

10-18-2007

# Comparing three methods for teaching Newton's third law

Trevor I. Smith

Rowan University, smithtr@rowan.edu

Michael C. Wittmann

Follow this and additional works at: [http://rdw.rowan.edu/csm\\_facpub](http://rdw.rowan.edu/csm_facpub)



Part of the [Physics Commons](#)

## Recommended Citation

Smith, Trevor I. and Wittmann, Michael C., "Comparing three methods for teaching Newton's third law" (2007). *Faculty Scholarship for the College of Science & Mathematics*. 86.

[http://rdw.rowan.edu/csm\\_facpub/86](http://rdw.rowan.edu/csm_facpub/86)

This Article is brought to you for free and open access by the College of Science & Mathematics at Rowan Digital Works. It has been accepted for inclusion in Faculty Scholarship for the College of Science & Mathematics by an authorized administrator of Rowan Digital Works. For more information, please contact [jiras@rowan.edu](mailto:jiras@rowan.edu), [rdw@rowan.edu](mailto:rdw@rowan.edu).

## Comparing three methods for teaching Newton's third law

Trevor I. Smith and Michael C. Wittmann

*Department of Physics and Astronomy, College of Education and Human Development, Center for Science and Mathematics Education Research, University of Maine, Orono, Maine 04469, USA*

(Received 20 April 2006; revised manuscript received 11 December 2006; published 18 October 2007)

Although guided-inquiry methods for teaching introductory physics have been individually shown to be more effective at improving conceptual understanding than traditional lecture-style instruction, researchers in physics education have not studied differences among reform-based curricula in much detail. Several researchers have developed University of Washington-style tutorial materials, but the different curricula have not been compared against each other. Our study examines three tutorials designed to improve student understanding of Newton's third law: the University of Washington's *Tutorials in Introductory Physics* (TIP), the University of Maryland's *Activity-Based Tutorials* (ABT), and the *Open Source Tutorials* (OST) also developed at the University of Maryland. Each tutorial was designed with different goals and agendas, and each employs different methods to help students understand the physics. We analyzed pretest and post-test data, including course examinations and data from the Force and Motion Conceptual Evaluation (FMCE). Using both FMCE and course data, we find that students using the OST version of the tutorial perform better than students using either of the other two.

DOI: [10.1103/PhysRevSTPER.3.020105](https://doi.org/10.1103/PhysRevSTPER.3.020105)

PACS number(s): 01.40.Fk, 01.40.G-, 01.40.gb

### I. INTRODUCTION

Studies conducted in recent years have shown that guided-inquiry, University of Washington-style tutorials can be effective supplements to traditional introductory physics instruction.<sup>1-4</sup> However, little comparative research has been done concerning how different types of tutorials compare in effectiveness or vary in structure and content. We wish to address the former issue, recognizing that addressing the latter issue might explain the differences. In the context of Newton's third law, three tutorials exist and can be compared in style and effectiveness: a pencil-and-paper tutorial from the University of Washington's *Tutorials in Introductory Physics* (TIP);<sup>5</sup> a microcomputer-based laboratory tutorial developed at the University of Maryland as part of the *Activity-Based Tutorials* (ABT);<sup>6-9</sup> and a refining raw intuitions tutorial that is part of the *Open Source Tutorials* developed at the University of Maryland.<sup>10-12</sup> This study aims to compare the effectiveness of each tutorial in improving student understanding of Newton's third law. Throughout, we will refer to them by their abbreviations (TIP, ABT, and OST) even when a phrase like "ABT tutorial" is redundant (like PIN number).

Research for this study was conducted within the General Physics I (PHY 111) course at the University of Maine (UMaine) during the fall of 2004. The course is the first half of an algebra-based introductory course and enrolls approximately 100 students per semester ( $N=107$  in this study). The student population enrolled in PHY 111 consists primarily of life science and earth science majors, and roughly three-quarters of the students have previously taken a calculus course. The course consists of two fifty-minute large-class lectures, two fifty-minute tutorial sessions, and one two-hour laboratory period per week. The student population is divided into six sections for the tutorial portion of the course. The laboratory portion is also divided into six sections, but the populations of these sections are independent of those of

the tutorial sections. The majority of tutorials for the course come from TIP.<sup>5</sup>

In our study, we grouped the six tutorial sections so that each of the Newton's third law (N3) tutorials was completed by two sections (approximately one-third of the class), and no students completed more than one of the tutorials. Our goal was to observe differences in student learning using random assignment of a tutorial to a section and trying to control for all other variables in the course.

### II. INSTRUCTIONAL MATERIALS

#### A. The physics of Newton's third law

Newton's third law states, in its simplest form, that the force of one object on a second is the same size (magnitude) as that on the first by the second, but in the opposite direction. We analyze student reasoning about N3 in two categories: pushing situations and collision situations. Pushing situations occur when two objects are in contact for an extended period of time. The objects might be speeding up, moving at a constant speed, or slowing down. Also, the objects might be arranged such that the masses are unequal, with the leading or the trailing object having a larger mass. Each of these leads to the same result, but may cause problems for students, as described below. An example of a pushing situation can be seen in Fig. 1. Collision situations occur when two objects interact for a brief period of time. As with pushing situations, various combinations of the objects' relative velocities and masses may affect how students reason about the forces exerted. An example of a collision situation is presented in Fig. 2. Both types of situation are prevalent throughout this study and are commonly referenced during the instruction of N3.

