Uptake of solution checks by undergraduate physics students

Tiffany-Rose Sikorski,¹ Gary D. White,² and Justin Landay²

¹Department of Curriculum and Pedagogy, The George Washington University, 2134 G. Street NW, Washington, DC, 20052

²Department of Physics, The George Washington University, 707 22nd Street NW, Washington, DC, 20052

A persistent concern within physics education is students' apparent failure to check the reasonableness of their answers. In an effort to better understand how students' capacity for checking solutions develops, this paper examines data on solution checking in an upper-level undergraduate electricity and magnetism course. All students demonstrated the ability to check answers in multiple ways, but showed variability in how they chose to do so, with checking units the most easily activated check, and numerical values strikingly underutilized.

I. INTRODUCTION

Checking solutions is valued among physics educators as an aspect of problem-solving expertise. Solution checking has been examined as a differentiator between expert and novice performance [1], as an indicator or component of metacognition [2], and as a mechanism for developing new knowledge and understanding [3].

Multiple studies have documented students' apparent failure to check their solutions, even when following problem-solving protocols that contain answer-checking as an explicit step. For example, while interviewing students from an Electricity and Magnetism (E&M) course where the ACER framework was introduced, Wilcox *et al.* [4] found only 8% of students attempted to check their solution as indicated in the *reflection* ("R") step of ACER. Some students examined limiting cases, but most simply "made superficial statements about whether the solution looked familiar" [4].

Targeted instruction can encourage students to examine the reasonableness of their solutions. Chasteen *et al.* [5] report that students in PER-aligned, junior-level E&M courses were better able to describe limiting behavior for their solutions than students in standard courses. Similarly, Warren [6] found that explicit instruction increased students' use of unit analysis and special-case analysis on a variety of problem types.

Extending this line of research, we introduced junior-level students in an E&M course to three solution-checking activities to determine whether: (1) a proposed solution to a physics problem has appropriate units ("check units"), (2) limiting cases of the solution match the student's physical intuition ("check limits"), and (3) numerical values are consistent with the student's previous experience or well-known constants ("check values"). We aim to understand: (1) In what ways will students employ the checks?, (2) Which check(s) are most challenging for students?, and (3) Could the checks serve as productive resources for further development of problem-solving expertise?

II. STUDY DESIGN

Complementing the large, quasi-experimental design work in Warren *et al.* [6], we present a small, exploratory investigation of solution-checking behaviors in a first-semester, junior-

level E&M course for which the second author and third author are lead instructor and Learning Assistant, respectively.

Participants were 12 students (11 male, 1 female). Nine (9) of the students had previously taken a course with the lead instructor and thus had some prior exposure to the checks. By focusing on a single classroom and small number of students, we aim to document small but potentially meaningful variations in how students assess the reasonableness of solutions.

Throughout the course, the instructor emphasized checking solutions (see Section III), and student use of the checks was assessed; this study focuses on three parallel in-class exercises collected on Day 1, Day 8, and Day 20 (see Section IV). Students also completed a survey at the end of the course describing their perspectives on the checks.

III. INTRODUCING THE CHECKS

Physics instructors are likely familiar with the three checks, and may introduce their own students to similar checks. Thus, in this section, we describe how the checks were introduced in the focal classroom, in part to more clearly define the checks in our context, and also to further motivate our interest in studying how students take up the checks over time.

The three checks were introduced on the first day of class and reinforced throughout the semester via homework, in class, and on exams. For example, as a pre-assessment on the first day of class, students were asked to examine a possible expression for the tension in a string for the symmetric suspended two charged ball problem (see Eq. 1).

$$T = mg\sqrt{1 + \left(\frac{kq^2}{4mgL^2}\right)^{\frac{2}{3}}} \quad (1)$$

Students received specific instructions for each check: (1) "Show that this formula has the right units to be a tension, considering the terms in both the numerator and denominator; (2) Show that this formula makes sense in the limit as the charge goes to zero, explaining why the mathematical answer matches your intuition about what the tension should be when there is no charge; (3) Calculate a numerical value for the tension in the case that the mass is 1 gram, the string is 1 m, and

the charge is 10 nC and comment on whether the numerical value you get is plausible by comparing to some well-known force."

