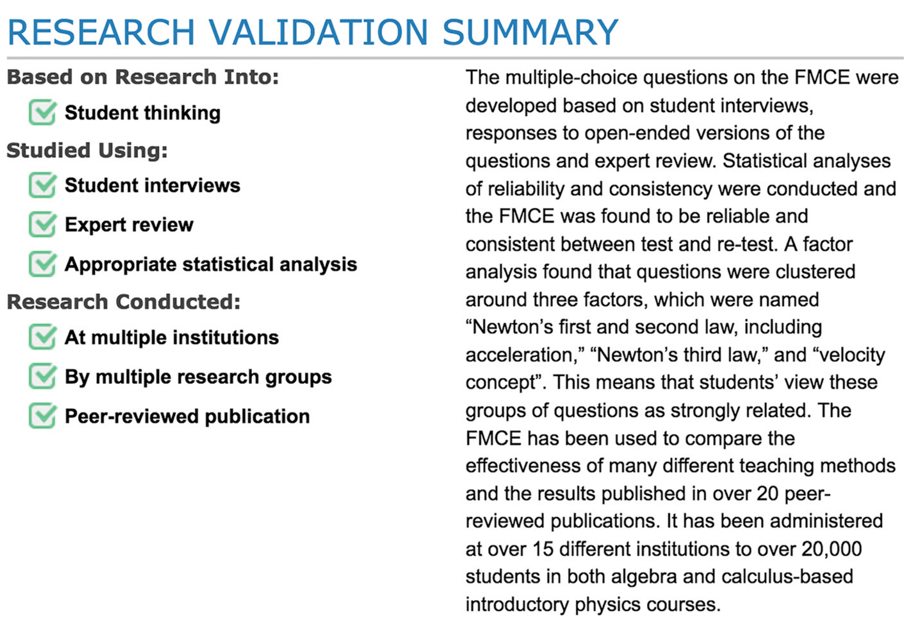

Fig. 1. Examples of research validation summary for the FMCE from PhysPort.

Table III. Introductory mechanics assessments.

Title	Content	Intended population	Research validation	Purpose
Kinematics and forces				
Force Concept Inventory (FCI)	Kinematics, forces: 1D and 2D	Intro college, high school	Gold	To assess students' understanding of the most basic concepts in Newtonian physics using everyday language and common-sense distractors
Force and Motion Conceptual Evaluation (FMCE)	Kinematics, forces: 1D	Intro college, high school	Gold	To assess students' understanding of Newtonian mechanics
Mechanics Baseline Test (MBT)	Kinematics, forces, energy, momentum	Intro college, high school	Bronze	To assess more formal dimensions of basic Newtonian physics
Inventory of Basic Conceptions-Mechanics (IBCM)	Kinematics, forces	Intro college	Silver	To assess the basic threshold of meaningful understanding of Newtonian theory
Test of Understanding of Graphs: Kinematics (TUG-K and TUG-K2)	Kinematics graphs	Intro college, high school (use TUG-K2)	Gold	To assess students' ability to interpret kinematics graphs
Force, Velocity and Acceleration Test (FVA)	Force, velocity, acceleration	Intro college	Bronze	To assess students' understanding of the relationships between force, velocity, and acceleration
Energy				
Energy and Momentum Conceptual Survey (EMCS)	Energy, momentum	Intro college	Gold	To assess conceptual understanding of energy and momentum for standard introductory mechanics courses
Energy Concept Assessment (ECA)	Energy principle, forms of energy, work and heat, absorption & emission spectra, specification of appropriate systems	Intro college	Silver	To assess conceptual understanding of students in the matter & interactions (M&I) Mechanics courses
Rotation				
Rotational and Rolling Motion Conceptual Survey (RRMCS)	Rotational motion	Intro college	Silver	To assess students' understanding of rotational and rolling motion concepts typically covered in a standard introductory physics course
Rotational Kinematics Inventory (RKI)	Rotational kinematics	Intro college	Bronze	To assess students' understanding of angular velocity and angular acceleration of a particle in standard introductory physics contexts
Density				
Density Survey (DS)	Mass, volume, density	Intro college, high school	Bronze	To assess students' understanding of density

A. Kinematics and forces

1. Overview of kinematics and forces assessments

There are six RBAIs which cover kinematics and forces: The Force Concept Inventory[4] (FCI), Force and Motion Conceptual Evaluation[13] (FMCE), Test of Understanding of Graphs in Kinematics[11] (TUG-K), Mechanics Baseline Test[14] (MBT), Force, Velocity, and Acceleration (FVA) test,[15] and Inventory of Basic Conceptions in Mechanics[16] (IBCM). Research and development of kinematics and forces RBAIs has been occurring since the early 1990s, with the FCI[4] being the first published RBAI in physics. The kinematics and forces RBAIs are all used in introductory classes at the university level, and some are also appropriate for high school students.

Table IV. Intermediate level mechanics assessment.

Title	Content	Intended Population	Research Validation	Purpose
Colorado Mechanics/Math Methods Instrument (CCMI)	Ordinary differential equations, Taylor series, potential energy, simple harmonic motion, Newton's laws	Intermediate, upper-level	Silver	To gauge student learning in your first semester classical mechanics course in a way that traditional exams do not allow and compare your students' skills to other.

The most commonly used test of forces and motion is the Force Concept Inventory[4] (FCI). This is a multiple-choice pre/post conceptual assessment about the most basic concepts of force and motion appropriate for introductory university level physics courses and high school courses. The FCI was the first RBAI in physics that presented answer choices consisting of Newtonian concepts and common-sense alternatives that were based on research into student thinking. Understanding which of these common-sense alternatives students choose is just as important as looking at the number of correct answers, as this information helps instructors learn how to improve their teaching. About half of the questions on the FCI come from an earlier test called the Mechanics Diagnostic Test[17] (MDT). Questions on the MDT were developed using students' ideas from open-ended responses.

There are several variations of the FCI: The Gender FCI[18,19] (or Everyday FCI) uses the same questions and answer choices as the original FCI, but changes the contexts to make them more "everyday" or "feminine." The Animated FCI[20] takes the original FCI questions and animates the diagrams, so it is given on a computer. The Representational Variant of the FCI[21] (R-FCI) takes nine questions from the original FCI and redesigns them using various representations (such as motion maps, vectorial and graphical representations). The Familiar Context FCI[22] presents the original FCI questions with everyday contexts, e.g., falling fruit instead of stones or colliding shopping carts instead of cars. The Simplified FCI[22] was adapted from the original FCI and made simpler for ninth grade physics. The Half-length FCI[23] (HFCI) uses the questions from the FCI (v95) and creates two equivalent tests (HFCI1 and HFCI2) that are about half of the length of the original FCI (14 versus 30 questions), but have virtually identical total scores. Several high-scoring FCI questions were removed from the Half-length versions, so the scores of on the HFCI1 and HFCI2 are about 5% lower than those for the FCI.

The Force Motion Conceptual Evaluation[13] (FMCE) is another multiple-choice pre/post conceptual assessment of forces and motion appropriate for introductory university physics courses. The questions on the FMCE are also based on research into student thinking. The FMCE has been used to show that traditional instruction does little to change students' conceptual understanding of forces and motion.

Many of the questions on the FMCE have a more complex question format, which includes a description of the problem context, a list of answer choices (often more than five), and then several questions about that problem situation. In order to give the FMCE in class, a special Scantron form[24] with room for ten answer choices is needed. The FMCE questions were developed based on student interviews, responses to open-ended versions of the questions and expert review.

Both the FCI and FMCE cover forces and motion, but they have different emphases. The FCI covers more topics than the FMCE, but the FMCE has more questions about each topic to more thoroughly assess students' understanding of each topic. Both tests assess one-dimensional kinematics and Newton's laws. The FCI also includes questions on two-dimensional motion with constant acceleration (parabolic motion), impulsive forces, vector sums, cancellation of forces, and identification of forces.[25] The FMCE includes questions about graphs of motion, whereas the FCI does not. FCI questions 15 and 16 present the same situation as FMCE questions 35–38. FCI question 28 is nearly identical to FMCE question 39. The questions on the FCI each have five answer choices, whereas some questions on the FMCE have more than five. Both tests have a strong research base. There is a strong correlation between FCI and FMCE scores.[25] As both of these tests are widely used, there is a large corpus of comparison data (FCI results have been published for over 50,000 students,[2] while FMCE results have been published for over 10,000 students[2]), which can help you understand how your students' scores compare to others. You can find a list of articles with FCI comparison data on the research tab of the FCI assessment page[26] on PhysPort.

The Mechanics Baseline Test[14] (MBT) is another multiple-choice conceptual pre/post assessment for introductory college mechanics courses. The MBT assesses more formal dimensions of basic Newtonian physics with some conceptual questions and some simple calculational questions. The MBT questions are based on research into student thinking. Some of the questions come from Advanced Placement (AP) exams.

The MBT is meant to be used alongside the FCI to get a well-rounded picture of students' understanding. The FCI questions can be answered with no previous physics training, whereas the MBT uses more formal language and includes graphical representations of motion and calculational problems that could not be answered without formal physics training. The MBT not only covers kinematics and forces, like the FCI, but also includes questions on energy and momentum, which are not covered in the FCI. The MBT includes just a few questions on Newton's first and third laws, since these are well covered in the FCI. The answer choices on the MBT include typical student mistakes but not common-sense alternatives like the FCI. The FCI has a stronger research base than the MBT. There is a strong correlation (0.68) between the FCI and the MBT for a group of university students.[14]

The pre/post multiple-choice questions on the Inventory of Basic Conceptions in Mechanics[16] (IBCM) also assess introductory students' conceptual understanding of Newton's laws and forces. The IBCM uses questions from the FCI, MBT, and MDT, but makes slight changes to the wording and answer choices. The IBCM was developed in Lebanon. Since the IBCM takes questions from the FCI and MBT, it is very similar to both of these tests. The IBCM concentrates on Newtonian theory with only two basic models: the free particle and uniformly accelerated motion. It does not include centripetal and centrifugal forces. There are no peer-reviewed publications presenting IBCM results.

The Force, Velocity, and Acceleration (FVA) test[15] is a pre/post multiple-choice conceptual assessment that probes students' understanding of the relationships between force, velocity, and acceleration. Each question presents a scenario with information about either the force, velocity, or acceleration vectors and then asks students about what this means for one of the other vectors. The FVA test provides a coherent picture of student understanding of the relationships between these three by probing six possible conditional relations between them. The FVA test questions were developed using students' responses to open-ended questions and revised using student interviews. The relationships between force, velocity, and acceleration on the FVA test are similar to those relationships probed in several questions on the FCI (questions 4, 7, and 9) and FMCE (questions 1, 3, and 12). The FVA is relatively newer than the FCI and FMCE, so there is not as much comparison data available. Also, the FVA test has been primarily used at the developers' institutions, so it has a lower level of research validation than the FMCE and FCI.

