

The use of representations in evidence-based and non-evidence-based physics activities

Joshua S. Von Korff,¹ Chang Zhan,¹ Birjoo Vaishnav,² Jacquelyn J. Chini,³ Ashley Warneke,¹ and Ozden Sengul¹

*¹Georgia State University, Department of Physics and Astronomy,
25 Park Place, Atlanta, GA, 30303*

²University of West Georgia, Department of Physics, 1601 Maple Street, Carrollton, GA, 30118

³University of Central Florida, Department of Physics, 4111 Libra Drive, Orlando, FL 32816

Abstract. Representations such as graphs, pictures, words, and equations play a crucial role in the pedagogy of introductory physics. Activities such as labs and tutorials often require students to analyze and produce these representations. We investigate the use of representations in a variety of activities. Specifically, we examine both labs and tutorials as well as both evidence-based and non-evidence-based activities. We define evidence-based activities as those that have been connected with positive learning outcomes in published research articles. Our evidenced-based activities are inspired by physics education research, while our non-evidence-based activities are more traditional. We compare and contrast the use of representations among these types of activities. We investigate several uses of representations, including representations produced by the student; representations provided by the activity; and representations that the student re-examines at a later point in the activity.

PACS: 01.40.Fk

I. INTRODUCTION

A major contribution of Physics Education Research (PER) has been to produce carefully designed curricular materials in the form of activities for students. Because most PER instructional strategies rely on collaborative group work, activities are a key feature of these strategies. We classify activities as “evidence-based” (EB) when they have been associated with high learning gains in a research publication and “not evidence-based” (NEB) otherwise. For example, high gains might be associated with the Force Concept Inventory (FCI) or a similar assessment [1].

The use of multiple representations is often said to be important for learning physics [2], so we hypothesized that evidence-based activities would use representations differently from non-evidence-based activities. Designers of activities might be aware of the literature on multiple representations or they might come to their own conclusions about representations through experimentation. The current paper describes the results of an effort to test this hypothesis.

We utilize two kinds of activities: tutorials and labs. We have chosen to compare all EB activities to all NEB activities, whether they are tutorials or labs. We chose to compare tutorials directly to labs for two reasons:

(1) The distinction is not absolute: for instance, many tutorials involve equipment. Also, EB labs are often more

tutorial-like than NEB labs. Thus, there is a continuum from lab to tutorial.

(2) Tutorials and labs can compete for scheduled time in a physics course. For example, instructors may have to choose whether to run a three hour lab or a one hour tutorial with a two hour lab. They may also compete for instructional resources, since teaching assistants may have a limited amount of time to supervise one or the other.

We do not claim that EB tutorials and EB labs are always similar. That said, we focus on features that differentiate EB activities, collectively, from NEB activities, because we aim to uncover widely applied principles that are implicit in the design of EB activities.

In this paper, we will review the literature, then describe how we selected our activities. We will discuss our coding scheme and statistical method. Finally, we will discuss our results. We will propose a heuristic quantity (“V / Q” – the ratio of verbal to quantitative student responses) that is empirically found to be large for our evidence-based activities and small for our non-evidence-based activities.

II. LITERATURE REVIEW

Here, we describe two themes in prior studies: (1) use of representations in introductory physics courses and (2) comparisons of “cookbook” and “interactive engagement” style documents.

REPRESENTATIONS: Researchers have taken many approaches to describing the use of representations in physics [3,4]. An important question related to the present study is whether instructional use of representations translates to student representational fluency. Kohl and Finkelstein found that students in a reform style introductory physics course learned a broader set of representational skills than those in a more traditional course [2]. The researchers used observations and document analysis to describe the time and emphasis devoted to various representations and found the reformed course made use of a richer set of representations and more frequent use of multiple representations. This suggests that exposing students to more representations may improve their ability to use those representations.

COOKBOOK: Royuk and Brooks defined cookbook activities as those with “excessively detailed instructions that allow students merely to follow a recipe without having to think about what they’re doing” [5]. For example, hallmarks of a cookbook laboratory exercise may include step-by-step experimental procedures, “fill-in-the-blank” data tables, and questions posed to the student mainly at the end of the experiment. On the other hand, more interactive activities are marked by “heads-on” activities and peer and instructor interaction. Thus, an interactive laboratory exercise may integrate more reflective questions into the procedure and emphasize concept formation. Royuk and Brooks found that students who used RealTime Physics had higher FCI scores than those who completed traditional cookbook labs. Karelina and Etkina looked at the effect of integrating required student explanations about topics such as assumptions and uncertainties into cookbook labs and found that students treated these superficially, likely because they did not see them as related to the lab [6].