#### B. Common aspects of all tutorial instruction

All three tutorials implemented during this study use guided-inquiry methods as a basis for teaching N3. In tuto-

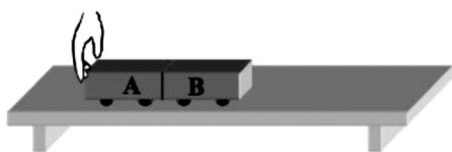


FIG. 1. Carts *A* and *B* are on a table as shown. A constant force is exerted on cart *A* by the hand so that the carts speed up at a constant rate. Compare the magnitudes and directions of the force exerted on cart *B* by cart *A* with the force exerted on cart *A* by cart *B*.

rials, students work in groups of three or four to complete worksheets that ask them to examine the qualitative nature of the physics with less emphasis on problem solving than is typically found in textbooks. The instructor acts as a facilitator, asking appropriate questions to engage the students in the tutorial without explicitly telling them the correct answers.

Each of the tutorials was originally designed to be completed within one fifty-minute period, but students rarely complete them in the allotted time.<sup>13</sup> As a result, students in the UMaine course are allowed two fifty-minute periods per week to complete each tutorial.

For our study, all students took the same ungraded pretest and answered identical homework questions (derived from the ABT materials<sup>6,14</sup>). Furthermore, all students received the same instruction during the lecture and laboratory sessions. The only difference in instruction occurred during the tutorial periods themselves.

### C. Describing the three curricula

The TIP tutorial for N3 was used for the control group for our study due to the use of TIP tutorials in the rest of the course. The TIP tutorial emphasizes a process of eliciting student responses, confronting incorrect answers, and resolving inconsistencies as a way of dealing with student misconceptions.<sup>3,15</sup> The students are presented with a situation in which objects of unequal mass are pushed by an external force, where the external force acts on a large mass that is in contact with (and pushes on) a small mass. In the tutorial, students must draw free-body diagrams for each object and make their diagrams consistent with the motion of the system of objects. Different physical situations are studied, including constant speed motion and accelerated motion. For comparative reasons, we note that the TIP tutorial contains only pushing situations.



FIG. 2. Cart *B* is at rest on a table as shown. Cart *A* approaches cart *B* from the left and collides with cart *B*. Immediately after the collision, the two carts separate. Compare the magnitudes and directions of the force exerted on cart *B* by cart *A* with the force exerted on cart *A* by cart *B* while the carts are in contact.

The ABT tutorial was developed at the University of Maryland<sup>16</sup> and utilizes microcomputer-based laboratory (MBL) data acquisition techniques.<sup>6,14</sup> The ABT tutorial employs low-friction carts and force probes to allow students to perform qualitative experiments involving both pushing and collision situations. In the pushing situation, students carry out experiments similar to those in the TIP tutorial. The carts have unequal masses, with the more massive cart pushing the less massive cart in accelerating motion. In the collision situations, students examine the forces exerted during a collision in which one cart has an initial nonzero velocity and the other is initially at rest. Two situations are tested: one in which the carts have equal masses and another in which they have unequal masses. Data are gathered from the force probes and plotted on a computer screen during the experiments. The students observe visual information that shows that N3 holds for several situations. Students quickly learn to expect that the graphs will have equal magnitude (in opposite directions) even as their intuitions tell them otherwise.

The OST tutorial was also developed at the University of Maryland and emphasizes the refinement of students' intuitions when studying N3.<sup>10,11</sup> The tutorial begins with a collision situation in which a massive truck collides with a stationary, lower-mass car. Students are asked to compare the force exerted by the more massive truck on the car with that exerted by the lower-mass car on the truck. Typically, students believe that the truck exerts the larger force, though their reasoning varies from student to student. Some students state that this force is larger because the truck is moving, while the car is still. Others state that it is larger because the mass of the truck is larger than that of the car and will, therefore, do more damage. In sum, students tend to state that the car will "react more" during the collision than the truck; therefore, more force must be exerted on the lower-mass object. The tutorial takes the students through a series of observations (including a MBL experiment similar to that in the ABT tutorial) and thought exercises designed to allow them to refine their raw intuitions. By helping students define conflicts in intuitions and consider ways to resolve these conflicts, the tutorial helps students gain an understanding of the difference between force and acceleration. The OST tutorial originally contained only collision situations. To create common situations in all tutorials (allowing common homework and examination questions), we added a section including a pushing situation to the end of the tutorial as a way for students to practice what they had learned during instruction. This is very similar to the pushing situations in the ABT tutorial, but without the use of MBL.

