Probing student ability to construct reasoning chains: a new methodology

J. Caleb Speirs,¹ William N. Ferm, Jr.,¹ MacKenzie R. Stetzer,¹ and Beth A. Lindsey²

¹*Department of Physics and Astronomy, University of Maine, 5709 Bennett Hall, Orono, ME, 04469-5709*

²*Department of Physics, Penn State Greater Allegheny, 4000 University Drive, McKeesport, PA, 15132*

Students are often asked to construct qualitative reasoning chains during scaffolded, research-based physics instruction. As part of a multi-institutional effort to investigate and assess the development of student reasoning skills in physics, we have been designing tasks that probe the extent to which students can create and evaluate reasoning chains. In one task, students are provided with correct reasoning elements (i.e., true statements about the physical situation as well as correct concepts and mathematical relationships) and are asked to assemble them into an argument that they can use to answer a specified physics problem. In this paper, the task is described in detail and preliminary results are presented.

I. INTRODUCTION

For more than 30 years, research-based materials developed by the physics education research (PER) community have helped transform introductory physics instruction. These materials typically focus on the development of student conceptual understanding, place considerable emphasis on qualitative inferential reasoning, and scaffold that reasoning via carefully crafted sequences of questions. The effectiveness of these materials on improving conceptual understanding is well documented in the literature [1]. While it is often assumed that these materials also improve student reasoning skills, there is not yet a solid empirical base in the literature in support of such claims.

Research has shown that students who show a solid conceptual understanding on one physics task may fail to apply that same knowledge on a closely related task; moreover, the reasoning patterns of such students shift dramatically between the two tasks [2]. Such inconsistencies in student reasoning have recently been examined through the lens of dual-process theories of reasoning [2–4]. These theories suggest that two distinct processes are involved in student reasoning: process 1, which is fast, subconscious, and automatic; and process 2, which is slow, analytical, and rule-based [5, 6]. When faced with a new physics task, process 1 generates a first-available “intuitive” response, typically based on salient problem features and contextual cues. If there is no intervention by process 2 or if reasoning biases preclude the productive exploration of alternative mental models by process 2, the “intuitive” response from process 1 will become the student’s answer. These studies as well as related work suggest that poor student performance on certain physics tasks may stem more from the nature of student reasoning itself than from specific conceptual difficulties.

As a result, there is a real need to investigate the reasoning skills of students throughout physics instruction in a systematic manner. Such work is critical for the identification of productive reasoning approaches employed by students as well as the kinds of reasoning with which they struggle at various points in instruction; the findings from this work may be used to inform the development of research-based instructional materials optimized to support student reasoning.

As such instruments capable of measuring student reasoning skills must be developed in order to investigate the extent to which these skills are changing over the course of instruction.

The use of qualitative, inferential reasoning chains is ubiquitous in research-based instructional materials such as *Tutorials in Introductory Physics* [7]. A qualitative, inferential reasoning chain (or QIR chain) is a series of conclusions, usually starting from first principles, which lead to an ultimate qualitative comparison. Given how integral such QIR chains are to some research-based instructional materials, it is critical to gain insight into students’ inferential reasoning skills. These skills differ substantively from the scientific reasoning skills assessed via existing instruments such as Lawson’s Classroom Test of Scientific Reasoning [8].

For this reason, as part of a multi-institutional effort to investigate student reasoning in physics, we (along with colleagues at three other institutions) have been developing a series of probes to document the extent to which students are able to follow, replicate, evaluate, and generate chains of reasoning before, during, and after scaffolded, research-based instruction. An important goal of these tasks is to disentangle, to the extent possible, students’ reasoning from their conceptual understanding. During the course of this work, we have developed a new task, the “chaining” task, in which students are asked to assemble existing knowledge elements into “chains”. This task has demonstrated considerable potential to shed light on student abilities to construct QIR chains. In this paper, we provide an overview of the chaining task along with preliminary results from two different implementations.