In the following days, many opportunities arose, both graded and ungraded, in class and on homework, in which students were asked to check their answers in the "usual three ways." The word "usual" is chosen deliberately, partly to convey that this is common practice among physicists, and the suggestion, explicitly mentioned by the instructor, is that practicing physicists do this even when no one is looking, because it is a way that one learns physics [3]. For example, seeing that in the limit $q \rightarrow 0$ that the tension tends toward the weight of the ball not only provides confidence in the expression since it matches one's intuition, but the formula given in Eq. 1 also suggests that the next leading order term for small charges goes as something that would be difficult to predict otherwise. The instructor explicitly details these kinds of discoveries, highlighting what was learned by doing the check, and commenting on the weirdly satisfying feeling that one gets when, for example, a limiting case works out to match one's physical intuition as expected. While these are some of the key ways that checks are introduced in the course, we cannot assume that the students attribute the instructor's intended meaning to the checks; hence, our interest in studying students' use of the checks over time.

IV. ASSESSING USE OF THE CHECKS

Anecdotally, class discussions and homework assignments indicated that students adopted the checks. In order to develop a more systematic source of evidence, multi-layer prompts were administered to all students as a pre-assessment on Day 1, and follow-up parallel exercises on Day 8 and Day 20 of the class (the class met for ~ 30 sessions).

Each prompt contained two layers, with students turning in Layer 1 before seeing Layer 2. In Layer 1, students were offered 4 different algebraic formulae as solutions to a relevant problem and asked to select the most "plausible," explaining their reasoning. In Layer 2, students were offered one purported solution to the same problem and asked to "Check to see if this formula is sensible in as many ways as you can think of, explaining your thinking clearly." Layer 2 was selected for coding because it was most consistent layer across days, and seemed to provide the most informative data on student use of the checks.

Problems were selected such that solutions were relatively difficult, perhaps unsolvable by the typical student in the allotted time, but within reach conceptually. Problems varied to align with material students were examining in class; as a result, the solutions that students were asked to check varied in mathematical and conceptual complexity across days.

The three authors reviewed Layer 2 responses, tracking whether a specific check was attempted (0 or 1; Table I) and how the check was implemented. Disagreements about coding arose when one or more authors failed to notice a check

that was not clearly marked, and when responses contained major algebraic or numerical errors that made it difficult to tell which check was attempted. Still, consensus was reached in all cases.

To get a sense of what the students thought about the three checks, we also report on a brief survey that they completed on the penultimate day of the course, a survey framed as feedback to the instructor on the "merits and detriments of checking your answers as implemented in class."

V. FINDINGS

Students checked units most often, followed by limits, and then reasonable values. All students except 2, 5, and 10 implemented each check at least once (Table I). Student 2 did not submit the Day 1 task; Student 5 withdrew before Day 8.

TABLE I. Students' use of the three checks over time.

Student	Units			Limits			Values		
	Day 1	Day 8	Day 20	Day 1	Day 8	Day 20	Day 1	Day 8	Day 20
1*	1	1	1	0	0	1	0	0	1
2	-	1	1	-	1	1	-	0	0
3	1	1	1	0	1	1	0	1	0
4	0	1	1	0	1	1	0	1	1
5*	0	-	-	0	-	-	0	-	-
6	1	1	1	1	1	1	1	1	0
7	0	1	1	0	1	1	0	1	1
8	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	0	1	1
10*	0	0	1	0	1	1	0	0	0
11	1	1	-	1	1	-	0	1	-
12	1	1	1	1	1	0	0	1	0
Total	7	10	10	5	10	9	2	8	5

*Student who had *not* previously taken a course with the instructor.

A. Student use of "checking units"

Consistent with Warren's study, we found that unit checks were the easiest check for students; in fact, most of our students employed a version of this check on Day 1 in Layer 2 described above. Occasionally, the unit check was incomplete (compare Fig. 1 to Fig. 2) or the students attempted dimensional analysis but substituted incorrect units, but both of these kinds of responses, along with fully successful unit checks, are indicated by a "1" in Table I; a "0" means that there was no evidence that a unit check was attempted. Warren [6], using a more nuanced rubric with scoring from 0 to 3, observed over time a decrease in the use of unit checks on critical thinking tasks in his experimental group, from 89% to 53% to 38%, a decline attributed to the decision to spend less time on unit analysis after observing students' early success. In contrast, in the present study all three checks were

emphasized throughout the semester, and no similar decline was evident. On the survey, one student noted, "Units are now my primary method of doing a solution check."

Check to see if this formula is sensible in as many ways as you can think of, explaining your thinking clearly.

1) Units: $T(\text{force}) = m(kg) \cdot g(m/s^2) \rightarrow N = N \checkmark$

FIG. 1. Unit check *without* explicit verification of terms (Day 1).

$$= kq \cdot \frac{m}{a} \sqrt{1 + \left(\frac{N \cdot g^2}{4mg^2} \right)^{2/3}}$$

$$= kq \frac{m}{a} \sqrt{1 + \left(\frac{N \cdot g^2}{4mg^2} \right)^{2/3}}$$

$$= kq \frac{m}{a} \sqrt{1 + \left(\frac{1}{4} \right)^{2/3}}$$

$$= \text{Mass} \cdot \text{acceleration}$$

$$= \text{Force}$$

$$= \text{acceleration} \checkmark$$

FIG. 2. Unit check *with* explicit verification of terms (Day 1).