We note that the three tutorials were written for very different populations from the one at UMaine. Both the TIP and the ABT tutorials were designed for calculus-based introductory physics courses, though neither of these two tutorials explicitly requires any knowledge of calculus. OST tutorial author Andrew Elby's profession as a high school physics teacher encouraged him to create the OST tutorial with that population in mind. As with the TIP and ABT tutorials, no advanced mathematics is needed to successfully complete the OST. All tutorials assume knowledge of Newton's second law and kinematics, and all assume that N3 has been discussed in class prior to tutorial instruction. The OST tutorial,

however, can function independently of this constraint.

### III. RESEARCH STUDY DESIGN

#### A. Procedures

The tutorial portion of the PHY 111 course was divided into six sections. Each type of tutorial was administered to two sections during regularly scheduled tutorial periods. Sections were randomly chosen. N3 had been covered in the lecture portion of the course prior to the tutorial periods. We gathered four types of data, from post-lecture, pre-tutorial pretests, post-tutorial homework, course examinations, and the Force and Motion Conceptual Evaluation (FMCE).<sup>17</sup> We describe these data in more detail below.

All data were gathered before grading occurred, and analysis was kept completely separate from the grading process.

#### B. Data gathered

All pretest and post-test (homework and examination) assessments were the same for all students regardless of instruction.

The pretest and homework questions came from materials accompanying the ABT tutorial and contained both pushing and collision situations.<sup>7</sup> The students completed the pretest during the first ten minutes of the first tutorial period. Thus, all students for whom we have pretest data also participated in the tutorials, and our matched data include only data from students who were in tutorials. At the end of the second tutorial period, students were given the homework. They had approximately one week to complete it.

The examination question was developed locally and contained only a pushing situation (this was the motivation for adding the pushing situation to the OST tutorial). The question was administered as part of an 11-question (four open response, seven multiple choice) midterm examination approximately two weeks after the tutorials had been completed. Students were required to answer the question on N3.

The Force and Motion Conceptual Evaluation<sup>17</sup> was administered at the beginning and end of the semester as part of the regular course work. The N3 cluster of the FMCE contains both pushing and collision situations. We compared incoming student performance on the overall scores of the FMCE to show that students were generally equal before instruction.

### IV. DATA ANALYSIS

The preliminary pretest data were examined to find a way to categorize the errors students were making in their thinking about N3.<sup>18</sup> The original goal was to quantify the errors students were making and determine by how much their use of these errors changed from the pretest to the post-tests. The errors we uncovered are well aligned with a description based on cognitive resources<sup>11,19</sup> and facets of reasoning.<sup>20</sup> Most student responses could be classified into three categories: action dependence, velocity dependence, and mass dependence responses. Occasionally answers were given that

did not fit into these categories, but such occurrences were rare and did not merit an additional category. Though defined independently, our categories are very similar to the contextual features developed by Bao, Hogg, and Zollman.<sup>21</sup>

#### A. Defining common facets of reasoning

We describe student responses in terms of facets of reasoning.<sup>20</sup> Facets, as described using the numbering code provided by Facet Innovations, Inc., describe many common ways in which students respond to questions. They are “lightly abstracted” from what students actually say in the classroom. While the entire facet cluster about “Forces as Interactions” contains elements not relevant to our study, several facets are very important to our study.

We use the common numbering system of facets, with higher values (up to 99) more problematic to instruction and the lowest values (facet 00, or variations numbered 01 or 02, etc.) being correct. Most important to our study is the facet 60 cluster of facets. Facet 60 states, “The student indicates that the forces in a force pair do not have equal magnitude because the objects are dissimilar in some property (e.g., bigger, stronger, faster).”

There are several variations to this cluster, including:

61; the “stronger” object exerts a greater force;

62; the moving object or a faster-moving object exerts a greater force;

63; the more active or energetic object exerts more force;

64; the bigger or heavier object exerts more force.

In particular, facets 62, 63, and 64 were found to be important to our study and analysis. Based on our preliminary analysis of the types of responses given by students, we group all student responses to the facet 60 cluster into three categories: the action dependence facet (similar to facet 63), the mass dependence facet (similar to facet 64), and the velocity dependence facet (similar to facet 62). We describe these three in more detail below.

All tutorials addressed elements of the facet 60 cluster. The TIP and ABT tutorials had pushing situations with unequally massed objects. The ABT and OST tutorials involved collision situations with unequally massed objects and unequal pre-contact velocities for the masses.

#### 1. Action dependence facet

The action dependence facet (ADF) embodies the notion that one object causes a force, and the other object feels that force. This facet is most likely influenced by the common (mis)statement of N3, “for every action there is an equal, but opposite, reaction.” Typically, students are more likely to focus on the action-reaction aspect of this statement rather than the equal-opposite portion.

The ADF manifests itself slightly differently in pushing and collision situations. In pushing situations, a student might state that the object doing the pushing is exerting a greater force than the object being pushed. In collision situations, a student might state that the object that initially has a greater speed exerts more force than the object initially at a slower speed. In Figs. 1 and 2 above, a student stating that the force that *A* exerts on *B* is greater than the force that *B*



exerts on  $A$  and giving the appropriate (incorrect) explanation could be classified as using the ADF. The ADF was the most commonly used incorrect reasoning in our study.