III. SELECTING DATA SOURCES

We analyzed 14 activities, including two from each of seven sources – seven about force and seven about acceleration – with a total length of 154 pages. To find our eight EB activities, we selected four well-known collections of physics activities: (1) the Tutorials in Introductory Physics, (2) the Open Source Tutorials, (3) Workshop Physics, and (4) RealTime Physics [7-11]. All of these collections are available for purchase or free download, all have been mentioned in peer-reviewed journals or conference proceedings, all have been disseminated outside of the institution that developed them, and all have been associated with a high learning gain on the Force Concept Inventory (FCI) or Force and Motion Conceptual Evaluation (FMCE) [5,12-14]. Thus, we feel justified in classifying all four as evidence-based. We also selected six NEB activities from three unpublished collections. All were acknowledged by individuals at their respective institutions to be traditional or “cookbook” laboratory activities. All of the activities are relatively old; two make

heavy use of “fill-in-the-blank” data tables at the end of the lab, and two (not the same two) have been updated or are no longer in use at their institution in the form analyzed in this paper. This is unsurprising, as we primarily contacted institutions with active PER programs; it is possible that at institutions without such programs, traditional labs may still be in use. Regardless, our interest is in analyzing how labs have changed due to PER, so “outdated” traditional labs remain relevant.

IV. CODING ACTIVITIES

To code the activity documents, we first analyzed themes in the literature about document features, and we found that representations were commonly discussed. We selected a coding scheme based on the literature but we also allowed ideas to emerge from the data. In our scheme, each document segment could relate to one or more representations, including Graph, Picture, Quantitative, or Verbal. The Quantitative category includes numbers and equations in any format. It is further broken down into a Number/Equation subcategory for individual numbers or equations and a Table subcategory for tables of numbers or equations. These codes could be used for three purposes:

P1. A representation given in the text; a given number would be coded “Given-Number/Equation,” or “GIV-Num/Eq” for short.

P2. A representation that the student produces in response to a question; e.g. “Student Produced-Number/Equation” or “STP-Num/Eq” for short

P3. A representation that the student produced which is referenced at a later point in the text; e.g. “REFS-Num/Eq,” for “reference to student number.” An example of (P3) would be if the student performs a computation and is later asked to explain in words whether the number is reasonable. Such an instance would be coded as REFS-Num/Eq as well as STP-Verbal.

We did not have a GIV-Verbal code, since almost all document units included at least some text. We coded references to experimental equipment (GIV-Equipment) since it is often the object being represented in the graphs, pictures, etc. Finally, we included codes for a variety of student response types beyond the basic representations listed above. These included a specific instruction to interact with group members or with an instructor as well as students designing an experiment, ranking two or more items, or selecting from multiple well-defined choices.

For both (P2) and (P3), we recorded not only the representations themselves, but also later textual references. For example, we found that the Tutorials in Introductory Physics would often draw a picture of a physical situation, then refer to that picture again and again. In our scheme, GIV-Picture was therefore coded multiple times even though the picture only appeared once. We considered that this was a better characterization of the tutorial’s usage of the picture than simply coding it once when it first appears.

Tutorials and labs were “unitized” or broken into pieces, each of which was assigned one or more codes. (A “unit” is one of these pieces.) We unitized each document according to a few principles. First, sections which provided a blank space for a student response were always unitized, with one codeable unit per blank space. Second, outlined or numbered sections (such as “3” or “II.A”) as well as significant introductory sections were always coded even if no student response was called for.

Because our coding scheme allows any number of codes to be assigned to each data unit, we applied the method of Smith et al. to compute inter-rater reliability (IRR) between two raters [15]. The resulting Cohen’s Kappa value was 0.879, indicating excellent agreement [16]. For IRR, we coded 89 units, or 13.2% of the length of our data, thus surpassing both of Lombard’s suggested criteria of 50 units or 10% of the data [17].

V. STATISTICAL METHODS

To compute the relative frequency of each code, we first divided the number of occurrences of that code for a given activity by the total number of codes for that activity with a given purpose. In other words, since the Tutorials in Introductory Physics force tutorial had 37 “student produces” (STP) codes, of which 7 were pictures, we said that the relative frequency of STP-Picture in that force tutorial was $7/37 = 19\%$. In Table I, we show the relative frequency for each code. The frequencies are shown for four groups of activities: (i) EB tutorials, (ii) EB labs, (iii) all EB activities combined, (iv) NEB activities. There are four EB tutorials represented, four EB labs, and six NEB activities. We have averaged the relative frequencies over all relevant activities. The Wilcoxon Rank-Sum Test (WRST) was applied to determine whether the EB and NEB activities are significantly different.