The pre/post multiple-choice questions on the Test of Understanding of Graphs in Kinematics[11] (TUG-K) focus on introductory college and high school students' conceptual understanding of position, velocity, and acceleration versus time graphs. Questions ask students to find displacement, velocity, or acceleration from a given graph or select a graph corresponding to the one given or a textual description. The TUG-K has been validated for high school students, but the TUG-K2 variant was written specifically for high school students. The TUG-K questions were based on the objectives that came from banks of test questions, introductory textbooks, and informal interviews with instructors. Multiple-choice options were written based on the previously studied student difficulties with kinematics graphs. The TUG-K is similar in content and format to the FMCE, which contains 17 out of 47 questions about graphs of motion, including graphs of force versus time, velocity versus time, and acceleration versus time. Both the FMCE and TUGK have a strong research validation.

13. "Assessing student learning of Newton's laws: The force and motion conceptual evaluation and the evaluation of active learning laboratory and lecture curricula," R. K. Thornton and D. R. Sokoloff, Am. J. Phys. **66**(4), 338–352 (1998). (E)

14. "A mechanics baseline test," D. Hestenes and M. Wells, Phys. Teach. **30**(3), 159–166 (1992). (E)

15. "Systematic study of student understanding of the relationships between the directions of force, velocity, and acceleration in one dimension," R. Rosenblatt and A. F. Heckler, Phys. Rev. ST Phys. Educ. Res. **7**(2), 020112 (2011). (I)

16. "Evaluation of the impact of the new physics curriculum on the conceptual profiles of secondary students," I. A. Halloun, <http://www.halloun.net/wp-content/uploads/2016/10/LU-Summative-Report-10-07.pdf>. Beirut, Lebanon (2007). (E)

17. "The initial knowledge state of college physics students," I. A. Halloun and D. Hestenes, Am. J. Phys. **53**(11), 1043–1055 (1985). (E)

18. "Gender differences in student responses to physics conceptual questions based on question context," L. McCullough, in *ASQ Advancing the STEM Agenda in Education, the Workplace and Society*, Stout, WI (2011), pp. 1–10, <http://asq.org/edu/2011/06/continuous-improvement/gender-differences-instudent-responses-to-physics-conceptual-questions-based-on-questioncontent.html?shl=105037>. (E)

19. "Differences in male/female response patterns on alternative-format versions of the force concept inventory," L. McCullough and D. Meltzer, in *Physics Education Research Conference 2001*, Rochester, NY (2001), pp. 103–106, <http://www.compadre.org/per/items/detail.cfm?ID=4324>. (E)

20. "Impact of animation on assessment of conceptual understanding in physics," M. H. Dancy and R. Beichner, Phys. Rev. Spec. Top. - Phys. Educ. Res. **2**, 10104 (2006). (E)

21. "Force concept inventory-based multiple-choice test for investigating students' representational consistency," P. Nieminen, A. Savinainen, and J. Viiri, Phys. Rev. ST Phys. Educ. Res. **6**(2), 020109 (2010). (E)

22. "Can assessment of student conceptions of force be enhanced through linguistic simplification? A Rasch model common person equating of the FCI and the SFCI," S. E. Osborn Popp and J. C. Jackson, in *Annual Meeting of the American Educational Research Association*, San Diego, CA (2009), pp. 1–11, <http://www.compadre.org/per/items/detail.cfm?ID=14025>. (I)

23. "Dividing the force concept inventory into two equivalent half-length tests," J. Han, L. Bao, L. Chen, T. Cai, Y. Pi, S. Zhou, Y. Tu, and K. Koenig, Phys. Rev. ST Phys. Educ. Res. **11**(1), 010112 (2015). (I)

24. "Scantron-forms," http://www.scantron.com/scanners-forms/forms/all-forms

25. "Comparing the force and motion conceptual evaluation and the force concept inventory," R. K. Thornton, D. Kuhl, K. Cummings, and J. Marx, Phys. Rev. Spec. Top. - Phys. Educ. Res. **5**, 010105 (2009). (I)

26. "Force concept inventory," www.physport.org/assessments/FCI

2. Recommendations for choosing a kinematics and forces assessment

Use the FCI if you want a broad understanding of what your students understand about kinematics and Newton's laws, and lots of comparison data. Use the FMCE if you want a more thorough understanding of what your students understand about kinematics and Newton's laws in one-dimension. Use the MBT in conjunction with the FCI to assess more formal parts of your course. Use the FVA if you want to know about your students' understanding of the relationships between force, velocity, and acceleration vectors. Use the TUG-K if you want to thoroughly assess your students' understanding of motion graphs.

B. Energy

1. Overview of energy assessments

There are two RBAIs that cover energy: the Energy and Momentum Conceptual Survey[27] (EMCS) and the Energy Concept Assessment[28] (ECA). Research and development of energy RBAIs has been occurring since the early 2000s to develop these pre/post multiple-choice assessments for introductory classes at the university level.

The Energy and Momentum Conceptual Survey[27] (EMCS) was designed for use in standard first-semester introductory physics courses. It emphasizes energy and momentum in common contexts that your students are likely to have seen in their courses, e.g., carts on tracks, cart filling with rain, bouncing balls, etc. The multiple-choice questions on the EMCS were developed by planning the content and complexity to be tested, getting expert feedback then writing questions. Student responses to open-ended versions of the questions were collected and these responses along with findings from student interviews were used to create the multiple-choice options.

The Energy Concept Assessment[28] (ECA) was designed specifically to assess conceptual understanding of students in the Matter & Interactions (M&I) mechanics course.[29] This is a first-semester introductory physics course with a radical change in content and emphasis, focusing on the power of fundamental principles, on both the macroscopic and the microscopic levels. Because of this, only about half of the questions on the ECA align well with the topics in a standard introductory course. The other half of the questions are not emphasized or covered in a standard course, for example, relativistic energy including rest mass, quantized energy levels, and photon emission and absorption.

27. "Multiple-choice test of energy and momentum concepts," C. Singh and D. Rosengrant, Am. J. Phys. **71**(6), 607–617 (2003). (E)

28. "Designing an energy assessment to evaluate student understanding of energy topics," L. Ding, Ph.D. dissertation, North Carolina State University (2007), <https://repository.lib.ncsu.edu/handle/1840.16/4050>. (I)

29. "Matter & interactions," R. Chabay and B. Sherwood, in *Reviews in PER Vol. 1: Research-Based Reform of University Physics*, edited by E. F. Redish and P. Cooney (American Association of Physics Teachers, College Park, MD, 2007). (E)

2. Recommendation for choosing an energy assessment

The ECA contains questions about non-standard introductory course topics (discussed above) while the EMCS contains more standard questions about energy and momentum. Use whichever test more closely matches the content in your course. Both tests were rigorously developed, tested, and found to be reliable.

C. Rotation

1. Overview of rotation assessments

There are two tests about rotation: Rotational and Rolling Motion Conceptual Survey[30] (RRMCS) and the Rotational Kinematics Inventory[31–33] (RKI). Research and development of rotational motion RBAIs has been occurring since the mid-2000s. Both are multiple-choice conceptual tests that use some physics formalism, which means that the pre-test scores are likely not meaningful because students do not have enough background knowledge to understand the questions.

The Rotational and Rolling Motion Conceptual Survey[30] assesses students' understanding of rotational kinematics and kinetic energy, moment of inertia, torque, and rolling motion. It is appropriate for introductory college students in both algebra-based and calculus-based courses. The RRMCS questions were developed using student ideas from demonstration-based interviews.

The Rotational Kinematics Inventory has three parts: "Part 1: Particles"[31] assesses students' understanding of angular velocity and acceleration of a particle in various standard contexts (the hands of a clock, orbiting plants, swinging pendulum, etc.); "Part 2: Particle in rectilinear motion"[32] assesses students' understanding of the angular velocity and acceleration of a particle moving along a straight line where the origin is not located on that line; "Part 3: Rigid body about a fixed axis,"[33] assesses students' understanding of the rotational kinematics of rigid bodies like pulleys and Ferris wheels. Some of the RKI questions use vector calculus including the cross product. The RKI has been tested with high school students and upper-division college students. Parts of this assessment would also be appropriate for introductory college students. You can use all three parts of the RKI, or only the parts match the content you cover in your course. The RKI questions were developed by creating a map of content and complexity to be tested and a literature review. The RKI was developed in India.

The RRMCS and RKI both cover rotational motion topics but with different emphases. The RRMCS focuses on rotational motion concepts commonly taught in introductory courses. The RKI covers these standard topics and also includes some non-standard topics, e.g., a particle in rectilinear motion, and higher-level math (vector calculus) that is more difficult than the content tested on the RRMCS. Both have a similar level of research validation.

30. "Student understanding of rotational and rolling motion concepts," L. G. Rimoldini and C. Singh, Phys. Rev. Spec. Top. - Phys. Educ. Res. **1**, 10102 (2005). (E)

31. "An inventory on rotational kinematics of a particle: Unravelling misconceptions and pitfalls in reasoning," K. K. Mashood and V. A. Singh, Eur. J. Phys. **33**(5), 1301–1312 (2012). (E)

32. "Rotational kinematics of a particle in rectilinear motion: Perceptions and pitfalls," K. K. Mashood and V. A. Singh, Am. J. Phys. **80**(8), 720–723 (2012). (I)

33. "Rotational kinematics of a rigid body about a fixed axis: Development and analysis of an inventory," K. K. Mashood and V. A. Singh, Eur. J. Phys. **36**, 45020 (2015). (E)

2. Recommendations for rotation assessments

Use the RRMCS to assess standard topics in calculus- and algebra-based introductory physics courses and compare to others. Use the RKI if the content matches what you teach in your course.

D. Density

The Density Survey[34] (DS) is a pre/post conceptual assessment of basic density concepts meant for high school and introductory college students. Most of the questions are standard multiple-choice questions, but there are two questions that require simple calculations. The questions on the DS were developed based on a survey of the literature and one-on-one interviews with several high school students. One question is from the Third International Mathematics and Science Study, and three are adapted from research on electric charge density. Use the density survey if you want to assess the change in your students' understanding of density before and after covering it in your course.

34. "Student understanding of density: a cross-age investigation," R. E. Yeend, M. E. Loverude, and B. L. Gonzalez, in Proceedings of the 2001 Physical Eduction Resarch Conference (2001), <http://www.compadre.org/per/items/detail.cfm?ID=4313>. (E)

E. Intermediate mechanics

The Colorado Mechanics/Math Methods Instrument[35,36] (CCMI) is an open-ended assessment of topics and skills commonly taught in a first-semester intermediate classical mechanics course, including the ability to visualize a problem, correctly apply problem-solving methods, connect math to physics, and describe the limiting behavior. The CCMI covers both content and mathematical skills though the questions are largely conceptual, including reasoning, explanation, graphing, and sketching. The CCMI does not cover all content in intermediate classical mechanics, but rather a sample of important skills. There is an optional shorter pre-test and longer post-test. Both are graded using rubrics. The CCMI questions were developed based on a set of learning goals produced by faculty and observed student difficulty with these concepts. This is the only RBAI for intermediate classical mechanics.