### 2. Velocity dependence facet

The velocity dependence facet (VDF) arises from a confusion between velocity and acceleration. Students often think of force as an intrinsic property of a body in motion (similar to momentum) rather than a product of the interaction of two bodies.<sup>17,22</sup>

In pushing situations, a student might express the thought that the forces the bodies exert on each other are equal only if the two bodies are moving at a constant velocity. The VDF would lead to a correct answer for incorrect reasons. In collision situations, a student might discuss the force of a moving car being transferred or imparted to a stationary one as it starts to move. Applications of the VDF are relatively rare in our study.

### 3. Mass dependence facet

The mass dependence facet (MDF) expresses the notion that more massive bodies always exert more force than less massive bodies. Students will often cite Newton's second law ( $\Sigma F=ma$ ) as evidence of this; however, students often forget that Newton's second law deals with the net force on an object, not each individual force. The MDF is utilized similarly in pushing and collision situations.

### 4. Using multiple facets

We, like Bao *et al.*,<sup>21</sup> have found that students may use two or more of these facets in situations dealing with N3. A more massive object smashing into a smaller, stationary object might elicit all three facets in a single problem. In such situations, students may not completely describe their reasoning because each facet leads to a consistent result. We have found that it is easier to determine which facets students used in questions where the application of different facets yields conflicting results. For example, in Fig. 2, a student guided by the ADF will likely state that object  $A$  exerts more force on  $B$  than  $B$  exerts on  $A$ , while a student guided by the MDF will state just the opposite. It is possible that students will give the correct answer by compensating between the two arguments: the ADF and MDF balance such that forces exerted are equal. We find that, in written explanations, students only rarely write down an explanation that uses more than one type of facet. Further work in this area is warranted, especially in studying the existence of false positives in standardized test results like the FMCE.

#### B. Coding student extended response answers

Having defined the three major facets we determined were used by students, we analyzed and coded the pretests to find which of these facets were used by each of the students in this sample. We quantified facet use by describing both the facet and the situation in which that facet was applied. For each type of situation (pushing or collision) the student was given a score representing the number of facets used. Re-

TABLE I. Pretest data. Tutorial populations are reported as well as the mean and standard deviation of the number of facets students used. The mean number of facets used on pushing questions is also reported, as it will impact the analysis of our examination data. No statistically significant differences are evident among the pretest results between each of the tutorials.

	$N$	Mean		Standard deviation
		Push	Total	Total
TIP	34	1.25	3.50	1.83
ABT	34	1.28	3.12	1.67
OST	27	1.48	3.81	1.47

gardless of how often each facet was used in a series of questions, one point was assigned for each type of facet (ADF, VDF, or MDF) for a maximum score of three facets per situation type, or six facets total. The emphasis lay on whether a student used a facet while reasoning about a particular type of situation, not on how often the facet was used across several questions. In the rare occurrence that a student presented an incorrect response that did not fit into the coding scheme, one point was added to the total score for the situation type in which the response occurred. There was never a time in which all three facets appeared concurrent with unidentifiable responses. As such, the maximum number of facets per student remained at three per situation type and six total.<sup>23</sup> Analysis of results displayed in Table I showed no significant differences in students' base understanding of N3 among the tutorial sections. Over 90% of the students in the class used at least one incorrect facet on the pretest.

The coding and quantifying scheme described above was used to analyze the homework and examination data. The homework contained pushing and collision questions, allowing for six possible facets to be used. Because the examination data contained only pushing and not collision situations, the maximum number of facets used was three (not six). Furthermore, while the entire homework was compared to the entire pretest, the examination was compared only to the pushing portion of the pretest data.

We created a measure called the "facet difference score" to indicate how much each student improved as a result of the tutorial. Facet differences were calculated simply by subtracting the post-test number of facets used from the pretest number of facets used. Scores for all students were averaged according to the type of tutorial instruction they received. These scores were used as a measure of the effectiveness of each tutorial: a higher difference in facets means greater average improvement by the students in the course and indicates a more effective tutorial.

We used the inferential statistic analysis of variance (ANOVA) to compare students' facet difference scores based on the type of tutorial used during instruction. The statistics take into account different class sizes. The ANOVA allowed us to compare all three tutorials at the same time to look for statistically significant differences and justify these differences as applying to broader ranges of students. We were also able to examine each pair of tutorials separately in order

to determine how each tutorial's student performance compared to each of the other tutorials. The threshold for statistical significance for this experiment was set at a probability level of  $p \leq 0.05$ . This indicates a 95% certainty that our data do not represent a random result, but rather a true difference in the effectiveness of the tutorials. We used other data sources to support our conclusions.

### C. Analyzing FMCE responses

Data collected from the FMCE were analyzed using an analysis template created by one of the authors (M.C.W.) and available online.<sup>24</sup> Modifications were made, as described below. We used the overall score to characterize similarities in student populations. We focused all other elements of our analysis on the ten questions dealing with N3. Four of these are included by Thornton but should not be used in analyzing student responses.<sup>25</sup> Of the six remaining questions, four are collision questions (30–32 and 34) and two are pushing questions (36 and 38).

