Development of a Mechanics Reasoning Inventory

Andrew Pawl*, Analia Barrantes†, Carolin Cardamone†, Saif Rayyan† and David E. Pritchard†

*Department of Chemistry and Engineering Physics, University of Wisconsin-Platteville, Platteville, WI 53818
†Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139

Abstract. Strategic knowledge is required to appropriately organize procedures and concepts to solve problems. We are developing a standardized instrument assessing strategic knowledge in the domain of introductory mechanics. This instrument is inspired in part by Lawson's Classroom Test of Scientific Reasoning and Van Domelen's Problem Decomposition Diagnostic. The predictive validity of the instrument has been suggested by preliminary studies showing significant correlation with performance on final exams administered in introductory mechanics courses at the Massachusetts Institute of Technology and the Georgia Institute of Technology. In order to study the validity of the content from the student's perspective, we have administered the instrument in free-response format to 40 students enrolled in calculus-based introductory mechanics at the University of Wisconsin-Platteville. This procedure has the additional advantage of improving the construct validity of the inventory, since student responses suggest effective distractors for the multiple-choice form of the inventory.

Keywords: problem solving, assessment, introductory physics
PACS: 01.40.Fk, 01.40.G-

MOTIVATION

There is a need for a standardized assessment of problem solving expertise that is valid across institutions. It is not obvious, however, what form such an assessment should take (see, *e.g.*, [1, 2, 3]). One of the earliest findings of physics education research is that experts are adept at classifying physics problems by the principles that will most readily lead to a solution, while novices fixate on surface features of problems which have little bearing on the solution [4, 5]. The favored instrument in this early research was the classification of a large number of problems into groups based on similarity of solution. Unfortunately, this procedure is difficult to standardize due to the large number of problems involved and the complex set of potential responses. Of course, one would expect that strong conceptual understanding would be correlated with the ability to classify problems based upon conceptual structure. This supposition has produced a variety of very successful standardized tests of conceptual understanding in mechanics (*e.g.* [6, 7]) that test student understanding of the application of various principles in physics such as Newton's laws of motion, conservation of mechanical energy and conservation of momentum. They fail, however, to probe one important aspect of expertise, namely the understanding of when specific fundamental principles do *not* apply in a problem. Experts must go beyond understanding the definition of the principles and the procedural application of the principles in familiar situations. They must additionally be able to synthesize these definitions and procedures in order to surmise whether a given principle is likely to apply in an unfamiliar situation. This level of understanding has been called "strategic knowledge" [8].

We are designing a new type of inventory that explicitly tests strategic knowledge in the domain of Newtonian mechanics[1]. It will be a multiple-choice instrument requiring less than 50 minutes to complete. Our approach is to focus on a small number of conceptually rich problems and deeply probe student understanding of the applicability *and* non-applicability of the fundamental principles of mechanics. One important inspiration for the construction of this inventory was Lawson's Classroom Test of Scientific Reasoning [9]. We have adopted Lawson's approach of asking multiple linked questions about each situation presented, typically asking whether a given principle applies and then why it does or does not. Another inspiration was Van Domelen's Problem Decomposition Diagnostic [10]. Decomposition of multi-concept problems is a natural forum for determining if students understand the conditions for applicability of mechanics principles.

DESIGN OF THE INVENTORY

The inventory has been designed in three groups of questions containing a total of eight situations to analyze. Group 1 looks at the applicability and non-applicability of conservation of momentum and conservation of mechanical energy in three different situations. Fig. 1 shows a typical situation and the corresponding questions from group 1. Group 2 (see Fig. 2) looks at the application

---

[1] Available at `http://www.uwplatt.edu/~pawla/MRI`.

of Newton's laws of motion and contains two situations. Group 3 (see Fig. 3) involves decomposing problems and contains three situations.

An astronaut is holding onto a long aluminum antenna attached to a deep-space probe which is floating freely far from any other object. The astronaut is initially at rest, but then begins to climb out along the antenna. The next two questions refer to this situation.

7.) Throughout this process, the linear momentum of the system consisting of the astronaut and the space probe (including the antenna) together is conserved because:
a.) All the forces are internal.
b.) All the forces are conservative.
c.) Linear momentum is always conserved.
d.) The statement is false. Linear momentum is not conserved for this system.