B. Student use of "checking limits"

Across the three exercises, we observed multiple kinds of limiting case checks. Some students commented on the resemblance of the solution to a known equation or mathematical form, for example the Pythagorean Theorem. Most students presented a mathematical form of a limit accompanied by a statement such as "limit goes to zero as expected" or a " $\checkmark$ ", but without providing a physical basis for the expectation (Fig. 3). A few students showed evidence of a mathematical limit *and* invoking physical intuition to argue that the limited behavior is expected (Fig. 4).

Limiting cases

$$T = mg \sqrt{1 + \left(\frac{kq^2}{4mg^2} \right)^{2/3}}$$

if $m \rightarrow \infty, T \rightarrow \infty$ & if $g \rightarrow 0, T \rightarrow 0 \checkmark$

if $q \rightarrow \infty, T \rightarrow \infty$ & if $L \rightarrow 0, T \rightarrow mg$

if $L \rightarrow \infty, T \rightarrow mg$

the limiting cases make sense

FIG. 3. Limits *without* explicit link to physical intuition (Day 1).

All eleven students that completed the course attempted limiting case analysis of some kind in Layer 2 as the semester progressed (see Table I), again consistent with Warren's [6] record of the increasing student use of special case analysis in response to explicit instruction. We were especially interested in what the students cited as the source of their expectation for what the answer should be. In surveys, students said they determined "expected behavior" via mathematical forms (e.g., "you should see if it behaves like some of the known behaviors of that type of equations. Such as the electric field of a point charge approaches $\frac{1}{r^2}$ "), everyday logic (e.g., "If you

Limit to infinity

$a = \frac{kq^2 a p_0^2}{m(1+2g)} - g$ as $L \rightarrow \infty$ it begins to not feel reasonable by the look but only goes down because of gravity

FIG. 4. Limits *with* explicit link to physical intuition (Day 8).

have an equation for how well a student might do in class, and the equation depends on hours spent studying, it would not make sense for performance to go to ∞ as hours goes to zero"), or general "intuition." One student wrote about using limiting cases *to develop a sense of* expected behavior ("understanding how limits are approached helps to visualize the physical properties or behavior for a problem.")

C. Student use of "checking reasonable values"

Students seemed to have the most difficulty checking reasonable values. When students did attempt the numerical values check, they most often plugged in numbers and made a generic statement of reasonableness without explicit comparison to well-known results or to classroom experience (Fig. 5). Less often, students plugged in numbers and showed some concrete evidence that the numerical values chosen were rooted in some physics experience and/or that the final numerical value is sensible compared to a known value (Fig. 6). In surveys, students wrote about their difficulties in developing an intuition for what values are reasonable, (e.g. "the reasonable answer check requires previous research which isn't available on tests/exams. This is the only check I don't always subscribe to and sometimes leave out."; "Sometimes the reasonable values check doesn't offer much insight because there are such wide range of charge, field [and] voltage values that could be considered 'reasonable' I could be off by a factor of 100 and it wouldn't sound any alarms.")

The former comment is somewhat worrisome because the student does not seem to view the check as an opportunity to build a sense of what values might be reasonable. However, the latter comment about reasonable values for voltage could indicate some developing sophistication—recognizing that different checks can yield more or less insight for particular problems.

possible values $kq^2 = 10^{-6}$

$$\frac{(10^{-9})(1)(10^{-6})(10^{-6})}{(10^{-3})(10)^2} = 10 = \frac{10^{-3}}{10^{-1}} = 10^{-2} \text{ m/s}^2 = 1 \text{ cm/s}^2$$

FIG. 5. No explicit comparison to a known value (Day 8).

D. Affordances and limitations of the checks for developing problem-solving expertise

The three checks were introduced as a *starting point* for more sophisticated practices of solution checking. Over time,

3) Sensible values $k = 9 \times 10^9$ $q = 1 \times 10^{-9}$ $q_0 = 1 \times 10^{-9} \text{ C/m}^3$ $R = 1 \text{ m}$
 $m = 1 \text{ kg}$ $L = 1 \text{ m}$ $\rho = 10^3 \text{ kg/m}^3$
 $a = \frac{(9 \times 10^9)(1 \times 10^{-9})(1 \times 10^{-9})(1 \times 10^3)}{(1)(1)(1)(1)} = 10^{-6} = \frac{9 \times 10^{-6}}{2} = 10^{-6}$
 This seems to make sense since a small 1 nC charge would not be likely to have much effect on the acceleration.