For the N3 FMCE data, we first quantified the number of correct responses each student gave. We computed the normalized gain for each student in place of the facet difference score used previously. The normalized gain is the ratio of each student's improvement divided by his or her capacity for improvement.<sup>26</sup> For example, a student who went from two correct responses at the beginning of the semester to five correct responses at the end of the semester improved three points out of a possible four (since the maximum is six correct). This student's normalized gain would be 0.75. We again used ANOVA at the  $p \leq 0.05$  level to test for statistical significance using our computed gain values. We distinguished between pushing and collision situations as well as looking at the overall N3 cluster score.

Since the FMCE is a multiple choice survey, we analyzed the offered distractors in terms of the facets that might lead a student to give that answer. This analysis led to important results on three questions: 30, 31, and 32. In these questions, a truck and a car collide. The truck is much more massive than the car. In question 30, the car and truck are moving at the same speed. In question 31, the car is much faster than the truck. In question 32, the truck is motionless when the car collides with it. (Note that none of these are similar to the OST question of a truck hitting a stationary car.)

The offered responses for these questions are all from the same list. Answers indicate a reliance on different facets to guide one's reasoning about N3. For example, response *A* states "The truck exerts a larger force on the car than the car exerts on the truck." Giving this response indicates use of the MDF, since the mass of the truck is always larger than that of the car. Question 30 is commonly answered with response *A*, indicating that the force exerted by the truck is larger when car and truck move at the same speed. Response *B* states, "The car exerts a larger force on the truck than the truck exerts on the car." When given in response to question 32, this indicates the ADF, since the car is moving and acting on the stationary truck.

For question 31, answers are often correct but indicate incorrect reasoning. We infer that students' responses to

TABLE II. Mean facet difference score for each tutorial on each post-test assessment. Larger value indicates more improvement. The maximum correction value for each assessment is displayed at the bottom.

	HW	Exam
TIP	1.29	0.94
ABT	0.96	0.83
OST	1.86	1.36
Maximum	6	3

questions 30 and 32 indicate what they might be thinking on question 31. For those who use MDF on question 30 (truck wins) and ADF on question 32 (moving car wins), we find many who correctly say that the forces are equal when the faster car hits the heavier, slower truck (question 31). Others answering MDF on 30 and ADF on 32 say that there is not enough information to properly answer 31. In both cases, we believe that the students are not using N3 correctly to answer the question but are instead balancing the MDF and ADF, compensating in order to arrive at balanced forces in that situation. To characterize these students' responses accurately, based on our analysis, we assigned values to the use of different facets in answering the questions. So, if a student used the MDF in 30, the ADF in 32, and answered correctly in 31, we assigned 1.5 questions to MDF, 1.5 questions to ADF, and none to the correct answer.

For all the collision and pushing questions, we added up all the instances of MDF and ADF use, as well as those who gave the correct answer of forces being equal in any form of collision or pushing. We then compared improvement in each area: the normalized gain in correct responses, the improvement in MDF (meaning, the fraction of instances of MDF that changed), and the improvement in ADF (the fraction of instances of ADF that changed).

## V. RESULTS

Consistent with the literature related to the three curricula from which we pulled the tutorials, all three tutorials were found to improve student understanding of N3. We report on data comparing pretests to homework and exams, as well as the FMCE data as described above.

### A. HW and exam facet difference scores

Table II shows the facet difference score for each type of instruction, indicating the average improvement in student understanding for each of the tutorials as measured by each of the post-test assessments compared with their respective pretests. The "HW" and "Exam" columns refer to the homework and examination assessments, respectively. Larger values indicate a greater overall student improvement and a more effective tutorial. The value of maximum improvement for each assessment is displayed at the bottom of the table. For the homework and exam data, the maximum improvement possible would correspond to a student who used all

TABLE III. Probability ( $p$ ) values computed using ANOVA. Values for the main effect and each post hoc comparison are displayed. Values displayed in italics are significant,  $p \leq 0.05$

	HW	Exam
Main effect	0.14	<i>0.04</i>
TIP vs ABT	0.42	0.61
TIP vs OST	0.17	<i>0.05</i>
ABT vs OST	0.06	<i>0.01</i>

three incorrect facets on the pretest and no incorrect facets on the post-test assessment.

In Table III, we display the results of the ANOVA comparisons in terms of the calculated probability ( $p$ ) values for all the data presented in Table II. The main effect indicates the comparison between all three tutorials to determine if significant differences exist. Each of the pairwise comparisons then corresponds to the differences between those two particular tutorials. The highlighted values indicate results that were found to be significant ( $p \leq 0.05$ ).

Based on our ANOVA comparisons, we find that the facet difference score for homework is not significant, while the examination scores are significant. Data indicate results that are significant for comparisons between the OST tutorial and each of the other two.