II. “CHAINING” TASK DESCRIPTION

A. General Overview

The new chaining task allows students to focus on arranging conceptual knowledge into a logical progression of inferences. We accomplish this using a modified card sorting task in which we: (1) provide the student with a list of reasoning elements; (2) indicate that all of the statements within these elements are true and correct; and (3) ask the student to construct a solution to a physics problem by selecting el-

Items	Reasoning Space
$x(t_f) = x_0 + \int_0^{t_f} v(t)dt$	
$v(t_f) = v_0 + \int_0^{t_f} a(t)dt$	
$v = \frac{dx}{dt}$	
$a = \frac{dv}{dt}$	
the integral, $\int_a^b f(x)dx$, is the area under the graph of f vs. x .	Connecting Words
the derivative, $\frac{df}{dx}$, is the slope of the f vs. x graph	and
slope of position vs time graph is velocity	and
slope of velocity vs time graph is acceleration	but
area under velocity vs time graph is the displacement	because
area under acceleration vs time graph is the velocity	therefore
the lines intersect at time B	therefore
slopes are the same at time A	so
Car 1 is going faster than Car 2 initially, but eventually is going slower than Car 2	Conclusions to use in reasoning space
	the speeds are the same at time A
	the speeds are the same at time B
	the speeds are the same at time C
	the speeds are never the same

FIG. 1. Online chaining version of the Kinematics Graph Task (KGT).

elements from the list, ordering them, and, as needed, incorporating provided connecting words (“and”, “so”, “because”, “but”). The reasoning elements primarily consist of observations about the problem setup, statements of physical principles, and qualitative comparisons of quantities relevant to the problem. Everything the student needs to produce a complete chain of reasoning is present in the elements, including statements of requisite conceptual knowledge. Irrelevant elements (which, again, are true statements) are also sometimes included. Thus, the student’s task is to pick from given conceptual pieces and directly assemble a reasoning chain.

At UMaine, tasks incorporating the chaining format have primarily been implemented online using Qualtrics’ “Rank/Group/Sort” question format. This online format is illustrated in Figure 1. Reasoning elements from the “Items” column, connecting words, and final conclusions can be dragged and dropped into the “Reasoning Space” box; the box increases in size vertically as elements are added.

B. Specific Implementation: Kinematics Graph Task

1. Kinematics Graph Task: Original Version

In order to pilot the chaining format, we selected a task that had been used to highlight the role reasoning may play on student performance even when requisite conceptual understanding is possessed by the student. The Kinematics Graph Task (KGT), shown in Figure 2, was administered by Heckler as part of a two question sequence [3]. He used this task to show that salient features may impact student performance on physics problems. In the original study, students were first

FIG. 2. Kinematics Graph Task [3].

presented with a position vs. time graph that had two straight lines of equal slope, and were asked when the speeds were the same; performance was nearly perfect. When presented with the graph shown in Figure 2 and asked the same question, however, student performance dropped to 60% correct. The other 40% of students incorrectly chose time B, the point where the graphs intersect, as the time at which the speeds were the same. The salience of the intersection impacted how students performed on the task, with the first-available response from process 1 effectively precluding the application of the slope-based approach used on the first task.

2. Kinematics Graph Task: Chaining Version

The online implementation of the chaining version of the KGT is shown in Fig. 1. The development of the list of elements provided to students is described below. From typical written student responses to the KGT, we selected reasoning elements that reflected independent logical components of correct chains of reasoning. For instance, in response to the KGT, one student wrote, “The slope of either curve at any point describes the change in position over change in time [or the derivative], which gives us their respective speeds. We can compare the slopes of both curves at a certain time to see if they are equal. This happens at time A.” From this, we broke out four logical components, modifying them slightly in wording to make them technically correct: (1) the slope of the position versus time graph is the derivative dx/dt , (2) the derivative is the velocity, (3) the slope of the position versus time graph is the velocity, and (4) the slopes are the same at time A. These four independent logical elements thus served as the structure of the most complete chain that could be formed in the chaining version of the KGT. Observation elements were created from possible observations that can be made about the problem and associated graphs. For this task, we included both the relevant feature (the slopes are the same at time A) and the irrelevant but salient feature that the lines intersect at time B. We also provided an element that could be used to describe both correct and incorrect interpretations of the graphs (*i.e.*, mapping the speed to either the slope or the vertical height of the graph): “Car 1 is going faster than Car 2 initially, but eventually is going slower than Car 2.”

For the KGT, we intentionally created a list of reasoning elements (*i.e.*, the reasoning elements in the “Items” column) that allowed students to work from first principles directly (“ $v = \frac{dx}{dt}$ ” and “the derivative, $\frac{dx}{dt}$, is the slope of the f vs. x graph”) or to work from a “chunked” knowledge piece [9], or *derived heuristic* (“slope of position vs. time graph is velocity”). By providing this combination of elements, we could determine empirically the extent to which students drew upon first principles in their reasoning.