FIG. 6. Explicit comparison to a known value (Day 8).

we hoped students' use and the meaning ascribed to the checks would become more sophisticated as they tried the checks in new contexts and gained insights from implementing them. Thus, our third research question considers the extent to which this starting point is helpful or productive.

Evidence that the three checks were supporting students' development of problem-solving expertise might include: more sophisticated implementation of a check over time, the development of physical intuition from implementing the checks, acting on the information gleaned from the checks to develop new solution paths, or exploring the insights that particular checks might provide for different kinds of problems [7]. Here, we were limited by our prompts; students were not required to "act on" the information gleaned from the checks, so we were only able to track how students' use of the checks changed across the three prompts.

Perhaps counter to our hopes, we observed "script-like" implementation of the checks. For example, though not instructed to do so, students labeled their checks (e.g., Fig. 2, 3, 6) and most conducted the checks in a fixed order (units, limiting cases, reasonable values). We also observed a decrease in "other" kinds of checks that students used; even when asked to check solutions "in as many ways as they could think of", students checked for the three ways. The script-like implementation stands in contrast to more fluid application that we might expect of expert physicists, and hope to develop in our students.

Despite concerns that we might be reducing the epistemologically rich practice of solution checking to three "rote" activities, there is also evidence of some more desirable evolution. As an example, consider Student 3, whose work is displayed in Figs. 2 and 4. On Day 1, the only mode of checking that he exhibits is a fairly regimented unit check. But, by Day 8, he utilizes the limiting case check thoroughly with an explicit remark about how the mathematical form matches the

physical behavior expected (not to mention that he invokes the other two checks as well, see Table 1). Again on Day 20, Student 3 explores multiple limiting case scenarios, with various and extended commentary about whether the proposed solution matches his expectations, suggesting the student is developing some sophistication with the limiting cases check.

In the written survey, Student 3 mentions that the "units check" is an "extremely simple way to verify your solution" and that it is his "primary method of doing a solution check." He also mentions that "limits are definitely beginning to make more sense but knowing how something approaches a limit is still a little difficult" and that "it helps differentiate between different problems" and "understanding how the limits are approached helps to visualize the physical properties or behavior." These comments suggest Student 3's use of the limiting case check is more than rote procedure; he uses the check for sensemaking. Finally, on the final exam, in response to this prompt: "Derive and defend a formula for the instantaneous acceleration of the little sphere" Student 3 wrote, after completing a derivation, "checked for units and assuming q was computed correctly, am missing units of meters. will come back if time." Here, Student 3 shows an inclination to act on the information gleaned from a check.

In conclusion, we again note that previous work has lamented students' apparent failure to check their solutions, even when specifically prompted to do so. Operating under the assumption that professional physicists check solutions for sensibility, and that all students are able to meaningfully engage in this practice under supportive conditions, the aim of this study was to determine if the three checks could be taken up as resources for problem-solving in a junior-level E&M course. Following this very small n investigation, we are cautiously optimistic. When asked to check the sensibility of a solution, most students are able to implement at least one of the checks, even for difficult E&M problems for which they are unable to generate complete solutions. In future work, we plan to interview students to better understand if and how they utilize solution checks while problem solving.

ACKNOWLEDGMENTS

The authors thank three anonymous reviewers, Mark Eichenlaub, Paul Hutchison, Erik Kuo, and Joe Redish for their comments.

- [1] Simon, D.P. and Simon, H.A. In *Children's Thinking: What Develops?*, Edited by R. S. Siegler (Lawrence Erlbaum Associates, Hillsdale, NJ, 1978).
- [2] A. H. Schoenfeld, *Cognitive Science and Mathematics Education* (Lawrence Erlbaum, Hillsdale, NJ, 1987), Chap. 8.
- [3] E. F. Redish and E. Kuo, *Sci. and Ed.* **24**, 5 (2015).
- [4] Wilcox, B. R., Caballero, M. D., Pepper, R. E., and Pollock, S. J. In *AIP Conference Proceedings, 2013*, edited by P.V. Engelhardt, A.D. Churukian, and N. S. Rebello, (AIP, 2013), 418.
- [5] Chasteen, S. V., Pollock, S. J., Pepper, R. E., and Perkins, K. K. *Amer. J. Phys.* **80** 10 (2012).
- [6] A. Warren. *Phys. Rev. ST Phys. Educ. Res.* **6** 020103 (2010).
- [7] J. Bolton and S. Ross. *Phys. Educ.* **32** (1997).