### B. FMCE gains and improvements

We report on three different elements of the FMCE data, using the analysis methods described above. We focus on six of the ten N3 questions, four collision and two pushing. We analyze overall N3 results as well as results in each cluster. As described below, we find that the OST materials lead to larger improvements than the other two tutorials, which are statistically similar to each other. For each of Figs. 3–5, we plot three forms of data per tutorial. The first is the normalized gain in specified set of questions. The second is the improvement in the use of the MDF, where we normalize to the proportion of students who used that facet before instruction and calculate the fraction of that group that improved on the post-test. The third is the improvement in the use of the ADF, calculated similarly.

Statistical significance of these measures is shown in Table IV. We italicize those data that are significant ( $p \leq 0.05$ ). We do not discuss the data from the pushing cluster because the main effect comparisons showed no statistical significance. This came for several reasons. First, there are only two questions analyzed in the pushing cluster, making the statistics very coarse grained. Second, very few students use the MDF before instruction, and nearly none after. Thus, the data on the MDF were relatively meaningless. Finally, the overall normalized gain in the pushing questions was essentially identical for all tutorials. We comment further on the pushing cluster data below.

When looking at Fig. 3, three observations are clear. First, the OST tutorial performed much better than the other two in overall performance and in addressing the use of the MDF.

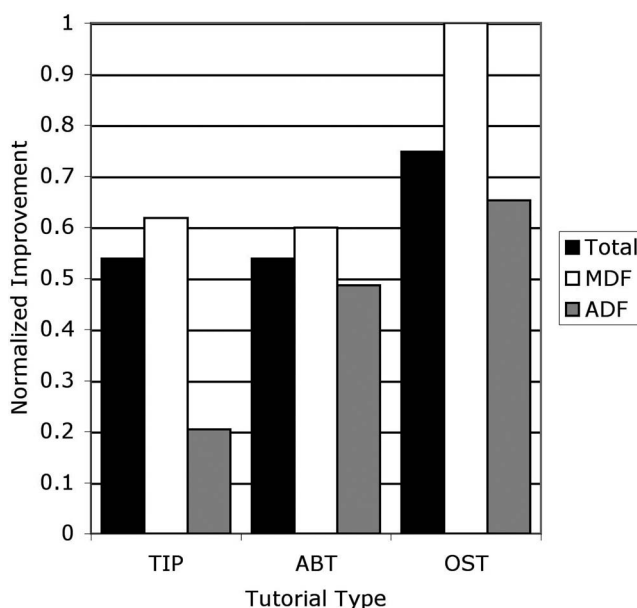


FIG. 3. FMCE data, overall N3 score. Comparisons of three curricula, with separate columns for normalized gain in the N3 cluster, improvement (fractional possible gain) in use of MDF, and improvement in use of ADF.

Second, the TIP and ABT tutorials are very similar, except in the improvement on use of the ADF. Here, the TIP students were far behind both the ABT and the OST students. Finally, all populations had the lowest improvements on use of the ADF. The FMCE questions on collisions allow for several ADF-guided responses, while the FMCE questions on pushing obviously include ADF scenarios (the one object pushing the other or resisting the push exerted on it). The ADF seems to be most difficult for students to address in their thinking.

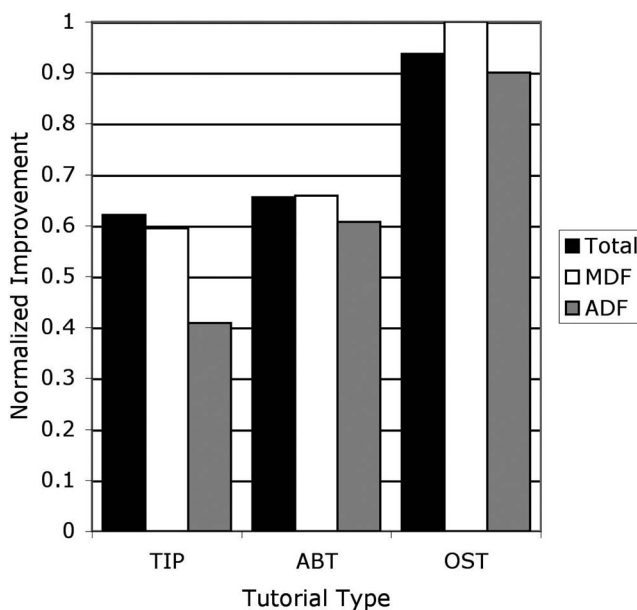


FIG. 4. FMCE data, collision situations in N3. Comparisons of three curricula, with separate columns for normalized gain on collision questions, improvement (fractional possible gain) in use of MDF, and improvement in use of ADF.