35. "Assessing student learning in middle-division classical mechanics/math methods," M. D. Caballero and S. J. Pollock, in *Physics Education Research Conference 2013*, Portland, OR (2013), pp. 81–84, <http://www.compadre.org/per/items/detail.cfm?ID=13113>. (I)

36. "Issues and progress in transforming a middle-division classical mechanics/math methods course," S. J. Pollock, R. E. Pepper, and A. D. Marino, AIP Conf. Proc. **1413**, 303–306 (2012), <http://www.compadre.org/per/items/detail.cfm?ID=11872>. (I)

IV. ELECTRICITY AND MAGNETISM ASSESSMENTS

RBAIs on Electrostatics and Magnetism (E&M) for introductory courses have been around since the late 1990s. There are six research-based assessments that cover electrostatics and magnetism. Four of these are for introductory courses: The Brief Electricity and Magnetism Assessment[37,38] (BEMA), the Conceptual Survey of Electricity and Magnetism[38,39] (CSEM), the Diagnostic Exam for Introductory, Undergraduate Electricity and Magnetism[40] (DEEM), and the Electricity and Magnetism Conceptual Assessment[41] (EMCA). There is one assessment specifically about symmetry and Gauss's law: the Symmetry and Gauss's Law Conceptual Evaluation[42] (SGCE). There is one assessment which covers just magnetism concepts: the Magnetism Conceptual Survey[43] (MCS). There is also the Electromagnetics Concept Inventory (EMCI) suite of assessments which includes EMCI-waves, EMCI-fields, and EMCI-waves and fields,[44] which were developed for engineering courses and would not be discussed further here.

For circuits, there are three: the Determining and Interpreting Resistive Electric Circuits Concepts Test[45] (DIRECT), the Electric Circuits Conceptual Evaluation[46] (ECCE), and the Inventory of Basic Conceptions-DC Circuits[47] (IBCDC). The CSEM also contains some questions about circuits, but this is not its main focus.

More recently, RBAIs for upper-level courses have been developed. We discuss three: the Colorado Upper Division Electrostatics Diagnostic-Free Response[48] (CUE-FR), the Colorado Upper Division Electrostatics Diagnostics-Coupled Multiple Response[49,50] (CUE-CMR), and the Colorado Upper-Division Electrodynamics Test[51] (CURrENT).

Introductory E&M RBAIs are summarized in Table V; upper-level ones are in Table VI.

A. Introductory level electricity and magnetism

1. Electrostatics and magnetism

The two most commonly used RBAIs for introductory electricity and magnetism courses are the Brief Electricity and Magnetism Assessment (BEMA) and the Conceptual Survey of Electricity and Magnetism (CSEM). Both are multiple-choice and can be given both a pre- and post-tests to measure student learning, or given as a post-test only, since students do not have much initial knowledge of these topics before instruction and average pre-test scores are usually similar at different institutions.

Table V. Introductory electricity and magnetism assessments.

Title	Content	Intended population	Research validation	Purpose
Introductory electrostatics and magnetism				
Brief Electricity and Magnetism Assessment (BEMA)	Circuits, electrostatics, magnetic fields and forces	Intro college	Gold	To assess students' qualitative understanding of basic concepts in electricity and magnetism.
Conceptual Survey of Electricity and Magnetism (CSEM)	Electrostatics, magnetic fields and forces, Faraday's law	Intro college	Silver	To assess students' knowledge about topics in introductory electricity and magnetism.
Diagnostic exam for introductory undergraduate electricity and magnetism (DEEM)	Electric and magnetic fields and forces, electrostatic potential, Maxwell's equations, induced currents	Intro college	Bronze	To assess students' understanding of basic concepts of electricity and magnetism.
Electricity and Magnetism Conceptual Assessment (EMCA)	Electrostatics, electric fields and forces, circuits, magnetism, induction	Intro college	Bronze	To assess basic concepts in an introductory electromagnetism course, using terms that will feel familiar to students on the pre-test and without overly difficult questions that might discourage students from pursuing physics.
Symmetry and Gauss's law conceptual evaluation (SGCE)	Symmetry, electric field, electric flux	Intro college, upper-level graduate	Bronze	To assess students' ability to identify situations where Gauss's Law is applicable and use it to calculate electric field strength.
Magnetism Conceptual Survey (MCS)	Magnetic fields and forces, Faraday's law	Intro college	Silver	To assess difficulties students have with magnetism concepts.
Introductory circuits				
Determining and Interpreting Resistive Electric Circuits Concepts Test (DIRECT)	DC circuits	Intro college	Gold	To evaluate students' understanding of Direct Current (DC) resistive electric circuits concepts.
Electric Circuits Conceptual Evaluation (ECCE)	DC and AC circuits	Intro college	Bronze	To assess students' understanding of simple circuit concepts.
Inventory of Basic Conceptions-DC Circuits (IBCDC)	DC circuits	Intro college	Silver	To assess basic conceptions of DC circuits.

Table VI. Upper-level electricity and magnetism assessments.

Title	Content	Intended population	Research validation	Purpose
Upper-level electrostatics and magnetism				
Colorado Upper Division Electrostatics Diagnostic-Free Response (CUE-FR)	Electrostatics, magnetostatics, choosing a problem-solving method	Upper-level	Silver	To assess skills that faculty teaching this course value, such as the ability to visualize a problem, correctly apply problem-solving methods, connect math to physics, and describe limiting behavior, through conceptual questions involving reasoning, explanation, graphing, and sketching. Time intensive to score
Colorado Upper Division Electrostatics Diagnostics-Coupled Multiple Response (CUE-CMR)	Electrostatics, magnetostatics, choosing a problem-solving method	Upper-level	Silver	To assess skills that faculty teaching this course value, such as the ability to visualize a problem, correctly apply problem-solving methods, connect math to physics, and describe limiting behavior, through conceptual questions involving reasoning, explanation, graphing, and sketching. Quick to score
Upper-level electrodynamics				
Colorado Upper-Division Electrodynamics Test (CURrENT)	Vector calculus, Maxwell's equations, charge and energy conservation, plane waves, transmission and reflection	Upper-level	Silver	To assess fundamental skills and understanding of core topics from advanced undergraduate electrodynamics

The BEMA[37] covers the main topics discussed in both the traditional calculus-based E&M physics curriculum and the Matter and Interactions[29] curriculum including basic electrostatics, circuits, magnetic fields and forces, and induction. BEMA questions are mostly conceptual, but there are few questions that require simple calculations. The BEMA questions were developed based on student difficulties with relevant concepts.

The CSEM[39] is an assessment of students' knowledge of electricity and magnetism. It aims to assess a range of topics across the standard introductory course content, but without assessing every single topic covered in an introductory course. It is a combination of a test of alternative conceptions and knowledge. It also has a combination of questions about the phenomena of electricity and magnetism and questions about the formalism explaining the phenomena. The questions on the CSEM are based on the questions from two earlier tests, the Conceptual Survey of Electricity (CSE) and the Conceptual Survey for Magnetism (CSM). The questions on the CSE and CSM were developed by a group of college physics professors.

The BEMA and CSEM both cover basic topics covered in introductory electricity and magnetism courses. They share six questions that are identical or nearly identical. The topics covered on the BEMA and CSEM vary somewhat. The CSEM does not cover circuits, whereas the BEMA does (7 out of 31 questions). CSEM questions have only five answer choices, while BEMA questions have up to ten possible choices of answers on some questions. Both have similarly strong research validation. CSEM and BEMA scores were compared for one group of students, and on average both pre- and post-test CSEM scores were higher than BEMA scores by 5%–6%, a statistically significant difference, with a moderate effect size.[38] But the absolute and normalized gains were similar for the BEMA and CSEM, so for this group of students, both instruments measure learning in a similar way.

There are two other electricity and magnetism tests that have not been as commonly used and validated: the Electricity Magnetism Conceptual Assessment (EMCA) and Diagnostic Exam for Introductory, Undergraduate Electricity, and Magnetism (DEEM). The multiple-choice pre/post conceptual questions on the DEEM[40] measure students' understanding of basic concepts of electricity and magnetism including electric and magnetic fields and force, electrostatic potential and potential energy, Maxwell's equations, and induced currents. The questions align well with the topics commonly taught in an introductory E&M course. The multiple-choice questions on the DEEM were developed based on student interviews, expert input, instructional objectives, literature review, and observations of students.

The DEEM is much longer than the CSEM or BEMA (66 questions versus 31 and 32 questions, respectively), so it covers topics much more thoroughly. The DEEM also contains follow-up questions, where students should answer a subsequent question only if they chose a certain answer(s) to a previous question. The DEEM, like the CSEM, does not cover circuits. It also does not cover graphical representations of vector fields, or conductors and insulators. About half the questions on the DEEM ask about the direction of the electric field, magnetic field, velocity, electric potential, or force for different situations.

The EMCA[41] is a multiple-choice assessment of standard second-semester introductory physics concepts including electrostatics, electric fields, circuits, magnetism, and induction. The authors developed the EMCA so that it aligned well with the topics taught in their course and so that it produced similar pre-test scores as the FCI for their student population. The EMCA is easier than the BEMA or CSEM. The authors designed the test this way so that on the pre-test students know the answers to some questions and gain confidence in the course (as opposed to the BEMA and CSEM, which many faculty give only as a post-test because students often score near guessing on the pre-test because they are not familiar with the material), but the post-test can still be used to show mastery at the end of the course.

There is only one assessment specifically about symmetry and Gauss's law: the Symmetry and Gauss's Law Conceptual Evaluation[42] (SGCE) which is designed for students in introductory calculus-based physics, but can also be challenging to upper-level students. The SGCE assesses students' ability to identify situations where Gauss's law is applicable and use it to calculate electric field strength. The SGCE questions are multiple-choice, and primarily conceptual, asking students about when and how to use Gauss's law, but not to explicitly calculate values. The BEMA has one question on Gauss's law, and the CUE-CMR and CUE-FR also ask questions which use Gauss's law and that are aimed at upper-level students.

The Magnetism Conceptual Survey[40] (MCS) was developed to help instructors assess difficulties their students have with magnetism concepts in introductory algebra-based and calculus-based courses. It assesses standard topics in introductory courses up to Faraday's law. The MCS only covers magnetism and not electrostatics, so it follows that it has more questions about magnetism than the BEMA, CSEM, DEEM, or EMCA. The BEMA, CSEM, and MCS all cover charges in magnetic fields and magnetic field from current carrying wires.