8.) Throughout this process, the mechanical energy of the system consisting of the astronaut plus space probe together is conserved because:
a.) All the forces are internal.
b.) All the forces are conservative.
c.) Mechanical energy of an isolated system is always conserved.
d.) The statement is false. Mechanical energy is not conserved for this system.

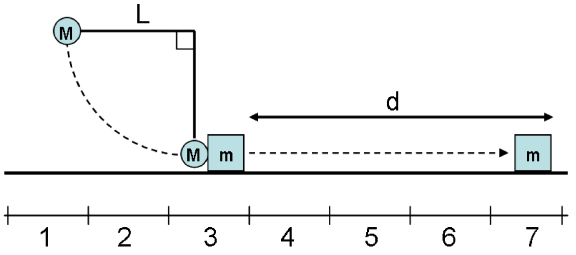
FIGURE 1. Questions 7 & 8 (group 1 situation 2).

A person is trying to move a very heavy safe by pushing it along the ground. The force applied by the person to the safe is perfectly horizontal. Neither the person nor the safe is moving. The following four questions refer to this situation.

11.) The friction force acting on the safe from the ground and the force on the safe from the person are:
a.) Equal in size because of Newton's 2nd Law for the person.
b.) Equal in size because of Newton's 2nd Law for the safe.
c.) Equal in size because of Newton's 2nd Law for the safe plus the person as a single system.
d.) Equal in size because of Newton's 3rd Law.

(The other questions are different pairings of the friction on the person, friction on the safe, the force from the person on the safe and from the safe on the person.)

FIGURE 2. Question 11 (group 2 situation 5).



20.) The pendulum shown in the figure above is released from rest when the string is perfectly horizontal and swings down to hit the box (mass m > M). The pendulum string is vertical when the collision occurs. The pendulum stops after it hits the box, and the box slides a distance d along a rough horizontal surface until it stops. The drawing shows the beginning and end of the motion of the pendulum and the box. You are asked to find the coefficient of kinetic friction between the box and the surface using the quantities described in the problem plus the gravitational acceleration g. What is the most appropriate way to decompose this problem?
a.) One problem: 1-7
b.) Two sub-problems: 1-3, 3-7
c.) Three sub-problems: 1-3, 3, 3-7

21.) My answer to question 20 is justified because:
a.) The work done by the collision forces is unknown.
b.) Mechanical energy is conserved throughout.
c.) Mechanical energy is not conserved when the block is sliding.
d.) The collision is elastic.

FIGURE 3. Questions 20 & 21 (group 3 situation 8).

PRELIMINARY RESULTS

Administrations

The multiple-choice form of the inventory has been administered in several settings. The majority have been at MIT (see Table 1), including a post-course administration to one section of the mainstream introductory mechanics course (8.01), a pre- and post-course administration to the full off-semester introductory mechanics course (8.011) and pre- and post-ReView administrations to two years' worth of a special January ReView for students who received a D in 8.01. Each of these, except for the most recent post-ReView and the pre-8.011 administrations, has been separated from a high-stakes administration of a standard 8.01 final exam by a brief period (one to three weeks) during which essentially no physics instruction took place. The multiple-choice form of the inventory has additionally been administered post-course to 80 students enrolled in the regular freshman mechanics course at the Georgia Institute of Technology.

TABLE 1. Administrations of the inventory at MIT.

Course	Term	N_{pre}	N_{post}	N_{total}
8.01	Fall 2009	N/A	42	42
ReView	January 2010	41	37	78
ReView	January 2011	45	45	90
8.011	Spring 2011	57	53	110

Encouraging Predictive Validity

The one time during the course that students are forced to consider the applicability of the full array of principles learned in mechanics to a wide variety of problems is a

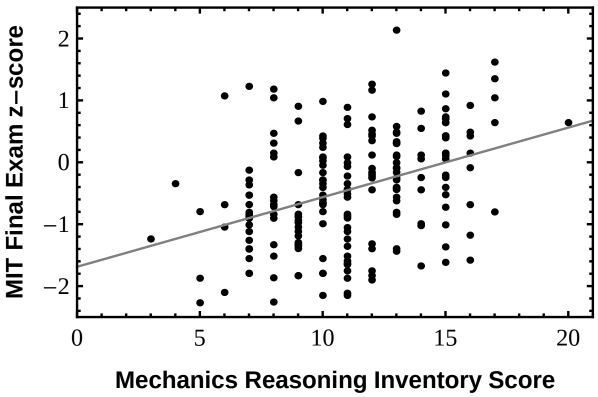


FIGURE 4. Correlation of inventory score with MIT final exam performance ($r = 0.38$, $N = 206$).