The WRST is a nonparametric test; its null hypothesis is simply that the fraction of (e.g.) STP-Verbal codes within activities comes from the same probability distribution for EB and NEB activities. Our alternative hypothesis is that the two distributions are not equal [18]. The WRST does not assume any particular probability distribution for the number of codes within documents. This is important because we do not know how activity designers decide how many verbal responses to include in a given document; the decision process may be very complex. Regardless of the distribution, if – as per the null hypothesis – the distribution is the same for all 14 activities, then it follows that any ordering of the 14 activities (ranked in order of % prevalence of a particular code, say STP-Verbal) would be just as likely as any other ordering. Now if it turns out that the six NEB activities all have more STP-Quant codes than all eight EB activities, this particular ordering would be extremely unlikely to happen by chance. Such an event would therefore be considered to be highly significant and would have a small p-value.

We considered a Bonferroni correction: dividing the significance level $\alpha = 0.05$ by the number of tests performed, which is $N = 23$, so we require $p < 0.0022$ because $0.05 / 23 = 0.0022$ [19]. Two of our tests passed this criterion. However, Nakagawa suggests supplementing the Bonferroni correction with effect size, so we also present effect sizes whenever $p < 0.05$ [20].

VI. RESULTS AND DISCUSSION

The null hypothesis was rejected at $p < 0.05$ in five cases: students’ tabular (Table I.2.d) quantitative (Table I.2.e), or verbal response (Table I.2.k); or their referring back to a table (Table I.3.d) or quantitative response they have produced (Table I.3.e). Of these, STP-Quant and STP-Verbal passed with $p < 0.0022$. The relevant effect sizes are all extremely large; for STP-Table, Cohen’s

TABLE I. Relative use of representations given to students (GIV), representations produced by students (STP), or textual references to student-produced representations (REFS) for evidence-based and non-evidence-based activities. *: $p < 0.05$, **: $p < 0.0022$

	EB tutorials	EB labs	All EB	All NEB
<i>1. Representations given to students (GIV)</i>				
a. Graph	0.27	0.05	0.16	0.02
b. Picture	0.36	0.20	0.28	0.14
c. Number / eqn	0.29	0.22	0.26	0.40
d. Table	0.00	0.00	0.00	0.00
e. Quant: c+d	0.29	0.22	0.26	0.40
f. Equipment	0.08	0.53	0.30	0.44
<i>2. Representations produced by student (STP)</i>				
a. Graph	0.10	0.19	0.14	0.11
b. Picture	0.07	0.04	0.05	0.04
c. Number/ eqn.	0.03	0.14	0.09	0.25
d. Table*	0.00	0.04	0.02	0.24
e. Quant: c+d**	0.03	0.18	0.11	0.49
f. Group interact	0.06	0.02	0.04	0.00
g. Instr. interact	0.07	0.00	0.03	0.02
h. Design	0.01	0.01	0.01	0.00
i. Ranking	0.03	0.02	0.03	0.02
j. Mult. Choice	0.20	0.21	0.20	0.17
k. Verbal**	0.42	0.33	0.38	0.16
<i>3. Textual references to student-produced representations (REFS)</i>				
a. Graph	0.17	0.61	0.39	0.28
b. Picture	0.25	0.03	0.14	0.10
c. Number/eqn.	0.25	0.14	0.20	0.23
d. Table*	0.00	0.08	0.04	0.36
e. Quant: c+d*	0.25	0.22	0.23	0.60
f. Verbal	0.33	0.14	0.23	0.02

$d = 2.17$; for STP-Quant, $d = 3.81$; for STP-Verbal, $d = 2.59$, for REFS-Table, $d = 2.11$; and for REFS-Quant, $d = 1.22$.

Quantitative answers, counting numbers, equations, and tables, make up 49% of NEB student responses but only 11% of EB student responses. A verbal response (STP-VER) was used with a frequency of 38% in EB activities, but only 16% in NEB activities. Indeed, *all* eight EB activities used quantitative responses less often than *all* six NEB activities, while the same was nearly true in reverse for verbal responses. (Except that *one* NEB had more verbal responses than *one* EB.) In order to test whether the verbal effect came mostly from short or long verbal responses, we further categorized verbal responses into “one word” responses, where the student could respond adequately with a single word or a standard phrase (“kinetic energy”), and “long” responses, where multiple words were necessary. We found that long responses were more common overall, but the ratio of long to short responses did not vary substantially between EB and NEB activities.