37. "Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment," L. Ding, R. Chabay, B. Sherwood, and R. Beichner, Phys. Rev. Spec. Top. - Phys. Educ. Res. **2**, 10105 (2006). (E)

38. "Comparing student learning with multiple research-based conceptual surveys: CSEM and BEMA," S. J. Pollock, AIP Conf. Proc. **1064**, 171–174 (2008), <http://www.compadre.org/per/items/detail.cfm?ID=8109>. (E)

39. "Surveying students' conceptual knowledge of electricity and magnetism," D. P. Maloney, T. L. O'Kuma, C. J. Hieggelke, and A. Van Heuvelen, Am. J. Phys. **69**(7), S12–S23 (2001). (E)

40. "Creation of a diagnostic exam for introductory, undergraduate electricity and magnetism," J. D. Marx, Ph.D. dissertation, Rensselaer Polytechnic Institute (1998), <http://www.compadre.org/Per/items/detail.cfm?ID=3786>. (I)

41. "Electricity and magnetism conceptual assessment" M. W. Mccolgan, R. A. Finn, D. Broder, and G. Hassel (unpublished). (E)

42. "Student understanding of symmetry and Gauss's law of electricity," C. Singh, Am. J. Phys. **74**(10), 923–936 (2006). (E)

43. "Developing a magnetism conceptual survey and assessing gender differences in student understanding of magnetism," J. Li and C. Singh, AIP Conf. Proc. **1413**, 43–46 (2012), <http://www.compadre.org/Per/items/detail.cfm?ID=11808>. (E)

44. "Concepty inventory assessment instruments for electromagnetics education," B. Notaros, in *Proceedings of the IEEE Antennas and Propagation Society International Symposium* (2002), Vol. 1, pp. 684–687, <http://ieeexplore.ieee.org/document/1016436/>. (E)

2. Recommendations for choosing an electricity and magnetism test

When teaching an introductory electricity and magnetism (E&M) course, use either the CSEM, BEMA, or DEEM. If you would like to assess your students' understanding of circuits in addition to other standard E&M topics, use the BEMA or DEEM. The CSEM and BEMA are used more commonly, so if having comparison data is important to you, use one of these tests. (For a list of articles with BEMA or CSEM comparison data, see the research tab on the BEMA[45] or CSEM[46] assessment pages on PhysPort.) If you want to assess your students' understanding of magnetism separately from other introductory E&M topics, use the MCS. Use the SGCE if you are particularly interested in introductory physics students' understanding of Gauss's law, or if you are making a change to your teaching about Gauss's law and want to understand if that change helped your students.

45. "Brief electricity and magnetism assessment," www.physport.org/assessments/BEMA

46. "Conceptual survey of electricity and magnetism," www.physport.org/assessments/CSEM

B. Circuits

1. Overview of circuits assessments

There are three RBAIs of circuits: the Determining and Interpreting Resistive Electric Circuits Concepts Test[47] (DIRECT), the Electric Circuits Conceptual Evaluation[48] (ECCE), and the Inventory of Basic Conceptions-DC Circuits[49] (IBCDC). All three are multiple-choice pre/post assessments for introductory college classes. The CSEM also contains some questions about circuits, but this is not its main focus.

The Determining and Interpreting Resistive Electric Circuits Concepts Test[47] (DIRECT) was developed to evaluate students' understanding of direct current (DC) resistive electric circuits concepts. Most of the questions on the DIRECT apply well to a wide variety of introductory courses, though a couple of questions were designed to assess concepts around microscopic aspects of circuits, in a way that closely aligns with the way these concepts are taught in the Electric and Magnetic Interactions curriculum (part of the Matter and Interactions curriculum). The questions on the DIRECT were developed based on instructional objectives, literature review, and expert input.

The Electric Circuits Conceptual Evaluation[48] (ECCE) assesses students' understanding of both direct and alternating current circuits. About 80% of the questions are about DC circuits and cover standard concepts around current, voltage, resistance, and brightness of bulbs in circuits containing resistors and capacitors. The remaining 20% of the questions are about AC circuits and ask students to match current versus time graphs to different circuit configurations.[48] The multiple-choice questions on the ECCE were developed based on open-ended questions about circuits published in the literature as well as the author's personal experience with teaching circuits topics.

Both the DIRECT and the ECCE not only ask similar conceptual questions about standard introductory circuits topics but also ask a couple of questions about non-standard topics: microscopic aspects of current on the DIRECT and AC circuits on the ECCE. Some of the questions on the ECCE have up to ten answer choices. Also, some of the questions on the ECCE have boxes for students to explain their reasoning. If you grade these short answers, the ECCE could take longer to grade, but many instructors just skip grading these. The DIRECT has a higher level of research validation than the ECCE.

The IBCDC is a multiple-choice conceptual assessment of DC circuits developed in the US and Lebanon.[49] The questions on the IBCDC were developed based on a taxonomy of relevant topics decided by experts in physics. The content on the DIRECT and IBCDC is very similar, though the IBCDC only covers circuit concepts that would be taught in a standard introductory level course. The DIRECT has a stronger research base and more comparison data than the IBCDC. (For a list of articles with DIRECT comparison data, see the research tab on DIRECT[50] assessment page on PhysPort.)

47. "Students' understanding of direct current resistive electrical circuits," P. V. Engelhardt and R. J. Beichner, Am. J. Phys. **72**(1), 98–115 (2004). (E)
48. "Teaching electric circuit concepts using microcomputer-based current/voltage problems," D. R. Sokoloff, *Microcomputer-Based Labs: Educational Research and Standards*, Series F, Computer and Systems Sciences Vol. 156, edited by R. F. Tinker (1996), pp. 129–146, <http://link.springer.com/chapter/10.1007/978-3-642-61189-6_7>. (E)
49. "Inventory of Basic Conceptions - DC Circuits (IBCDC)," I. Halloun, at: https://www.physport.org/assessments/assessment.cfm?I=96&A=IBCDC. See also Ref. 16. (E)
50. "Determining and interpreting resistive electric circuit concepts test," www.physport.org/assessments/DIRECT

2. *Recommendations for choosing a circuits assessment*

Use the DIRECT if you want to assess standard introductory DC circuit concepts because it has a stronger research base and is more commonly used, thus providing you with more comparison data. Use the ECCE if you cover AC circuits and DC circuits. Use the IBCDC if the content matches what you teach in your course more closely.

C. Upper-level electricity and magnetism

1. *Electrostatics and magnetism*

The Colorado Upper Division Electrostatics Diagnostic-Free Response[51] (CUE-FR) contains open-ended, primarily conceptual questions that assess students' understanding of electrostatics topics (15 out of 17 questions) commonly covered in the first half of a standard upper-division electricity and magnetism course. It also contains two questions about magnetostatics. In addition to assessing E&M content, the CUE-FR assesses several key skills such as the ability to choose a problem-solving method and defend that choice, visualize a problem, connect math to physics, and describe the limiting behavior. The CUE-FR has open-ended questions where students show work and explain their reasoning. There is an optional 20-min pre-test consisting of the subset of questions that incoming juniors could be expected to know. If you want to check your students' knowledge coming into the course, and learning gains throughout the course, give the pre-test. The questions on the CUE-FR were developed based on previously established learning goals, expert input, and commonly observed student difficulties.

The Colorado Upper Division Electrostatics Diagnostics-Coupled Multiple Response[52] (CUE-CMR) was developed to cover the same content as the CUE-FR, but is easier to grade. The questions on the CUE-FR and CUE-CMR are almost identical, but the answer format is different.[53] The CUE-CMR is a coupled multiple-response assessment where students can choose multiple-responses to a given question and are awarded partial credit depending on the accuracy and consistency of their answer. Students are first asked to select the correct answer or easiest method to solve a problem, and then select a "reasoning element" that supports their initial answer. Students get full credit for selecting all the correct reasoning elements (and only the correct elements). Students can also receive partial credit. A rubric is used to grade the free-responses to the CUE-FR. Partial credit is also granted here. The CUE-CMR also has similar an optional 20-min pre-test consisting of the subset of questions that incoming juniors could be expected to know. The CUE-FR has 17 questions, while the CUE-CMR has 16 questions (it is missing question 15 from the CUE-FR). On average, students score similarly on the multiple-response version of the test as compared to the free-response version of the test. The CUE-CMR was developed based on the CUE-FR, so it has a slightly lower level of research validation (as research has not yet been conducted using it at other institutions or published by other researchers).

51. "Colorado upper-division electrostatics diagnostic: A conceptual assessment for the junior level," S. V. Chasteen, R. E. Pepper, M. D. Caballero, S. J. Pollock, and K. K. Perkins, Phys. Rev. Spec. Top. - Phys. Educ. Res. **8**, 20108 (2012). (I)
52. "Multiple-choice assessment for upper-division electricity and magnetism," B. R. Wilcox and S. J. Pollock, in *Physics Education Research Conference 2013*, Portland, OR (2013), pp. 365–368, <http://www.compadre.org/per/items/detail.cfm?ID=13154>. (I)
53. "Coupled multiple-response versus free-response conceptual assessment: An example from upper-division physics," B. R. Wilcox and S. J. Pollock, Phys. Rev. Spec. Top. - Phys. Educ. Res. **10**, 20124 (2014). (I)

2. *Recommendations for choosing an electricity and magnetism test*

If you are teaching an upper-division E&M and want an assessment that is easy to grade and compare to others, use the CUE-CMR. Use the CUE-FR if you want a more in-depth look at the details of your students' reasoning.

3. *Electrodynamics*

There is one assessment of upper-level electrodynamics: the Colorado Upper-Division Electrodynamics Test[51] (CURrENT). There is also the Electromagnetics Concept Inventory[44] (EMCI) which includes questions about electrodynamics, but was created for engineering courses, so it will not be discussed further here.

The CURrENT is designed to assess fundamental skills and understanding of core topics in the second semester of junior-level undergraduate electrodynamics covering topics in chapters 7–9 of Griffiths.[55] The CURrENT is free-response in order to assess the ability of upper-level students to generate and justify their own answers. The CURrENT has a conceptual focus, though some mathematical manipulations are required. The CURrENT pre-test contains three questions, while the post-test contains six questions, as students do not have a-priori familiarity with many of the topics before taking the course. The CURrENT is graded with a rubric. The CURrENT questions were developed based on previously established learning goals, expert input, and common student difficulties. Use the CURrENT to assess your students' understanding in second semester of junior-level undergraduate electrodynamics.

54. "Research-based course materials and assessments for upper-division electrodynamics (E&M II)," C. Baily, M. Dubson, and S. J. Pollock, AIP Conf. Proc. **1513**, 54–57 (2013), <http://www.compadre.org/per/items/detail.cfm?ID=12661>. (I)
55. *Introduction to Electrodynamics*, D. J. Griffiths, 3rd ed. (Prentice Hall, Upper Saddle River, NJ, 1999). (I)

Table VII. Modern physics assessments.

Title	Content	Intended population	Research validation	Purpose
Relativity				
Relativity Concept Inventory (RCI)	Special relativity	Intro college	Silver	Measure changes in students' conceptual understanding of special relativity and identify students' misconceptions
Intermediate quantum mechanics				
Quantum Physics Conceptual Survey (QPCS)	Photoelectric effect, wave particle duality, de Broglie wavelength, double slit interference, uncertainty principle	Intro college, intermediate	Silver	Investigate students' understanding of introductory quantum physics concepts
Quantum Mechanics Conceptual Survey (QMCS)	Wave functions, probability, infinite square well, one-dimensional tunneling, wave-particle duality, energy levels, uncertainty principle	Intermediate	Silver	Measure the effectiveness of different teaching methods at improving students' conceptual understanding of quantum mechanics, and to use such measurements to improve their teaching
Quantum Mechanics Concept Inventory (QMCI)	Wave functions, probability, 1D tunnelling	Intermediate, upper-level	Research-based	Assess students' alternative conceptions around 1D potential barriers, tunneling, and probability distributions

V. QUANTUM MECHANICS AND MODERN PHYSICS

There are seven tests covering modern physics and/or quantum mechanics content for sophomore, junior, senior, and graduate level courses. These tests were developed starting in the early 2000s and until very recently. All cover a broad range of topics. These tests are discussed below in groups based on the level of course they are appropriate for. There are two additional graduate quantum mechanics surveys, but these are not research-based and validated, so they will not be discussed further below.[56,57] Intermediate-level tests, such as for Modern Physics courses, are summarized in Table VII; ones for upper-level and graduate courses are in Table VIII.