cumulative final exam. We therefore consider correlation of the inventory score with performance on a cumulative final exam to be a necessary component of the predictive validity of the final form of the inventory. The preliminary version of the inventory exhibits a significant correlation with performance on the written analytic problems included in the MIT 8.01 final exam (Fig. 4), yielding $r = 0.38$ for 206 students. These results are in keeping with data from the administration of the inventory at the Georgia Institute of Technology, which yielded a correlation with final exam score of $r = 0.47$ for 80 students.

Another measure of predictive validity is correlation of improvement of inventory scores with improved problem solving ability. Since 2009 the authors have been responsible for a three-week January “ReView” course at MIT offered to students who have just completed the regular mechanics course (8.01) with a grade of D which is explicitly designed to provide intensive training in the type of reasoning that the inventory is designed to test. It has proved successful in elevating the problem solving ability of the students [11]. The inventory was administered as a pre- and post-ReView assessment in 2010 and 2011 to a total of 77 students. The performance on the inventory improves significantly ($p \sim 10^{-7}$) during the ReView, with the mean score improving from 9.8 ± 0.3 points to 12.3 ± 0.3 points out of 21. The ReView students exhibit significant improvement ($p \leq 0.05$) on 10 of the individual inventory items. Two items (questions 10 and 16) show statistically non-significant loss.

Item Statistics

Table 2 lists the fraction of correct responses and the discrimination for the data set consisting of all 333 administrations of the inventory to MIT students. The discrimination was calculated by differencing the number of correct responses among the top and bottom 111 inventory scores (1/3 of the total sample) and dividing by 111. (High discrimination means that better performing students get the question right while poor performers get it wrong.) Looking at group 1 (situations 1, 2 and 3), we note that situation 1 yields more promising discriminations than situations 2 and 3. In group 2 (situations 4 and 5), it appears that situation 4 has substantial defects. Finally, all the group 3 situations should be re-examined.

FREE-RESPONSE ADMINISTRATION

The Sample

We recruited 40 students enrolled in calculus-based introductory mechanics at the University of Wisconsin-Platteville to take a free-response form of the inventory during the final week of classes for the Spring 2011 semester. These students were offered $10 for 50 minutes of work on the questions, meaning that not all students completed the full inventory. The various groups of questions were given in random order so that each question was completed by a minimum of 31 students.

Findings

Questions are Easily Understood

One important finding is that the students interpreted the questions as intended by the authors. One exception is the decomposition problems, where 15 out of 38 students hurried through the instructions and failed to write out a decomposition in terms of the numbered intervals shown in the figures. This is not a concern on the multiple-choice form of the inventory, however. Another issue is that situation 4 (questions 5 and 6) is ambiguous. In it, students choose whether to consider two blocks connected by a rope as one system or separate systems. In the written responses, two students indicated they would analyze only one system but specified only one of the blocks as their system, which is the correct procedure but not the intended meaning of “one system” in the question. Given the poor discrimination associated with these questions as shown in Table 2, situation 4 will be dropped from the inventory.

Multi-Question Format is Important

The inventory is inspired by Lawson’s question format, where each regular question is followed by a question examining the reasoning employed. To reduce the length, however, the authors frequently employed a modified format as shown above in Figures 1 and 2, where an assertion is made by the *question* rather than by the student and the reasoning is demanded. (To be contrasted with the full format shown in Fig. 3.) Importantly, situation 1 (questions 1-4) did employ a full “Lawson format”. Questions 3 and 4 as a pair, for example, are very similar to the single question 8 (see Fig. 1). Question 3 asks whether mechanical energy is conserved in situation

TABLE 2. Aggregate data from preliminary administrations of the multiple-choice Mechanics Reasoning Inventory, $N = 333$.

Group	1								2							3					
Situation	1				2		3		4		5					6		7		8	
Question	1	2	3	4	7	8	9	10	5	6	11	12	13	14	15	16	17	18	19	20	21
% Correct	62	50	73	71	77	16	32	15	85	47	69	80	71	79	66	72	52	9	19	74	10
Discrimination (%)	41	64	45	44	32	23	32	23	20	23	51	15	51	33	38	34	29	16	25	28	15

1 and question 4 asks for the reasoning. Questions 1-4 exhibit more robust discrimination than the other group 1 situations. Additionally, the entire free-response version of the inventory used the full format (*e.g.* question 8 began by asking whether mechanical energy is conserved and then asked for a justification) and the response patterns were qualitatively different. The UW-Platteville students significantly underperformed the MIT students on question 3 (full-format for both) but question 8 was the only group 1 question on which the UW-Platteville students outperformed the MIT students significantly.