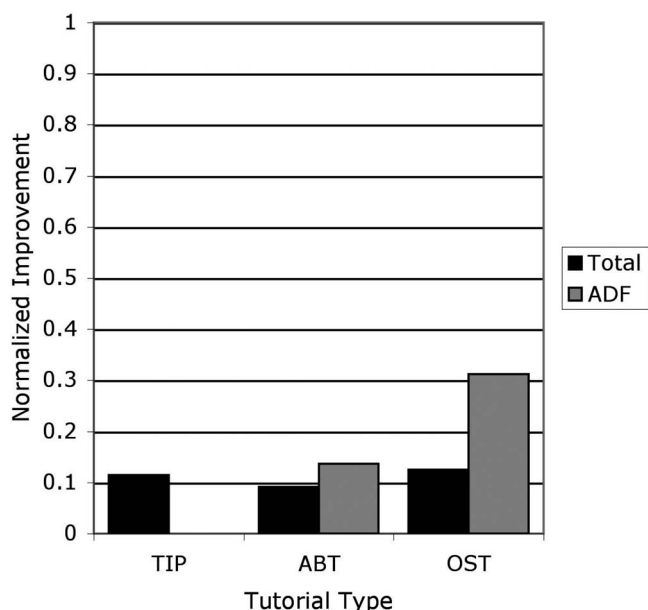


FIG. 5. FMCE data, pushing situations in N3. Comparisons of three curricula, with separate columns for normalized gain on pushing questions and improvement in use of ADF. Use of MDF is very rare in pushing situations, and only on pretests. Note that results are not statistically significant and are included for completeness only.

The data from Fig. 4 show similar results. On the collision questions, we find that the students in the TIP and ABT tutorials improve by roughly the same amount, except in the use of the ADF, while the students in the OST tutorials perform much better in every category. Perhaps this is to be expected, when the primary element of the OST tutorials is a collision between a truck and a car (though, as noted, not in the situation given in FMCE questions). We note that TIP students improved greatly on the collision questions (except in use of the ADF), which implies that learning about pushing situations in the tutorial helped these students learn about N3 in collision situations. Further exploration is required to understand the transfer of learning that may have occurred.

The data from Fig. 5 indicate that most of the improvement in the overall N3 scores came from improvements in the collision cluster, not in the pushing cluster. We note that the results are not significantly different in any of the tutori-

TABLE IV. Probability ( $p$ ) values on FMCE gains and improvements, computed using ANOVA. Values for the main effect and each *post hoc* comparison are displayed. Values displayed in italics are significant. For the pushing cluster and the overall ADF values, the main effect was well outside statistical significance; pushing cluster scores are not shown.

	Overall			Collision		
	Total	MDF	ADF	Total	MDF	ADF
Main effect	0.082	<i>0.008</i>	0.121	<i>0.011</i>	<i>0.004</i>	0.071
TIP vs ABT	0.240	0.124	XXX	0.113	0.087	0.169
TIP vs OST	<i>0.030</i>	<i>0.003</i>	XXX	<i>0.002</i>	<i>0.004</i>	<i>0.032</i>
ABT vs OST	<i>0.031</i>	<i>0.005</i>	XXX	<i>0.007</i>	<i>0.002</i>	<i>0.014</i>

als, though the zero fractional improvement in ADF use by TIP students is notable.

The data indicate that all tutorials show small improvement in the pushing cluster of N3 questions, even though all included a section on pushing. It seems that our modifications to the OST materials, adding a pushing situation, had little overall effect on student improvement in the ADF and normalized gain. Also, contrary to our expectations, students in the TIP tutorials improved primarily on collision questions, even though their materials did not include a collision situation.

VI. CONCLUSIONS

Our results show that the “refining raw intuitions” tutorial is more effective than either the TIP tutorial or the ABT tutorial at improving students’ understanding of Newton’s third law in an algebra-based physics course at UMaine. These results were consistent using multiple measures, including comparisons between types of facets used on ungraded pre-instruction quizzes and postinstruction graded assessments, and pre- and postinstruction use of the Force and Motion Conceptual Evaluation.

Improvement in student performance on FMCE questions evaluating understanding of Newton’s third law came primarily in the understanding of collisions and less in the area of objects pushing one another. All tutorials contained pushing situations (though the pushing scenario had to be added to the original OST materials). There were no meaningful differences between the groups in terms of improvement of their understanding of N3 in pushing situations, though. The FMCE data are inconsistent with the pretest-to-examination improvements, in which pushing situations were used to evaluate student learning. On the examination data, we found that all groups improved, but the students in OST tutorials most of all.

By analyzing results in terms of the mass dependence and action dependence facets, we were able to quantify the most common facets we found students using and describe the reasons for the student facets. We found that use of the ADF was least improved compared to MDF improvement and normalized gain of the overall N3 score (which also takes into account other responses, not just those motivated by the MDF and ADF). In particular, we found that students in the TIP tutorials showed the least improvement in the ADF, both in collision and pushing situations.

We have two seemingly contradictory suggestions for curriculum developers. Based on the FMCE data, we suggest that developers focus on pushing situations to help students learn N3 more effectively. This is the area with the most room for improvement. At the same time, we find that the ADF is the most common incorrect idea used by students. Our data suggest that addressing the ADF appropriately requires engaging students in more detailed discussions of collisions. A curriculum containing both pushing and collision situations is most likely necessary to help students develop a more complete view of N3.