56. "Graduate quantum mechanics reform," L. D. Carr and S. B. McKagan, Am. J. Phys. **77**(4), 308–319 (2009). (I)

57. "Student understanding of quantum mechanics at the beginning of graduate instruction," C. Singh, Am. J. Phys. **76**(3), 277–287 (2008). (I)

Table VIII. Upper-level quantum mechanics assessments.

Title	Content	Intended population	Research validation	Purpose
Upper-level quantum mechanics				
Quantum Mechanics Concept Assessment (QMCA)	Wave functions, probability, infinite square well, 1D tunneling, energy levels, measurement, time dependence	Upper-level	Silver	Assess students' knowledge about main topics of quantum measurement at the junior level. Also compare outcomes of different curricular approaches
Quantum Mechanics Survey (QMS)	Wave functions, probability, infinite square well, 1D tunneling, energy levels, measurement, time dependence	Upper-level and graduate	Silver	Assess students' conceptual understanding of quantum mechanics, specifically their proficiency with the formalism of quantum mechanics in 1D
Quantum Mechanics Formalism and Postulates Survey (QMFPS)	Formalism and postulates of quantum mechanics	Upper-level and graduate	Silver	Assess students' conceptual understanding of the formalism and postulates of quantum mechanics rather than their mathematical skills
Quantum Mechanics Visualization Instrument (QMVI)	Wave functions, probability, infinite square well, 1D tunneling, time dependence, momentum space, 2D potentials, visualization of the relationship between potentials and wave functions	Intermediate, upper-level and graduate	Silver	Probe the development of students' conceptual understanding of core topics in quantum mechanics across the undergraduate curriculum, especially their visualization skills
Quantum Mechanics Assessment Tool (QMAT)	Wave functions, probability, infinite square well, 1D tunneling, energy levels, measurement, time dependence	Upper-level	Bronze	Measure student learning of the quantum mechanics concepts most valued by faculty, assess student learning difficulties, and inform course improvement

A. Modern physics

1. Relativity

The Relativity Concept Inventory[58] (RCI) is the only RBAI that covers special relativity and is for introductory undergraduate courses that cover relevant relativity topics. This is a pre/post conceptual multiple-choice assessment where students are asked to also rate their confidence for each question. Topics covered include time dilation, length contraction, relativity of simultaneity, inertial reference frames, velocity addition, causality, and mass-energy equivalence. The questions were developed based on a list of concepts informed by the learning goals for a relevant course, textbooks, and the research literature. Use the RCI if you want to assess your students' conceptual understanding of special relativity and the effectiveness of your instruction.

58. "Relativity concept inventory: Development, analysis, and results," J. S. Aslanides and C. M. Savage, Phys. Rev. Spec. Top. - Phys. Educ. Res. **9**, 10118 (2013).

2. Intermediate quantum mechanics

There are three tests designed for sophomore-level quantum mechanics: the Quantum Physics Conceptual Survey[59,60] (QPCS), the Quantum Mechanics Conceptual Survey[61] (QMCS), and the Quantum Mechanics Concept Inventory[62] (QMCI). There is one additional quantum assessment, the Quantum Mechanics Visualization Instrument (QMVI), which can be used at multiple levels, including intermediate, upper-level, and graduate quantum, so it will be discussed in the "Upper-level quantum mechanics and beyond" section below.

The Quantum Physics Conceptual Survey[59,60] (QPCS) is a pre/post conceptual assessment that can be used at the introductory level (if you have covered these topics) and in a sophomore-level modern physics course. There are no equations on the QPCS and most questions focus on wave-particle duality and the photoelectric effect (this is the only quantum test which includes the photoelectric effect). Most of the questions are structured in a way that asks the students about what happens when they do a specific experiment. The multiple-choice questions on the QPCS were developed based on topics common across several introductory quantum syllabi, expert opinion, and student ideas that emerged through open-ended questions. It was developed in Thailand and tested in Thailand and Australia.

The Quantum Mechanics Conceptual Survey[61] (QMCS) is a highly conceptual multiple-choice assessment for sophomore-level students. The QMCS can be given as a post-test only at the end of the term in a sophomore-level modern physics course. It can be given as both a pre- and post-test to measure student learning in a junior-level course or higher. Some of the questions on the QMCS probe ideas that students have about quantum mechanics, as uncovered in student interviews. For example, one question asks about electrons moving in sinusoidal paths, because interviews found that this is how many undergraduates think about the motion of an electron. The QMCS does not explicitly include equations, but it does ask students to think about qualitative relationships in equations. The questions on the QMCS were developed based on faculty interviews, a review of textbooks and syllabi, observations of students, and a literature review of known student difficulties. A few of the questions on the QMCS come from other tests (questions 10 and 11 are from the QMVI). Further, the QMCS covers many quantum mechanics topics, but only has 12 questions, so is limited in what it can tell you about what your students learned.

The Quantum Mechanics Concept Inventory[62] (QMCI) is a pre/post multiple-choice assessment which is very conceptual in nature with no equations included and simple language. The question format gives statements from a hypothetical student about a given concept and your students have to pick which one they agree with. It was designed to diagnose students' alternative conceptions about quantum mechanics, so each answer choice is associated with a specific alternative conception. It is meant for sophomore and junior-level students. Questions are based on students' ideas about quantum as documented in the literature. The QMCI was developed in Sweden.

Unlike the QMCS, the questions on the QMCI are about a narrow range of topics, with most questions asking about tunneling through one-dimensional barriers. Similar to the QMCS, the QMCI is very conceptual in nature and only has a few questions (nine for the QMCI), so it is limited in what it tells you about what your students learned.

59. "Probing a deeper understanding of modern physics concepts," T. L. Larkin, P. Meade, and J. Uscinski, in *41st ASEE/IEEE 2011 Frontiers in Education Conference* (2011), pp. S2H–1–S2H–6, <http://ieeexplore.ieee.org/document/6143101/>. (E)

60. "Development and use of a conceptual survey in introductory quantum physics," S. Wuttiprom, M. D. Sharma, I. D. Johnston, R. Chitaree, and C. Soankwan, Int. J. Sci. Educ. **31**(5), 631–654 (2009). (E)

61. "Design and validation of the quantum mechanics conceptual survey," S. B. McKagan, K. K. Perkins, and C. E. Wieman, Phys. Rev. Spec. Top. - Phys. Educ. Res. **6**, 20121 (2010). (E)

62. "Developing a quantum mechanics concept inventory," J. Falk, Ph.D. dissertation, Uppsala University (2004), <http://www.compadre.org/quantum/items/detail.cfm?ID=13413>. (I)

3. Recommendations for choosing an intermediate quantum mechanics assessment

If you are teaching a sophomore-level modern physics course, use the QMCS if you want a broad overview of course topics and the QMCI if you want an in-depth test of one-dimensional potential barriers, tunneling, and probability distribution. Use the QPCS if you want to test photoelectric effect or a more in-depth treatment of wave particle duality. Use QMVI if you want a very detailed look at the relationship between the wave function and shape of potential. The QMVI contains questions from several levels of quantum mechanics, so expect your sophomore-level students to do poorly on most questions.

B. Upper-level quantum mechanics and beyond

There are four tests that are designed to assess students' understanding of quantum at the junior level: The Quantum Mechanics Concept Assessment[63,64] (QMCA), the Quantum Mechanics Assessment Tool[65] (QMAT), the Quantum Mechanics Survey[66] (QMS), and the Quantum Mechanics Formalism and Postulates Survey[67] (QMFPS). The Quantum

Mechanics Visualization Instrument[68] (QMVI) can be used at several levels and will also be discussed in this section.

The Quantum Mechanics Concept Assessment[63,64] (QMCA) is one of the newer quantum mechanics assessments for a first-semester junior-level quantum mechanics course. It assesses students' understanding of five main topics of quantum measurement: the time-independent Schrödinger equation, wave functions, boundary conditions, time evolution, and probability density. The QMCA includes math formalism, but most of the questions rely on qualitative understanding of the relationships between equations rather than quantitative calculations. It contains many questions about the Schrödinger equation and a few about measurement as a theoretical construct (e.g., given a wave function, make a measurement, what is the new wave function). There are many questions that use infinite square well potentials and a couple which ask students to think about non-standard potentials qualitatively. The developers recommend using the QMCA as a post-test for sophomore level modern physics classes. It could be used as a pre-test in graduate level quantum to see if students have sufficient conceptual understanding of undergraduate level quantum topics. The multiple-choice questions on the QMCA were developed using the open-ended questions on the QMAT as a starting point.

The Quantum Mechanics Assessment Tool[65] (QMAT) questions are open-ended and are a mix of conceptual and math intensive questions, where students are asked to solve equations in some of the questions. The QMAT covers the same five main topics of quantum measurement as the QMCA and is also meant for a first-semester junior-level quantum mechanics. It should be given as a post-test only at the end of the term. It was designed to measure student learning of concepts most valued by faculty, assess students' learning difficulties, and inform course improvement. The content of the QMAT is based on working with faculty to determine learning goals for quantum mechanics. A couple of the questions were taken from an early version of the QMCS. There is a rubric for grading the test, but the rubric requires extensive training to get acceptable inter-rater reliability. Further, because this is an open-ended assessment it is difficult to compare results to other institutions. There are limited validation studies of the QMAT, and it has been archived by the developers, so you should use the QMCA, unless you specifically want a short-answer test. Further, the QMCA has been more thoroughly researched and validated.

The Quantum Mechanics Survey[66] (QMS) is a multiple-choice assessment for the junior and graduate-level. The QMS has a wide range of topics including wave functions, the expectation value of a physical observable and its time dependence, the role of the Hamiltonian, stationary and non-stationary states and issues related to their time development, and measurements.[66] All questions are restricted to one-dimensional quantum mechanics models. The QMS should be given as a post-test only in a junior-level course, but can be given as a pre- and post-test in a graduate level quantum course. The QMS was designed not only to assess students' conceptual understanding of quantum mechanics but also contains an extensive mathematical formalism. Although students do not have to complete difficult integrals to solve any of the questions, they do need to understand the basics of linear algebra. Topics covered on the QMS are those that faculty find important for junior-level quantum mechanics courses.

The content covered by the QMS and QMCA is very similar, but the QMS is more difficult and mathematical than the QMCA, and contains a lot more equations. Both have similar formats and levels of research validation.

