

# Investigating student ability to follow and interact with reasoning chains

William N. Ferm Jr.<sup>1,2</sup>, J. Caleb Speirs<sup>1</sup>, MacKenzie R. Stetzer<sup>1,2</sup>, and Beth A. Lindsey<sup>3</sup>

<sup>1</sup>*Department of Physics and Astronomy, University of Maine, 120 Bennett Hall, Orono, ME, 04469-5709*

<sup>2</sup>*Maine Center for Research in STEM Education, University of Maine, 5725 Estabrooke Hall, Orono, ME, 04469-5727*

<sup>3</sup>*Physics, Penn State Greater Allegheny, 4000 University Drive, McKeesport, PA 15132*

The effectiveness of scaffolded, research-based instruction in physics has been extensively documented in the literature. However, much less is known about the development of students' reasoning skills in these research-based instructional environments. As part of a larger collaborative project, we have been designing and implementing tasks to assess the extent to which introductory physics students are able to logically follow and interact with hypothetical student reasoning chains in a variety of physics contexts. In this paper, we report preliminary results from a "Follow Reasoning" task in which students are asked to infer the conclusions that would be drawn from different lines of reasoning articulated by hypothetical students and provide justification for that inference.

## I. INTRODUCTION

For more than 30 years, the PER community has focused considerable attention on the development of student conceptual understanding, especially at the introductory level [1]. Drawing upon findings from investigations of the learning and teaching of specific physics concepts, numerous research-based and validated assessment instruments have been developed to evaluate student conceptual understanding (e.g., the Force Concept Inventory [2]). These investigations have also informed the development and refinement of research-based and research-validated instructional materials (e.g., *Tutorials in Introductory Physics* [3]); such materials have been shown to be very effective in improving student conceptual understanding (e.g., [4]).

However, even after research-based instruction, students who demonstrate a solid conceptual understanding on one physics task may subsequently perform poorly on another closely related task requiring application of that same knowledge [5]. Research on such inconsistencies has begun to suggest that poor performance on certain tasks may primarily be attributed to reasoning difficulties rather than conceptual difficulties [6].

Dual-process theories of reasoning (e.g., [7]) have been shown to be useful in accounting for, in a mechanistic fashion, inconsistent student performance across related tasks [8]. These findings suggest that research-based efforts explicitly designed to support and enhance student reasoning skills may be needed to further improve student performance in introductory physics and beyond.

Physics courses are generally believed to develop student reasoning and thinking skills [1], and such skills are expected of students in many Science, Technology, Engineering, and Math (STEM) related fields (e.g., [1]). To date, however, while there has been research on formal scientific reasoning (e.g., [9]), there has been little work on the distinct set of skills associated with qualitative inferential reasoning, which is ubiquitous in most research-

based instructional materials (e.g., [3]). Indeed, the extent to which scaffolded, research-based physics instruction impacts student ability to engage in such reasoning remains largely unknown. Additionally, there are currently no widely available instruments that can be used to assess these particular reasoning skills.

For this reason, we are collaborating with researchers at three other institutions (North Dakota State University, University of Washington, and Western Washington University) in an effort to develop instruments and methodologies capable of measuring targeted student reasoning skills. In particular, our collaboration has focused on designing and piloting tasks that may be used to document the extent to which students are able to follow, replicate, evaluate, and generate coherent chains of reasoning before, during, and after scaffolded, research-based instruction.

While a variety of novel tasks are currently under development (see, for example, [10]), this paper describes the development and implementation of a "Follow Reasoning" (FR) task, which was designed to assess the extent to which students in an introductory calculus-based physics course are able to follow provided reasoning chains. While this task has been implemented in several different physics contexts, we primarily focus our discussion on one particular FR task involving the application of Newton's 2<sup>nd</sup> law.