Prior research suggests that non-mathematical representation use is critical for problem solving success, suggesting that instructor emphasis on non-mathematical representations is important [21].

We have shown that students’ verbal responses have been a key feature of many successful EB activities to date. In particular, our EB activities used more than three times as many verbal responses as quantitative responses, while our NEB activities used three times as many quantitative responses as verbal responses. Our interpretation is that EB activity designers recognize that students cannot gain conceptual knowledge simply by recording, computing, or reporting numerical values. It is important for students to

explain what they are doing and why they are doing it. We suggest that the ratio of verbal to quantitative responses (“V / Q”) is a useful heuristic for categorizing activity documents. These are the two variables that met the strict standard of the Bonferroni test. For our EB tutorials, this ratio was 13 (dividing the total V by the total Q); for EB labs, 1.83, and for NEB labs, 0.33. In fact, each individual V / Q for the eight EB activities was greater than one, while each V / Q for the six NEB activities was less than one. Note that this does not show that a high V / Q ratio is sufficient to produce learning, only that it has been a common feature of many successful activities.

VII. FUTURE WORK

We have developed and applied our framework to published and disseminated evidenced-based activities as well as traditional non-evidence-based labs. However, we expect to see a spectrum of document types in between these possibilities, such as activities designed by faculty who are knowledgeable about PER but which may not be designed as rigorously as those appearing in peer-reviewed publications. We are interested in finding the V/Q ratio of such documents in order to ascertain whether the ratio is consistently correlated with PER-influenced documents.

ACKNOWLEDGEMENTS

We would like to acknowledge Brian Thoms for his help with this work. This work was funded by the National Science Foundation (Grant No. DUE 1347510).

- [1] R. R. Hake, *Am. J. Phys.* **66**, 64-74 (1998).
- [2] P. B. Kohl and N. D. Finkelstein, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010102 (2006).
- [3] D. E. Meltzer, *Am. J. Phys.* **73**, 463-478 (2005).
- [4] J. H. Larkin, *The Role of Problem Representation in Physics* (Carnegie-Mellon University, Department of Psychology, Pittsburgh, 1981).
- [5] B. Royuk and D. W. Brooks, *J. Sci. Educ. Technol.* **12**, 317-324 (2003).
- [6] A. Karelina and E. Etkina in *AIP Conf. Proc.* **883**, 93-96 (2007).
- [7] L. C. McDermott and P. S. Shaffer, *Tutorials in Introductory Physics* (Prentice Hall, Upper Saddle River, NJ, 2001).
- [8] A. Elby, R. E. Scherr, T. McCaskey, R. Hodges, E. F. Redish, D. Hammer, and T. Bing, *Open Source Tutorials in Physics Sensemaking: Suite I* (2007).
- [9] R. E. Scherr, A. Elby, R. M. Goertzen, and L. D. Conlin, *Open Source Tutorials in Physics Sensemaking: Suite II* (2010).
- [10] P. Laws, *Workshop Physics Activity Guide, The Core Volume with Module 1: Mechanics I*, (Wiley, Hoboken, NJ, 2004).
- [11] D. R. Sokoloff, R. K. Thornton, and P. W. Laws, *RealTime Physics Active Learning Laboratories: Module 1: Mechanics*, (John Wiley & Sons, Hoboken, NJ, 1999).
- [12] G. Duda and K. Garrett, *AIP Conf. Proc.* **1064**, 111-114 (2008).
- [13] K. M. Koenig and R. J. Endorf, *AIP Conf. Proc.* **720**, 161-164 (2004).
- [14] R. M. Goertzen, E. Brewster, L. H. Kramer, L. Wells, and D. Jones, *Phys. Rev. ST Phys. Educ. Res.* **7**, 020105 (2011).
- [15] M. K. Smith, F. H. M. Jones, S. L. Gilbert, and C. E. Wieman, *CBE Life Sci. Educ.* **12**, 618-627 (2013).
- [16] J. R. Landis and G. G. Koch, *Biometrics* **33**, 159-174 (1977).
- [17] M. Lombard, J. Snyder-Duch, and C. C. Bracken, <http://matthewlombard.com/reliability/> (2010).
- [18] M. P. Fay and M. A. Proschan, *Statist. Surv.* **4**, 1-39 (2010).
- [19] R. J. Cabin and R. J. Mitchell, *Bull. Ecol. Soc. Am.* **81**, 246-248, (2000).
- [20] S. Nakagawa, *Behav. Ecol.* **15**, 1044-1045 (2004).
- [21] C. J. De Leone and E. Gire, *AIP Conf. Proc.* **818**, 45-48, (2006).