The Quantum Mechanics Formalism and Postulates Survey[67] (QMFPS) is the newest quantum mechanics assessment. It assesses students' understanding of the formalism and postulates of quantum mechanics. The QMFPS is a pre/post multiple-choice test that is appropriate for junior/senior level quantum mechanics courses where Dirac notation has been covered similar to coverage in the first four chapters of Griffiths. The QMFPS is not meant to assess students' mathematical skills, but students do need to know the basics of linear algebra to answer the questions. The multiple-choice questions on the QMFPS were developed based on expert feedback about relevant topics, review of course materials, and a subsequent test blueprint.

The QMFPS, like the QMCA and the QMS, is meant to assess students' conceptual understanding of quantum mechanics, but the QMFPS focuses particularly on students' understanding of the formalism and postulates of quantum mechanics.

The Quantum Mechanics Visualization Instrument[68] (QMVI) is a multiple-choice exam and was the first quantum mechanics survey created. It was designed to assess students' understanding of quantum topics at all levels, from sophomore-level to graduate-level. Most of the questions are about the relationship between the shape of the potential and the wave function, with an emphasis on visualizing this relationship. There are a few questions about the uncertainty principle, and two questions about momentum space probability distributions. Some of the questions require "tricks" to figure out, e.g., making a symmetry argument makes a question very easy, but without the symmetry argument, it is very difficult. The questions are multiple-choice, and also ask students to give a 2–3 line written response and a rating of their confidence level. The QMVI contains 25 questions at all different levels, with very simple questions for sophomore-level students, and very difficult questions for graduate-level students. Because of the variety in difficulty of the questions, it can be used to track students' progress throughout the quantum course sequence. Since it contains questions at the graduate-level, it is a very difficult test. The QMVI contains extensive mathematical formalism. The developers recommend giving it as an extended take-home exam, as it can take up to two hours for students to complete. The topics covered are those that authors feel are important for students to learn in the quantum sequence.

63. "Constructing a multiple-choice assessment for upper-division quantum physics from an open-ended tool," H. R. Sadaghiani, J. Miller, S. J. Pollock, and D. Rehn, in *Physics Education Research Conference 2013*, Portland, OR (2013), pp. 319–322, <http://www.compadre.org/per/items/detail.cfm?ID=13190>. (E)

64. "Quantum mechanics concept assessment: Development and validation study," H. R. Sadaghiani and S. J. Pollock, Phys. Rev. Spec. Top. - Phys. Educ. Res. **11**, 10110 (2015). (I)

65. "Transforming upper-division quantum mechanics learning goals and assessment," S. Goldhaber, S. J. Pollock, M. Dubson, P. Beale, and K. K. Perkins, AIP Conf. Proc. **1179**(1), 145–148 (2009), <http://www.compadre.org/per/items/detail.cfm?ID=9475>. (I)

Table IX. Thermodynamics assessments.

Title	Content	Intended Population	Research Validation	Purpose
Thermodynamics Conceptual Survey (TCS)	Temperature, heat transfer, ideal gas law, 1st law of thermodynamics	Intro college, intermediate	Silver	To assess students' understanding of heat and temperature, the ideal gas law, the first law of thermodynamics and processes
Thermal Concept Evaluation (TCE)	Heat, temperature, heat transfer	Intro college, high school	Silver	To assess introductory college or 3rd-year high school students' understanding and application of thermodynamics concepts using common contexts that reflect students' own conceptions
Survey of Thermodynamic Processes and First and Second Law (STPFaSL)	1st and 2nd law of thermodynamics, thermodynamics processes	Intro college, upper-level graduate	Silver	To measure the effectiveness of traditional and/or research-based techniques of teaching the first and second laws of thermodynamics and thermodynamic processes
Heat and Temperature Conceptual Evaluation (HTCE)	Heat, temperature, specific heat capacity, phase changes	Intro college	Bronze	To assess students' understanding of heat and temperature concepts

66. "Surveying students' understanding of quantum mechanics in one spatial dimension," G. Zhu and C. Singh, Am. J. Phys. **80**(3), 252–259 (2012). (E)

67. "Improving the quantum mechanics content knowledge and pedagogical content knowledge of physics graduate students," E. M. Marshman, Ph.D. dissertation, University of Pittsburgh (2015), <http://d-scholarship.pitt.edu/25547/>. (I)

68. "Development and validation of an achievement test in introductory quantum mechanics: The Quantum Mechanics Visualization Instrument (QMVI)," E. Cataloglu, Ph.D. dissertation, Pennsylvania State University (2002), <https://etda.libraries.psu.edu/catalog/5937>. (I)

1. Recommendations for choosing an upper-level quantum mechanics assessment

If you are teaching a junior- or senior-level quantum mechanics course, which test you use depends on both the difficulty level and the range of topics you want to cover: In terms of difficulty, the QMCS is at the lowest level, followed by the QMCI and QPCS, the QMCA, and then the QMVI, QMFPS, and QMS. In terms of content, all quantum RBAIs cover some basic ideas about wave functions. The QMVI focuses in great depth on the relationship between the wave function and the shape of the potential. The QPCS is the only assessment that covers the photoelectric effect. The QMCI is entirely conceptual, whereas the QMCA, QMFPS, and QMS require some formalism. The QMFPS has a different focus than the other upper-level quantum assessments, should be used if you want to specifically assess your students' understanding of quantum formalism and postulates.

VI. THERMODYNAMICS ASSESSMENTS

A. Overview of thermodynamics assessments

There are four RBAIs for thermodynamics concepts: The Thermodynamics Conceptual Survey[69] (TCS), Thermal Concept Evaluation[70] (TCE), Heat and Temperature Conceptual Evaluation[71] (HTCE), and Survey of Thermodynamic Processes and First and Second Laws[72] (STPFaSL) (Table IX). All of these assessments were developed for introductory level courses. The Thermal and Transport Concept Inventory-Thermodynamics[73] (TTCI-T) and the Thermodynamics Concept Inventory[74] (TCI) were developed specifically for engineering courses, and will not be discussed further here. We are not aware of any research-based assessments on statistical mechanics.

The Thermodynamics Conceptual Survey[69] (TCS) is a multiple-choice pre/post conceptual assessment of heat and temperature, the ideal gas law, and the first law of thermodynamics for introductory physics courses. It consists of two parts with part one covering temperature, heat transfer, and the ideal gas law; and part two covering the first law of thermodynamics. It is split into two parts so that you can choose the part(s) that most closely match the content covered in your course. The questions on the TCS are all either adapted from other thermodynamics tests or studies of students' understanding of thermodynamics topics. In addition to assessing your students' understanding of these thermodynamics topics, the authors suggest that the questions may be used as teaching materials to help students overcome conceptual difficulties. The TCS was developed in Thailand.

The Thermal Concept Evaluation[70] (TCE) is a multiple-choice pre/post-conceptual assessment of heat transfer, temperature change, and thermal properties of materials. It was developed for third-year high school students and introductory college students in Australia. The multiple-choice answers allow students to choose from "everyday physics" answers or "classroom physics" answers. Many questions consist of a conversation between students and then statements about the opinions of the students involved in the conversation. There are no diagrams or graphs. The TCE questions were developed based on an inventory of students' alternative conceptions of thermodynamics from the research literature. The TCE was developed in Australia.

The Heat and Temperature Conceptual Evaluation[71,75] (HTCE) is a multiple-choice pre/post conceptual assessment of heat, temperature, and heat transfer for introductory physics courses. A majority of the questions are about heat transfer of various materials in cups and about a third have to do with graphing temperature versus time. The HTCE questions were developed based on research into student thinking.

The TCS shares many commonalities with the TCE and HTCE because its questions were adapted from various other RBAIs or interview tasks. TCS questions 2, 4, 5, and 6 are the same as TCE questions 8, 11, 14, and 6. TCS questions 1 and 3 are the same as HTCE questions 1 and 8. The TCS covers more thermodynamics concepts than either the TCE or HTCE. The TCS includes questions on the first law of thermodynamics and the ideal gas law, whereas neither the TCE nor HTCE contain these topics. Some of the questions on the TCS are also more complex than those on the HTCE or TCE. For example, there is an explanation of a five-step process of a gas being compressed by a piston, and students are asked questions about work, heat, and energy at various points in the process. The TCS is also the only thermodynamics test that asks students to interpret P vs V graphs.

The heat and temperature concepts covered on the HTCE are very similar to those covered on the TCE, though the questions on the TCE focus on students' everyday experiences of heat and temperature and many present conversations where students are asked to indicate who they agree with. The HTCE and TCS are more formal and focus on the content of thermodynamics in a physics course. The TCE would be better used as a pre-test, because it focuses on everyday language. The HTCE has three questions about temperature versus time graphs, whereas the TCE has no questions about graphs.

The Survey of Thermodynamic Processes and First and Second Laws[72] (STPFaSL) is the newest pre/post-multiple-choice conceptual thermodynamics assessments for introductory algebra-based and calculus-based physics courses. It can also be used in upper-level or graduate courses when relevant content is covered. It assesses the first and second laws of thermodynamics and thermodynamic processes. The questions on the STPFaSL were developed by consulting with instructors of introductory courses about topics and content goals, as well as materials for these courses, and a literature review of student difficulties with relevant topics.

Both the STPFaSL and the TCS contain questions on the first law of thermodynamics and ask students about pressure vs volume graphs. The STPFaSL also asks questions about the second law of thermodynamics. The TCS and STPFaSL have similar formats and levels of research validation.

69. "Development and implementation of a conceptual survey in thermodynamics," P. Wattanakasiwich, P. Taleab, M. D. Sharma, and I. D. Johnston, Int. J. Innovations Sci. Math. Educ. **21**(1), 29–53 (2013), <http://www.compadre.org/per/items/detail.cfm?ID=13523>. (E)

70. "Introductory thermal concept evaluation: Assessing students' understanding," S. Yeo and M. Zadnik, Phys. Teach. **39**(8), 496–504 (2001). (E)

71. "Surveying Thai and Sydney introductory physics students' understandings of heat and temperature," C. Tanahoung, R. Chitaree, C. Soankwan, M. Sharma, and I. Johnston, in Proceedings of the Assessment in Science Teaching and Learning Symposium (2006), <http://www.compadre.org/per/items/detail.cfm?ID=14034>. (E)

72. "Developing and assessing research-based tools for teaching quantum mechanics and thermodynamics," B. R. Brown, Ph.D. dissertation, University of Pittsburgh (2015), <http://d-scholarship.pitt.edu/25903/>. (I)

73. "Rigorous methodology for concept inventory development: Using the "assessment triangle" to develop and test the thermal and transport science concept inventory (TTCI)," R. A. Streveler, R. L. Miller, A. I. Santiago-Román, M. A. Nelson, M. R. Geist, and B. M. Olds, Int. J. Eng. Educ. **27**(5), 968–984 (2011). (E)

74. "Development of engineering thermodynamics concept inventory instruments," K. C. Midkiff, T. A. Litzinger, and D. L. Evans, in 31st ASEE/IEEE Frontiers in Education Conference, F2A–3, Reno, NV (2001), <http://ieeexplore.ieee.org/document/963691/>. (E)

75. "Surveying Thai and Sydney introductory physics students' understandings of heat and temperature," C. Tanahoung, M. D. Sharma, I. D. Johnston, R. Chitaree, and C. Soankwan, in Australian Institute of Physics 17th National Congress, Brisbane (2006), <http://www.compadre.org/PER/items/detail.cfm?ID=14034>. (E)

B. Recommendations for choosing a thermodynamics assessment

Use the TCS if you want to assess the first law of thermodynamics in addition to other topics such as temperature, heat transfer, phase change, and thermal properties of materials. Use the STPFaSL is you are interested in assessing the second law of thermodynamics in addition to the first law and thermodynamic processes. The TCE uses everyday language and ideas that would be familiar to students before a physics course, so it would be appropriate to use as a pre-test with students who have not seen this content before. Further, the format of the TCE where students answer questions about a student discussion could help students get into the frame of mind of discussion and not test taking, which might help you understand their ideas more deeply. You could also use the STPFaSL and TCS as pre-tests in courses where you think students will understand the formal terms before taking the course.