## II. RESEARCH CONTEXT

This study was conducted in an introductory calculus-based mechanics course offered at the University of Maine. There were 284 students enrolled in this course, and the majority of students were pursuing degrees in STEM fields. The course consisted of lecture, recitation, and laboratory components, with approximately two hours per week dedicated to each component. Typically, *Tutorials in Introductory Physics* were implemented in one hour of recitation each week, with graduate TAs and undergraduate LAs serving as instructors. All data were collected from

online exam review assignments, administered after all relevant content instruction. Students received extra credit for completing these optional assignments.

### III. OVERVIEW: FOLLOW REASONING (FR) TASK ON NEWTON'S SECOND LAW

In a Follow Reasoning (FR) task, students are: (1) presented with a physics question to consider, which they are not asked to solve; (2) prompted to predict the conclusions that follow from provided hypothetical student reasoning chains (HSRCs); and (3) asked to justify their predictions. In this particular FR task, students were provided with a physics question, shown in Figure 1, in which two blocks, A and B, are pushed to the right at constant speed. Block A has twice the mass of block B. The question asks whether the net force on block A is *greater than, less than, or equal to* that on block B. Students were then shown, one at a time, four HSRCs in response to the original physics question. The HSRCs provided were incomplete in that the final concluding statement of each reasoning chain was omitted. Students were subsequently prompted, for each HSRC, to: (1) select the appropriate concluding statement, thereby completing the chain of reasoning provided, and (2) explain why they think the HSRC would lead to that particular net force comparison.

Since both blocks are moving at constant speed in the physics question, the magnitudes of both accelerations are equal to zero. Thus, from Newton's second law, both net forces are equal to zero, and therefore the correct comparison is that the net force on A is equal to that on B.

Although each HSRC provides enough information so that it is possible for a student to predict the omitted conclusion, not all conclusions correspond to correct net force comparisons. (See Table 1.) Collectively, the conclusions associated with all four HSRCs span all possible net force comparisons (*i.e.*,  $F_{(net)A} > F_{(net)B}$ ,  $F_{(net)A} < F_{(net)B}$ , and  $F_{(net)A} = F_{(net)B}$ ).

The HSRCs also represent four different approaches to answering the physics question. While S1 applies Newton's

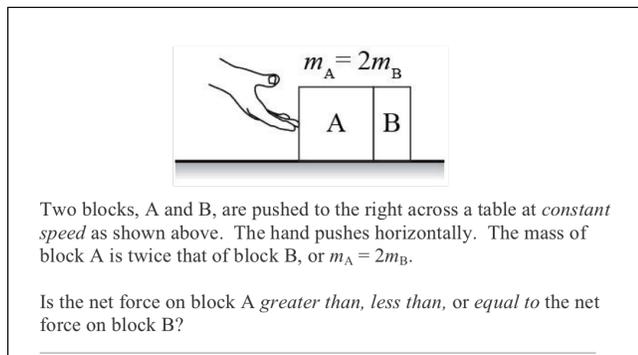


FIG 1. Physics question on Newton's second law, which served as the context for the Follow Reasoning (FR) task.

<p style="text-align: center;">S1</p> <p>[1] The acceleration of A and B must be the same because the blocks move together, or <math>a_A = a_B</math></p> <p>[2] <math>F_{(net)A} = m_A a_A</math> and <math>F_{(net)B} = m_B a_B</math></p> <p>[3] and <math>m_A &gt; m_B</math></p> <p>[4] so ...</p>	<p style="text-align: center;">S2</p> <p>[1] <math>m_B = m_A/2</math></p> <p>[2] Objects with half the mass have twice the acceleration</p> <p>[3] so <math>a_B = 2a_A</math></p> <p>[4] <math>F_{(net)} = ma</math> implies that ...</p>
<p style="text-align: center;">S3</p> <p>[1] A and B are moving at constant speed (<math>\Delta v = 0</math>)</p> <p>[2] so <math>a_A = 0</math> and <math>a_B = 0</math></p> <p>[3] and therefore...</p>	<p style="text-align: center;">S4</p> <p>[1] Smaller masses accelerate more than bigger masses</p> <p>[2] so <math>a_A &lt; a_B</math></p> <p>[3] and <math>F_{(net)}</math> is proportional to <math>a</math></p> <p>[4] so ...</p>