The written responses also imply that switching to the full Lawson format in situation 5 (Fig. 2) could help the discriminations by probing for the claim that certain pairs of forces are *not* equal. In free-response form the "not equal" response rate was over 40% for two of the pairings involving friction forces.

Potential for Reduced Jargon

Jargon like "conservative forces", "external forces" and "Newton's 2nd (or 3rd) Law" appears in the inventory. Of these, "conservative force" is the most misleading. Some students and faculty equate "conservative" with non-dissipative while others reserve the term for forces giving rise to potential energy. This confusion seems responsible for the poor performance observed on question 10 of the inventory. "Newton's 2nd(3rd) Law" can be replaced by discussing acceleration or with the standard action-reaction phrasing, respectively. On question 12, which involves an action-reaction pair, only 33% of the students who clearly gave the correct reasoning mentioned the 3rd Law and an equal number exhibited confusion about which of the three laws was the action-reaction law. "External" is the most interesting. In the clearly isolated situation 2 (Fig. 1) 25% of the written responses discussed the absence of outside or external forces. However, in the very similar situation 1 where objects move along frictionless ground instead of deep space, only 10% mentioned internal or external forces.

Need for Better Correspondence of Answers

A possible reason for the poor performance of students on the decomposition problems of group 3 is that the justifications do not explicitly correspond to a specific number of stages (see Fig. 3). 75% of the students who gave the correct written response to question 20 provided a justification that explicitly had three stages.

SUMMARY

The Mechanics Reasoning Inventory being developed shows promising predictive validity and alignment with instruction. The free-response administration indicates that the questions are understood by the students and suggests several ways to improve the inventory.

ACKNOWLEDGMENTS

The authors thank D. Caballero for administering the inventory at Georgia Tech. Development of the inventory was supported by PHY-0757931 and DUE-1044294 from the NSF and 1-RC1-RR028302-01 from the NIH. The free-response study was funded by a University of Wisconsin-Platteville SAIF grant.

REFERENCES

1. T. French and K. Cummings, in *Proceedings of the 2001 Physics Education Research Conference*, edited by S. Franklin, J. Marx and K. Cummings, PERC Publishing, Rochester, NY, 2001, pp. 87-91.
2. W.K. Adams and C.E. Wieman, in *Proceedings of the 2006 Physics Education Research Conference*, edited by L. McCullough, L. Hsu and P. Heron, AIP Conference Proceedings, Melville, NY, 2007, pp. 18-21.
3. J. Marx and K. Cummings, in *Proceedings of the 2010 Physics Education Research Conference*, edited by C. Singh, M. Sabella and S. Rebello, AIP Conference Proceedings, Melville, NY, 2010, pp. 221-224.
4. M.T.H. Chi, P.J. Feltovich and R. Glaser, *Cogn. Sci.* **5**, 121-152 (1981).
5. P.T. Hardiman, R. Dufresne and J.P. Mestre, *Mem. Cognit.* **17**, 627-638 (1989).
6. D. Hestenes, M. Wells and G. Swackhamer, *Phys. Teach.* **30**, 141-158 (1992).
7. R. Thornton and D. Sokoloff, *Am. J. Phys.* **66**, 338-352 (1998).
8. W.J. Gerace, in *Proceedings of the 2001 Physics Education Research Conference*, edited by S. Franklin, J. Marx and K. Cummings, PERC Publishing, Rochester, NY, 2001, pp. 33-36.
9. A.E. Lawson, *J. Res. Sci. Teach.* **15**, 11-24 (1978).
10. D. Van Domelen, "The Development of the Problem Decomposition Diagnostic", Ph.D. Thesis, Ohio State University (2000).
11. A. Pawl, A. Barrantes and D.E. Pritchard, in *Proceedings of the 2009 Physics Education Research Conference*, edited by M. Sabella, C. Henderson and C. Singh, AIP Conference Proceedings, Melville, NY, 2009, pp. 51-54.