VII. OPTICS AND WAVES ASSESSMENTS

A. Optics

There is one assessment of geometrical optics, the Four Tier Geometrical Optics Test[76] (FTGOT) which is a pre/post-conceptual assessment for introductory college courses (Table X). The questions on the FTGOT ask about observing oneself and observing others with plane mirrors, spherical mirrors, and lenses. The FTGOT has "four tiers" of sub-questions for each main question. These ask students to answer a multiple-choice content question, rate their confidence in their answer, indicate their reasoning (also multiple-choice), and then rate their confidence in their reasoning. The test structure can give instructors more confidence that a correct answer to the content question does actually indicate understanding by the student. The FTGOT questions were developed based on a literature review, the developers experience teaching these topics and open-ended interviews with students. The FTGOT was developed in Turkey. Use the FTGOT if you want to assess your students' understanding of geometrical optics concepts at the introductory level.

76. "Development and application of a four-tiered test to assess pre-service physics teachers' misconceptions about geometrical optics," D. Kaltakci, Ph.D. dissertation, Middle East Technical University (2012), <http://www.compadre.org/per/items/detail.cfm?ID=14002>. (E)

B. Introductory waves assessments

There are four RBAIs about waves, three for introductory-level courses, the Mechanical Wave Conceptual Survey[77] (MWCS), the Mechanical Wave Conceptual Survey 2[78] (MWCS-2), the Wave Diagnostic Test[79] (WDT), and one for upper-level courses, the Wave Concept Inventory[80] (WCI) (Table X).

Table X. Optics and waves assessments.

Title	Content	Intended population	Research validation	Purpose
Optics				
Four Tier Geometrical Optics Test (FTGOT)	Plane mirrors, spherical mirrors, lenses	Intro college	Silver	To assess misconceptions in geometric optics
Waves				
Mechanical Wave Conceptual Survey (MWCS)	Mechanical waves, wave propagation, wave superposition, wave reflection, standing waves	Intro college, intermediate, high school	Silver	To identify students' alternative conceptions about mechanical waves before instruction and evaluate the effectiveness of instruction at the end of a course
Mechanical Wave Conceptual Survey 2 (MWCS-2)	Mechanical waves, wave propagation, wave superposition, wave reflection, standing waves	Intro college, intermediate, high school	Silver	To assess students' understanding of basic wave concepts using a standard multiple-choice questions format
Wave Diagnostic Test (WDT)	Waves	Intro college, intermediate, high school	Silver	To understand students' thinking about basic wave concepts
Wave Concept Inventory (WCI)	Visualization of waves, mathematical depiction of wave, wave definitions	Upper-level	Bronze	To assess students' understanding of wave phenomena in an integrated upper-division engineering course on electronic and electromagnetic topics

The Mechanical Wave Conceptual Survey[77] (MWCS) is a multiple-choice pre/post assessment of basic wave concepts covered in introductory courses, though it has also been tested with high school students. The MWCS has four subtopics including propagation, superposition, reflection, and standing waves. Several questions have more than 5 answer options and several questions ask students about their reasoning in addition to their answer (as a two part question). Because of this non-standard question format, a standard Scantron answer sheet would not work. The questions were created based on the open-ended questions from the WDT. The MWCS was developed in Thailand and Australia.

The Mechanical Wave Conceptual Survey 2[78] (MWCS-2) is a modification of the MWCS. The MWCS-2 is modified to make MWCS questions into the standard multiple-choice formats with five answer options each. The developers of the MWCS-2 made changes to the wording of questions as well as adding and removing answer choices for some questions. Further, for the MWCS questions with two parts (answer and reasoning), the MWCS-2 combines the answer and reasoning together, so that these questions just have five standard answer choices. Besides the modifications discussed above, the content tested on the MWCS and MWCS-2 is the same. The MWCS-2 was developed in Mexico.

The Wave Diagnostic Test[79] (WDT) has both free-response and multiple-choice questions about mechanical and sound waves topics covered in a typical introductory physics course. The main purpose of the WDT is to learn about students' thinking about waves, not to compare students' scores to a baseline. The WDT elicits rich and varied responses from students that show what they believe about waves and why. This makes the WDT very useful as a benchmark, and allows you to more accurately tailor your instruction to the incoming beliefs of your students. Because the WDT is meant to understand students' thinking, it is not scored. There are two parts to the WDT, and students should complete and turn in part 1 before completing part 2.

The questions on the WDT and MWCS are very similar, since the MWCS was developed from the WDT, but all the questions on the MWCS are multiple-choice, whereas many of the questions on the WDT are free-response. The MWCS is scored in the standard way (% correct), whereas the WDT is meant to be used to understand your students' ideas, and therefore is not scored.

77. "Developing, evaluating and demonstrating the use of a conceptual survey in mechanical waves," A. Tongchai, M. D. Sharma, I. D. Johnston, K. Arayathanitkul, and C. Soankwan, Int. J. Sci. Educ. **31**(18), 2437–2457 (2009). (E)

78. "Mechanical waves conceptual survey: Its modification and conversion to a standard multiple-choice test," P. Barniol and G. Zavala, Phys. Rev. Phys. Educ. Res. **12**(1), 10107 (2016). (I)

79. "Making sense of how students come to an understanding of physics: An example from mechanical waves," M. C. Wittmann, Ph.D. dissertation, University of Maryland, College Park (1998), <http://www.compadre.org/per/items/detail.cfm?ID=5687&Relations=1>. (E)

1. Recommendations for choosing a waves assessment

Use the MWCS-2 to assess students' understanding of mechanical waves in introductory physics courses if you want to compare students' scores before and after your course with an assessment that is quick and easy to score. Use the WDT for introductory courses if you want to understand students thinking about mechanical waves in a more in-depth way.

C. Upper-level waves assessments

The Wave Concept Inventory[80] (WCI) is a multiple-choice pre/post-assessment of upper-level wave phenomenon content including visualization of waves, mathematical depiction of waves, and wave definitions. It was designed to assess the effectiveness of an integrated electrical engineering course covering quantum mechanics and Schrödinger's wave equation as well as Maxwell's wave equations and their application to the propagation of electromagnetic waves, though could also be appropriate for an upper-division physics course. Some of the questions have more than one correct answer, which more thoroughly assess students understanding of the content. There are no calculational questions on the WCI, but students are asked about mathematical equations (e.g., which linear partial differential equation can be used to model wave propagation). The WCI questions were developed by the instructors of an integrated electrical engineering course.

The concepts covered on the WCI are for upper-level engineering courses, though could also be used at the upper-level in a physics department. The WDT and MWCS are meant for introductory courses, so the content and level of these tests are very different. Use the WCI for your upper-level course if the content on the test aligns with what you teach in your class.

80. "The wave concepts inventory-an assessment tool for courses in electromagnetic engineering," R. J. Roedel, S. El-Ghazaly, T. R. Rhoads, and E. El-Sharawy, in *8th Annual Frontiers in Education Conference* (1998), Vol. 2, pp. 647–653., <http://www.compadre.org/Per/items/detail.cfm?ID=13754> (I)

VIII. ASTRONOMY

There are eight RBAIs for astronomy, and all are designed for use in the introductory astronomy course. Three of these, the Astronomy Diagnostic Test 2.0[81] (ADT2), the Test of Astronomy Standards[82] (TOAST), and the Astronomical Misconceptions Survey[83] (AMS), contain questions about a wide range of topics covered in an introductory astronomy course and can be used to assess the overall effectiveness of your course. Five of these, the Star Properties Concept Inventory[84] (SPCI), the Light and Spectroscopy Concept Inventory[85] (LSCI), the Newtonian Gravity Concept Inventory[86,87] (NGCI), the Lunar Phases Concept Inventory[88] (LPCI), and the Greenhouse Effect Concept Inventory[89] (GECI) cover a more narrow range of content, and can be used to assess your students' understanding of specific content from your course. All of these RBAIs are multiple-choice. All astronomy assessments are summarized in Table XI.

A. General astronomy assessments

The Astronomy Diagnostic Test 2.0[81] (ADT2) is a multiple-choice conceptual pre/post-test for non-science majors taking an introductory astronomy course and covers content commonly found in the K-12 curriculum including seasons, lunar phases, motions in the sky, and size and scale. It was designed to help instructors assess their students' initial knowledge coming into a college astronomy course, as the topics included were likely covered in K-12. The multiple-choice questions on the most recent version of the Astronomy Diagnostic Test (ADT), version 2.0, come from an earlier version of the ADT, which consisted of questions from several earlier astronomy tests.

The Test of Astronomy Standards[82] (TOAST) is a multiple-choice broad conceptual assessment of general astronomy content knowledge that is built on and from earlier astronomy assessments (all the astronomy assessments included here). The content includes gravity, electromagnetic radiation, fusion and formation of heavy elements, evolution of the universe, star and stellar evolution, evolution and structure of the solar system, seasons, scale, yearly patterns, daily patterns, moon

Table XI. Astronomy assessments.

Title	Content	Intended population	Research validation	Purpose
General Astronomy Assessments				
Astronomy Diagnostic Test 2.0 (ADT2)	Seasons, lunar phases, motions in the sky, and size and scale	Intro college	Gold	To assess students' conceptual understanding of introductory astronomy topics
Test of Astronomy Standards (TOAST)	General astronomy content knowledge	Intro college	Silver	To measure students' mastery of core concepts in a general astronomy course
Astronomical Misconceptions Survey (AMS)	Misconceptions about introductory astronomy courses	Intro college	Research-based	To identify misconceptions introductory students hold and measure the effectiveness of instruction to dispel these misconceptions
Specific astronomy topic assessments				
Star Properties Concept Inventory (SPCI)	Stellar properties, nuclear fusion, star formation	Intro college	Gold	To measure student learning about the properties and formation of stars
Light and Spectroscopy Concept Inventory (LSCI)	Light, waves, spectroscopy	Intro college	Silver	To measure students' conceptual understanding of topics related to light and spectroscopy, and evaluate the effectiveness of instruction in introductory college astronomy courses
Newtonian Gravity Concept Inventory (NGCI)	Gravity	Intro college	Silver	To assess student understanding of Newtonian gravity and effectiveness of instruction in general education introductory college astronomy course
Greenhouse Effect Concept Inventory (GECI)	Types of greenhouse gases, energy equilibrium balance, greenhouse effect mechanisms, global warming vs greenhouse effect	Intro college	Silver	To assess pre- and post-instruction conceptual understanding of the greenhouse effect focusing on the physics of energy flow through Earth's atmosphere
Lunar Phases Concept Inventory (LPCI)	Phases of the moon	Intro college	Bronze	To assess college students' mental models of lunar phases

phases. The content on the TOAST was determined based on that which was deemed more important for introductory astronomy students as described in expert position statements from several professional organizations[90,91] and later reviewed by 28 experts in astronomy. This makes it a unique astronomy RBAI, as the topics are broad, covering the whole intro course, and are chosen based on based on consensus documents from the astronomy community. Further, most of the questions are taken from other astronomy RBAIs.