FIG 2. The four hypothetical student reasoning chains (HSRCs) that were provided to students in the FR task.

second law to the two block system and all statements made are generally true, the observation that both blocks are moving with constant speed is neglected. S2 uses circular reasoning to arrive at a correct net force comparison (although S2 does not recognize that these net forces are both equal to zero). Indeed, the second statement made by S2 is true if and only if the net forces are assumed to be equal and non-zero, an assumption that ultimately corresponds to the final net force conclusion associated with S2's reasoning chain. S3 arrives at the correct net force comparison using an appropriate reasoning chain; there is nothing incorrect about the chain, although we do not claim that it is an ideal and complete response. Finally, the reasoning provided by S4 is internally inconsistent. The first statement by S4 relies on the assumption that the net forces on the two blocks are equal (much like statement 2 given by S2). Later in S4's reasoning however, the final net force conclusion that stems from line 3 contradicts the implicit assumption of statement 1.

### IV. DATA ANALYSIS

All student responses were coded based on the predictions made for each HSRC. Student performance was measured by the percentage of students making the correct prediction for each student (*i.e.*, the logical conclusion following from each HSRC).

Upon repeated examinations of students' prediction justifications, four independent coding categories emerged from the data and were used to analyze those justifications. These categories are described in detail below.

**Engagement with and critiquing of the provided reasoning.** Engagement corresponded to synthesis, clarification, or extrapolation of the HSRC beyond a direct restatement (*i.e.*, light paraphrasing or copying and pasting) of that reasoning. The following student response illustrates engagement with the HSRC: "Since  $a_A = a_B$  and  $m_A > m_B$ ,  $F_{(net)A}$  needs to be greater than  $F_{(net)B}$  since there are no other variables that affect the equation." Critiquing was defined as identifying aspects of the HSRC that made the

reasoning adequate or flawed, or characterizing the overall approach used in the HSRC in addition to also engaging with the HSRC (e.g., “The student [S1] will arrive at the conclusion selected because they failed to account for the fact that the acceleration of the blocks is equal to zero. This is a special case in which the net forces are equal.”).

**Commentary regarding the correctness of the provided reasoning.** *Commentary on correctness* codes were assigned to students who explicitly agreed or disagreed with the provided reasoning (e.g., “Because the student [S1] is assuming correctly that acceleration of both blocks is the same.”).

**Type of physics arguments used in justification.** In their justifications, some students included no explicit physics arguments (*no physics*), whereas other students either solely articulated physics arguments found in the associated HSRC (*HSRC physics*) or drew upon at least one physics argument (which may or may not be correct) not found in the associated HSRC (*outside physics*). Students coded as using outside physics may also have included HSRC physics in their responses. The following is an example of a student response (in reference to S1) coded as *HSRC physics*: “Because if  $m_A$  is larger than  $m_B$  if they are both multiplied by the same thing the one multiplied by  $m_A$  will be larger.”

**Explicit reference to the hypothetical student.** This was a binary code in that students either did or did not explicitly reference the hypothetical student (e.g., “They [S1] said  $m_A > m_B$ .” vs. “Because the blocks are moving at a constant velocity with no acceleration.”).

## V. RESULTS

Student performance results are shown in Table 1. Over three quarters of the students were successful at predicting the logical conclusions associated with the reasoning provided by S1, S3, and S4 (See Table 1.) Student prediction performance for S2, however, barely surpassed 50%.