C. Research Context

Students completed the chaining version of the KGT as part of an extra-credit online exam review before exam 1 in the first semester of the introductory calculus-based physics sequence at the University of Maine. Prior to the exam review, all relevant instruction on kinematics in lecture, laboratory, and recitation had been completed; students had also been given homework assignments on the topic. For the KGT, $N = 157$, corresponding to nearly 55% participation of the class as a whole.

III. RESULTS FROM KGT IN CHAINING FORMAT

Key results for the KGT are described in Table I, II and Figure 3. As shown in Table I, performance on the KGT mirrored that reported by Heckler at Ohio State University [3], with 40% of UMaine students concluding that the speeds were the same at time B.

For those who chose time B, we found that roughly 50% constructed chains that simply linked the observation that “the lines intersect at time B” to the conclusion “the speeds are the same at time B” with a “therefore,” “so,” or “because”. Another 20% of those who chose time B seemed to be demonstrating knowledge that the slope was the velocity and justifying why this was the case via first principles before indicating that speeds were the same at time B. An example of this kind of response is shown in Figure 3. Close to 15% of those who chose time B included the statement that the slope was the velocity but did not attempt to justify why this was the case from first principles. Thus, approximately one third of the students citing the intersection point and choosing time B selected correct conceptual elements that were inconsistent with their incorrect final conclusions. Significantly, none of the students who chose time B as their answer placed the element “the slopes are the same at time A” in their chain.

When examining the chains constructed by students who correctly selected time A on the KGT, we identified two distinct categories of reasoning chains. Reasoning chains in the first category included elements corresponding to first principles either with or without the derived heuristic “the slope of the position vs. time graph is velocity”. The second category included reasoning chains in which the derived heuristic (“slope of position vs. time graph is velocity”) was not justifi-

TABLE I. Results from Kinematics Graph Task (original and chaining versions) as well as results from isomorphic Magnetic Flux Task in chaining format.

Task and Format	Correct (A)	Intersection (B)
KGT Original (OSU, [3])	60%	40%
KGT Chaining (UMaine, $N = 157$)	52%	40%
MFT Chaining (UMaine, $N = 71$)	86%	10%

Reasoning Space slope of position vs time graph is velocity because $v = dx/dt$ the lines intersect at time B therefore the speeds are the same at time B	Reasoning Space $v = dx/dt$ therefore slope of position vs time graph is velocity and the slopes are the same at time A therefore the speeds are the same at time A
(a)	(b)

FIG. 3. (a) Incorrect student response in which correct conceptual knowledge elements accompany elements associated with the incorrect first-available response. (b) Correct student responses in which first principles are used to support the derived heuristic.

fied via first principles. The percentage of correct responses falling into each category is shown in Table II.

IV. ISOMORPHIC FLUX TASK AND RESULTS

We also created and administered a task that was isomorphic to the chaining version of the Kinematics Graph Task but focused on electromagnetic induction. In the Magnetic Flux Task (MFT), the same graphs were used to show the magnetic flux through two different loops (separated in space) as a function of time. Students were asked to determine the time at which the induced EMFs in the loops were the same, if ever. All reasoning elements from the KGT were directly mapped to their electromagnetic analogs (*e.g.*, “ $v = \frac{dx}{dt}$ ” became “ $\mathcal{E} = -\frac{d\Phi_B}{dt}$ ”).

Data were collected in the second semester of the introductory calculus-based sequence on an online final exam review assignment. $N = 71$, which was 43% of the class as a whole.

Student data from the MFT are shown in Table I and II. Overall, performance on the MFT was much stronger than

TABLE II. Nature of justifications provided in support of correct answers on the chaining versions of the KGT and MFT.

Task	Included first principles	Included derived heuristic without first principles
KGT	22%	52%
MFT	51%	18%

on the KGT, with a much smaller percentage of students selecting the intersection. In addition, as shown in Table II, a detailed analysis of students' reasoning chains revealed a greater tendency of students to support correct answers via first principles on the MFT than on the KGT.