The TOAST and ADT2 cover very similar content including phases of the moon, motions in the sky, seasons, scale, distances, sizes, properties, and lifecycles of stars, gravity, and the universe. There are several questions that are the same on both tests since the TOAST was created using questions from other astronomy assessments. The TOAST contains questions about production of light (emission, absorption, etc.), while both tests ask about the relative speed of electromagnetic waves. There TOAST asks about the Big Bang, and the ADT2 does not. The ADT2 has one question about global warming, and the TOAST does not. Both tests are general assessments for introductory astronomy. They have both been well validated. The ADT2 been used widely in introductory astronomy courses across the US, so there is a lot of comparison data available. The TOAST is a newer assessment, so there is less comparison data available now, but this will likely change in the near future.

The Astronomical Misconceptions Survey[83] (AMS) is a pre/post-conceptual multiple-choice survey of common misconceptions in introductory astronomy, e.g., the phases of the moon are caused by the earth's shadow or the seasons are caused by differences in the earth's distance from the sun. There are two versions of the AMS: the true/false version and the multiple-choice version. The true/false version can be used to help instructors understand the misconceptions their students come to their course holding. The multiple-choice version can be given to students to help instructors understand the misconceptions their students have or to assess the effectiveness of different types of instruction at addressing these misconceptions. The questions on the AMS are not about a particular topic, but instead a variety of topics for which students have commonly held incorrect beliefs. The questions on the AMS were developed from a list of 25 astronomy misconceptions, which were based on previous research on misconceptions.

Because the AMS is a test of students' misconceptions about astronomy, the topics covered and the focus of the questions is very different from the questions on the ADT2 and TOAST.

81. "Development of the astronomy diagnostic test," B. Hufnagel, Astron. Educ. Rev. **1**(1), 47–51 (2002), <http://www.compadre.org/PER/items/detail.cfm?ID=13745>. (E)

82. "The development and validation of the Test of Astronomy Standards (TOAST)," S. J. Slater, J. Astron. Earth Sci. Educ. **1**(1), 1–22 (2014), <https://www.cluteinstitute.com/ojs/index.php/JAESE/article/view/9102/9224>. (E)

83. "An astronomical misconceptions survey," B. M. C. Lopresto and S. R. Murrell, J. Coll. Sci. Teach. **40**(5), 14–22 (2011), <http://www.compadre.org/PER/items/detail.cfm?ID=14009>. (E)

B. Specific astronomy topic assessments

The Star Properties Concept Inventory[84] (SPCI) is a multiple-choice pre/post conceptual assessment of stellar properties, nuclear fusion, and star formation for introductory astronomy courses. The SPCI questions were developed based on exam and textbook questions and the authors experience with teaching the content. It was developed in response to research on students' alternative conceptions about stars.[92]

The Light and Spectroscopy Concept Inventory[85] (LSCI) is a multiple-choice pre/post conceptual test about the electromagnetic spectrum and the nature of light and is meant for introductory astronomy courses. These specific topics have been chosen because they were found to be central topics common across most introductory astronomy courses. The more narrow range of topics means that there are multiple questions probing each. Students usually score near guessing (25%) on the pre-test, implying that the LSCI is testing material unfamiliar to students. That said, most instructors still give it as a pre- and post-test. The LSCI questions were developed based on expert opinions about the important core knowledge around light and the electromagnetic spectrum, and research on student ideas about light and quantum phenomena.

The Newtonian Gravity Concept Inventory[87] (NGCI) is a multiple-choice pre/post-conceptual assessment of gravity, a foundational topic in introductory astronomy courses. The questions probe four conceptual dimensions including the directionality of gravity, the force law, independence of other forces (e.g., gravity is not affected by rotation), and thresholds related to gravity (e.g., there is not distance for which gravity suddenly stops). The NGCI was developed for use in introductory astronomy courses, but can also be used in introductory physics. The questions are based on student ideas about gravity.

The Lunar Phases Concept Inventory[88] (LPCI) is a multiple-choice pre/post conceptual assessment of lunar phases concepts including cause and period of lunar phases, period and direction of the Moon's orbit, and observational phenomena. It is designed to assess students' mental models of lunar phases using a mathematical technique called model analysis theory.[88] The result of this analysis is the probability of students in a course answering with the correct model as well as the probability of answering with one of several incorrect models. The LPCI can also be analyzed and scored in the more common way of finding the percent correct on the pre- and post-test and then calculating the normalized gain. Furthermore, since the test content was developed based on students' ideas about the lunar phases, as opposed to expert opinions about the most important content related to lunar phases, it is most appropriate to use the LPCI to understand your students' thinking and mental models, instead of how well their ideas match expert conceptions.

The Greenhouse Effect Concept Inventory[89] (GECI) is a multiple-choice pre/post-conceptual assessment about the physics of energy flow through Earth's atmosphere. Topics include types of greenhouse gases, types of electromagnetic energy, energy equilibrium balance, greenhouse effect mechanisms, global warming versus the greenhouse effect. The GECI can be used in introductory astronomy courses that cover relevant content. The questions were developed based on extensive research on students' beliefs about models of the greenhouse effect.

Because the content of these specific astronomy topic assessments is so different, we don't compare them.

84. "Development of a concept inventory to assess students' understanding and reasoning difficulties about the properties and formation of stars," J. M. Bailey, Astron. Educ. Rev. **6**(2), 133–139 (2007). (E)

85. "Development and validation of the light and spectroscopy concept inventory," E. M. Bardar, E. E. Prather, K. Brecher, and T. F. Slater, Astron. Educ. Rev. **5**(2), 103–113 (2006). (E)

86. "Development and calibration of a concept inventory to measure introductory college astronomy and physics students' understanding of Newtonian gravity," K. E. Williamson, Ph.D. dissertation, Montana State University (2013), <http://scholarworks.montana.edu/xmlui/handle/1/3027>. (E)

87. "Development of the Newtonian gravity concept inventory," K. E. Williamson, S. Willoughby, and E. E. Prather, Astron. Educ. Rev. **12**(1), 010107 (2013). (E)

88. "Developing the lunar phases concept inventory," R. S. Lindell and J. P. Olsen, in *Physics Education Research Conference 2002*, Boise, ID (2002), pp. 1–4, <http://www.per-central.org/items/detail.cfm?ID=4323>. (E)

89. "Part I: Development of a concept inventory addressing students' beliefs and reasoning difficulties regarding the greenhouse effect; part II: Distribution of chlorine measured by themars odyssey gamma ray spectrometer," J. M. Keller, Ph.D. dissertation, University of Arizona (2008), <http://hdl.handle.net/10150/193632>. (I)

90. *National Science Education Standards* (National Research Council, Washington DC, 1996). (E)

91. *Project 2061: Benchmarks for Science Literacy* (American Association for the Advancement of Science, Washington, DC, 1986). (E)

92. "Development and validation of the star properties concept inventory," J. M. Bailey, B. Johnson, E. E. Prather, and T. F. Slater, Int. J. Sci. Educ. **34**, 2257–2286 (2011). (E)

93. "Astronomy Diagnostic Test 2.0 (ADT2)," https://www.physport.org/assessments/ADT

94. "Light and Spectroscopy Concept Inventory (LSCI)," www.physport.org/assessments/LSCI

C. Recommendations for choosing an astronomy assessment

Use the TOAST or ADT2 if you are making changes to your entire introductory astronomy course, and want to measure the effectiveness of the change. Use the TOAST if you want to assess students' understanding of how light is produced in addition to other standard introductory concepts. Use the ADT2 as a pre-test if you want to understand the ideas your students bring to your course from their K-12 education. Use the AMS if you are particularly interested in understanding your students' misconceptions about astronomy.

If instead you are making changes to a specific portion of your course, use an assessment of specific topics that match the content you are changing (SPCI, LSCI, NGCI, LPCI, or GECI). The developers of the LSCI point out that the topics covered on the LSCI (electromagnetic spectrum and the nature of light) are foundational and central in many astronomy courses, so you could use this test as a proxy for understanding the effectiveness of your instruction for your course, even though it covers only a subset of the material. Furthermore, if comparing your students' scores to others is important to you, use either the ADT2 or LSCI, as there is a large amount of comparison data published. A list of articles with ADT2 and LSCI comparison data can be found on the research tab on their respective assessment pages[93,94] on PhysPort. The TOAST is a newer assessment, so there is less comparison data available now, but this will likely change in the near future.

Table XII. Summary of RBAs of physics and astronomy content.

Topic	Names	N
Mechanics		
Kinematic and Forces—Intro	FCI, FMCE, MBT, IBCM, TUG-K, FVA	6
Energy—Intro	EMCS, ECA	2
Rotation—Intro	RRMCS, RKI	2
Density—Intro	DS	1
Classical mechanics—Intermediate	CCMI	1
Electricity and Magnetism		
Electrostatics and magnetism—Intro	BEMA, CSEM, DEEM, EMCA, SGCE, MCS	6
Circuits—Intro	DIRECT, ECCE, IBCDC	3
Electricity and magnetism—Intermediate	CUE-FR, CUE-CMR, CURrENT	3
Quantum mechanics and modern physics		
Relativity—Intermediate	RCI	1
Quantum mechanics—Intermediate	QPCS, QMCS, QMCI	3
Quantum mechanics—Upper-level	QMCA, QMAT, QMVI, QMS, QMFPS	5
Thermodynamics		
Thermodynamics—Intro	TCS, TCE, STPFaSL, HTCE	4
Optics and waves		
Optics—Intro	FTGOT	1
Waves—Intro	MWCS, MWCS-2, WDT	3
Waves—Upper-level	WCI	1
Astronomy		
Astronomy—General—Intro	ADT2, TOAST, AMS	3
Astronomy—Specific topics—Intro	SPCI, LSCI, NGCI, GECI, LPCI	5
Total		50

IX. CONCLUSION

Table XII summarizes the 50 RBAIs of physics and astronomy content discussed in this resource letter. We have found RBAIs in nearly every major content area in physics, with the exception of statistical mechanics. Most topics have RBAIs at both introductory and the upper-levels.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the contributions of the other members of the PhysPort team and KSUPER who worked on this project: John D. Thompson, Jaime Richards, Devon McCarthy, and Brian Danielak. This work was partially supported by NSF Grant Nos. PHYS-1461251, DUE-1256354, DUE-1256354, DUE-1347821, and DUE-1347728.