Upon analyzing the student justifications, we found that 81% of students engaged with the reasoning of at least one hypothetical student, and 13% of students critiqued at least one HSRC. In addition, 15% of students explicitly indicated their agreement or disagreement with at least one HSRC. All students articulated some kind of physics argument at least once on this task. Approximately 9% of

**TABLE 1.** Student prediction performance on Newton’s 2<sup>nd</sup> law FR task. N = 144.

Hypothetical Student	Correct Predictions	Predicted Correctly
S1 – Reasoning that omits $a = 0$	$F_{(net)A} > F_{(net)B}$	76%
S2 – Circular Reasoning	$F_{(net)A} = F_{(net)B}$	54%
S3 – Correct Reasoning	$F_{(net)A} = F_{(net)B}$	88%
S4 – Inconsistent Reasoning	$F_{(net)A} < F_{(net)B}$	83%

students only articulated HSRC physics arguments when justifying all four predictions. Consequently, 72% of students used outside physics arguments in support of at least one justification. Finally, 69% of students explicitly referenced the hypothetical student when justifying at least one prediction.

When examining the relationships between prediction performance and the nature of the associated justifications, four general trends were of note. (A more rigorous statistical analysis of these data is currently underway.) First, not only did the overwhelming majority of students engage with at least one HSRC, but *every* student who predicted correctly engaged with the associated HSRC. Second, few students critiqued or commented on the correctness of the HSRC (13% critiqued at least one HSRC; 15% commented on the correctness of at least one HSRC), so it was not possible to identify a relationship between student performance and critiquing or commenting on the correctness of the HSRCs. Third, no relationship was apparent when looking at the type of physics argument used by students when justifying their prediction; in other words, prediction performance did not appear to depend on whether or not students used outside physics (either correct or incorrect) or limited themselves to HSRC physics arguments. Finally, no clear relationship appeared to exist between making an explicit reference to the hypothetical student and prediction performance.

## VI. DISCUSSION

With the exception of S2’s reasoning chain, students appeared to quite successful at following the HSRCs in the Newton’s second law context. To investigate possible sources of difficulty associated with S2’s reasoning chain (e.g., the circular nature of the reasoning), we carefully examined the distribution of prediction responses for S2 as well as the associated prediction justifications. Students who predicted incorrectly were split evenly between the two remaining conclusion options ( $F_{(net)A} > F_{(net)B}$ , 24%;  $F_{(net)A} < F_{(net)B}$ , 22%). In their justifications, many students appeared to display some mathematical comprehension issues (e.g., interpreting the relationship  $m_B = m_A/2$  as saying that  $m_B$  is twice as large as  $m_A$ , consistent with the algebra “reversal error” reported in the literature (e.g., [11])). Furthermore, some students who predicted incorrectly appeared to focus their attention either on lines 1 and 4 (thereby basing the net force comparison solely on the mass comparison), or on lines 3 and 4 (thereby basing the net force comparison solely on the acceleration comparison). These two phenomena, in which the students implicitly and inappropriately held one variable in a multi-variable algebraic expression constant, are consistent with the findings and discussions reported in [12]. Therefore, we hypothesize that student difficulties with S2’s reasoning more likely stemmed from the inherent mathematical structure of S2’s reasoning chain (which was more

quantitative than the other three chains) than from the circular nature of that reasoning.

Our findings also indicate that all students who made correct predictions demonstrated engagement with the associated HSRCs in their justifications. This is somewhat of an expected result in that students have a greater chance of following reasoning when they examine and process that reasoning in a non-superficial way. Though very few students critiqued or commented on the correctness of the HSRCs, they were not prompted to do so, which suggests that students do not appear particularly inclined to evaluate someone else's reasoning spontaneously or without prompting. In essence, this FR task reflects student *tendencies* regarding evaluating reasoning and not necessarily the *abilities* of students to do so. No relationship was found between the inclusion of explicit references to the hypothetical student and prediction performance. This finding provides us with preliminary evidence suggesting that distancing oneself from an HSRC is neither beneficial nor detrimental when following the associated HSRC.