V. DISCUSSION OF ALL RESULTS

The similarity in student performance on the chaining version of the KGT reported here and on the original version reported by Heckler in 2011 suggests that the chaining format did not significantly alter student reasoning patterns on the KGT. This is likely due to the fact that we have captured most of the reasoning paths students use in our collection of elements. Thus, properly constructed chaining tasks have the potential to probe student reasoning, and insights gained from such tasks may be generalized beyond this particular format, thereby yielding information about student reasoning patterns that may not be readily obtained from other formats.

Perhaps more importantly, the similarity in performance on the two versions of the KGT suggests that the presence of necessary conceptual information does not preclude the most common incorrect response (the time corresponding to the intersection point). This serves as further evidence that answering the KGT incorrectly cannot be solely attributed to conceptual difficulties. Approximately one third of the students who indicated that the speeds were the same at the intersection also selected relevant, correct conceptual elements, thereby suggesting that they were perhaps struggling to reconcile these elements with process 1's first-available "intuitive" response. The significant percentage of students selecting the intersection even after indicating that slope is related to velocity provides further evidence suggesting that such incorrect responses primarily stem from reasoning difficulties — not from conceptual difficulties.

Through the lens of dual-process theories of reasoning, the stronger performance of students on the MFT (in comparison to the KGT) seems to suggest that students are approaching the MFT with their analytic processes more productively engaged, as evidenced by fewer students selecting the intersection. It is likely that graphs of magnetic flux were not so familiar to students and perhaps they did not expect

that they should have an "immediate" answer; instead, they may have slowed down and reasoned systematically based on first principles (supported by the results in Table II), thereby suppressing the kind of quick, intuitive responses that were so prevalent on the KGT. Future studies will investigate this phenomenon in greater detail and will attempt to rule out alternative explanations (*e.g.*, the performance difference is due to the impact of additional physics instruction).

VI. LIMITATIONS AND CONCLUSIONS

There are some limitations to the chaining task, particularly the online version. For example, there is not a clear way to tell whether an element has been interpreted as intended. Interviews are currently being conducted using the online interface in order to probe student interpretations and to inform any necessary modifications. We have also implemented chaining tasks using 3" x 5" index cards in think-aloud interviews with pairs of students. This allows for greater flexibility and provides greater insight into the approaches used (*e.g.*, discarding certain elements).

In conclusion, we have demonstrated that our new chaining methodology effectively probes student reasoning approaches in a manner that helps decouple reasoning and conceptual understanding. The similarity in student performance on tasks administered in both traditional and chaining formats indicates that the new format does not significantly alter student reasoning patterns (at least on the tasks investigated), suggesting that insights gained may be generalized beyond this format. Finally, the observed differences in student reasoning on the KGT and the MFT could be accounted for using a dual-process framework, further illustrating the relevance of these dual-process theories (originally developed in the context of cognitive psychology) to student reasoning in physics.

ACKNOWLEDGMENTS

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] N. D. Finkelstein, & S. J. Pollock. *Phys. Rev. ST Phys. Educ. Res.* **1**, 010101 (2005).
 - [2] M. Kryjevskaja, M. R. Stetzer, & N. Grosz, *Phys. Rev. ST Phys. Educ. Res.* **10**, 020109 (2014).
 - [3] A. F. Heckler, *Psych. Learn. Motiv.* **55**, 227 (2011).
 - [4] M. Kryjevskaja, M. R. Stetzer, and T. K. Le, *2014 PERC Proceedings*, (Minneapolis, MN, July 30-31, 2014), edited by P. V. Engelhardt, A. D. Churukian, and D. L. Jones, 143-146 (2015).
 - [5] J. St. B. T. Evans, *Psychon. Bull. Rev.* **13**, 378 (2006).
 - [6] D. Kahneman, *Thinking, Fast and Slow* (Farrar, Straus and Giroux, New York, 2011).
 - [7] L. C. McDermott and P.S. Shaffer, the Physics Education Group at the University of Washington, *Tutorials in Introductory Physics, Preliminary Second Edition* (Prentice Hall, Upper Saddle River, NJ, 2011).
 - [8] A. E. Lawson, *J Res. Sci. Teach.* **15**, 11 (1978).
 - [9] *How People Learn: Brain, Mind, Experience, and School*, edited by J. D. Bransford, A. L. Brown, and R. R. Cocking (National Academy Press, Washington, DC, 1999).