## VII. RELATED FR TASK INVESTIGATIONS

While this paper has focused on an FR task related to Newton's second law, FR tasks focusing on Lenz' law, the work-energy theorem, and wave-pulse superposition have also been administered. In addition, we have explored the use of more authentic student responses, including a mixture of verbatim quotes and lightly modified student work. Across all tasks, we have observed a range of prediction performance (from 33% to 94%), though typically about 2/3 of students are able to predict correctly.

We were also interested in investigating the impact of student conceptual understanding on prediction performance. On an online review assignment for the final exam, students at UMaine were first asked to answer for themselves a question on the work-energy theorem (from [13]) before completing the associated FR task. The results

suggested the absence of a relationship between students' performance on the question itself and students' FR prediction performance. When administered at a different institution immediately after relevant instruction, the results differed considerably, suggesting a possible relationship. Given the mixed results, more work is ongoing in order to account for the observed discrepancy and to determine whether or not there is a relationship between student conceptual understanding and success on FR tasks.

## VIII. CONCLUSIONS

In summary, we have developed a task in which students were asked to follow the reasoning of hypothetical students, to use that reasoning to make predictions, and to justify these predictions. Performance overall was strong, and we found that all students who made correct predictions engaged with the associated HSRCs. We also observed that few students spontaneously critiqued or provided commentary about the correctness of the HSRCs. We are currently conducting more rigorous statistical analyses of both our coding scheme and the data corpus itself in order to better gauge the significance of our findings. The extent to which student conceptual understanding impacts student prediction performance on FR tasks remains unclear, so additional work is ongoing. Ultimately, we hope to use a collection of FR tasks to track student prediction performance over the course of an entire introductory physics sequence to characterize possible shifts in student abilities to follow reasoning.

## ACKNOWLEDGEMENTS

This material is based upon work supported by the National Science Foundation under Grant Nos. DUE-1431857, DUE-1431541, DUE-1431940, DUE-1432765, DUE-1432052, and DUE-0962805.

- [1] National Research Council, *Adapting to a Changing World—Challenges and Opportunities in Undergraduate Physics Education* (Washington, DC: National Academies Press, 2013).
- [2] D. Hestenes, M. Wells, and G. Swackhamer, *Phys. Teach.* **30**, 141 (1992).
- [3] L.C. McDermott and P. S. Shaffer, the Physics Education Group at the University of Washington (Prentice Hall, Upper Saddle River, NJ, 2011).
- [4] N. D. Finkelstein and S. J. Pollock, *Phys. Rev. ST Phys. Educ. Res.* **1**, 010101 (2005).
- [5] M. Kryjevskaja, M. R. Stetzer, and T. K. Lê, 2014 PERC Proceedings (Minneapolis, MN, July 30-31, 2014), edited by P. V. Engelhardt, A. D. Churukian, and D. L. Jones, 143-146 (2015).
- [6] A. Heckler, *Psych. Learn. Motiv.* **55**, (2011).
- [7] J. St. B. T. Evans, *Psychon. Bull. Rev.* **13**, 378 (2006).
- [8] M. Kryjevskaja, M. R. Stetzer, and N. Grosz, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020109 (2014).
- [9] A. E. Lawson, *J. Res. Sci. Teach.*, **15**, 11 (1978).
- [10] J. C. Speirs, W. N. Ferm, Jr., M. R. Stetzer, and B. A. Lindsey, *Physics Education Research Conference Proceedings, Sacramento*, (submitted) (2016).
- [11] E. Cohen and S. E. Kanim, *Am. J. Phys.* **73**, 1072 (2005).
- [12] S. Rozier and L. Viennot, *Int. J. Sci. Ed.* **13**, 159 (1991).
- [13] R. A. Lawson and L. C. McDermott, *Am. J. Phys.* **55**, 811 (1987).