ACKNOWLEDGMENTS

We thank the following people for their assistance in



implementing the tutorials and gathering data: Jeffrey Morgan, John Thompson, Eleanor Sayre, Andrew Paradis, R. Padraic Springuel, Michael Murphy, Brandon Bucy, Michael Dudley, Elizabeth Bianchi, Glen Davenport, Yokiko Muira,

Chelsea Lucas, and James Perry. We thank Gregory Smith for suggestions and advice regarding statistical methods. We thank Andrew Elby and five reviewers for their insightful comments and excellent questions.

- 
- <sup>1</sup>N. D. Finkelstein and S. J. Pollock, Replicating and understanding successful innovations: Implementing tutorials in introductory physics, *Phys. Rev. ST Phys. Educ. Res.* **1**, 010101 (2005).
- <sup>2</sup>L. C. McDermott and E. F. Redish, Resource letter: PER-1: Physics education research, *Am. J. Phys.* **67**, 755 (1999).
- <sup>3</sup>L. C. McDermott, Millikan Lecture 1990: What we teach and what is learned—closing the gap, *Am. J. Phys.* **59**, 301 (1991).
- <sup>4</sup>L. C. McDermott, Oersted Medal Lecture 2001: Physics education research—the key to student learning, *Am. J. Phys.* **69**, 1127 (2001).
- <sup>5</sup>L. C. McDermott, P. S. Shaffer, and the University of Washington's Physics Education Group, *Tutorials in Introductory Physics* (Prentice-Hall, Upper Saddle River, NJ, 2002).
- <sup>6</sup>E. F. Redish, J. M. Saul, and R. N. Steinberg, On the effectiveness of active-engagement microcomputer-based laboratories, *Am. J. Phys.* **65**, 45 (1997).
- <sup>7</sup>M. C. Wittmann, R. N. Steinberg, and E. F. Redish, *Activity-Based Tutorials Volume 1: Introductory Physics* (John Wiley & Sons, New York, 2004).
- <sup>8</sup>M. C. Wittmann, R. N. Steinberg, and E. F. Redish, Investigating student understanding of quantum physics: Spontaneous models of conductivity, *Am. J. Phys.* **68**, 218 (2002).
- <sup>9</sup>M. C. Wittmann, R. N. Steinberg, and E. F. Redish, Understanding and affecting student reasoning about sound waves, *Int. J. Sci. Educ.* **25**, 991 (2003).
- <sup>10</sup>D. Hammer and A. Elby, Tapping epistemological resources for learning physics, *J. Learn. Sci.* **12**, 53 (2003).
- <sup>11</sup>D. Hammer, Student resources for learning introductory physics, *Am. J. Phys.* **68**, (Physics, Education, Research, Suppl.), S52 (2000).
- <sup>12</sup>A. Elby, Helping physics students learn how to learn, *Am. J. Phys.* **69**, (Physics, Education, Research, Suppl.), S54 (2001).
- <sup>13</sup>This was found to be true for students at UMaine and has been reported anecdotally at other institutions.
- <sup>14</sup>M. C. Wittmann, R. N. Steinberg, and E. F. Redish, Making sense of students making sense of mechanical waves, *Phys. Teach.* **37**, 15 (1999).
- <sup>15</sup>L. C. McDermott and P. S. Shaffer, Research as a guide for curriculum development: An example from introductory electricity. Part I: Investigation of student understanding, *Am. J. Phys.* **60**, 994 (1992).
- <sup>16</sup>We make explicit that one author, M.C.W., is lead author on the published ABT materials (Ref. 7). The original authors of the tutorial were Jeffery T. Saul, John Lello, Richard N. Steinberg, and Edward F. Redish. We used the published version as revised by Michael C. Wittmann and Richard N. Steinberg.
- <sup>17</sup>R. K. Thornton and D. R. Sokoloff, Assessing student learning of Newton's laws: The Force and Motion Conceptual Evaluation and the evaluation of active learning laboratory and lecture curricula, *Am. J. Phys.* **66**, 338 (1998).
- <sup>18</sup>The original analysis was carried out as part of a senior thesis by one author, T.I.S., and used analysis methods he created. Only later did we revise this analysis into the language used in this paper.
- <sup>19</sup>D. Hammer, E. F. Redish, A. Elby, and R. E. Scherr, in *Resources, framing, and transfer*, in *Transfer of Learning: Research and Perspectives*, edited by J. Mestre (Information Age Publishing, Charlotte, NC, 2004).
- <sup>20</sup>J. Minstrell, *Facets of students' knowledge and relevant instruction in Research in Physics Learning: Theoretical Issues and Empirical Studies, Proceedings of an International Workshop, Bremen, Germany 1991*, edited by R. Duit, F. Goldberg, and H. Niedderer (IPN, Kiel, 1992), pp. 110–128.
- <sup>21</sup>L. Bao, K. Hogg, and D. Zollman, Model analysis of fine structures of student models: An example with Newton's third law, *Am. J. Phys.* **70**, 766 (2002).
- <sup>22</sup>J. Clement, Students' preconceptions in introductory mechanics, *Am. J. Phys.* **50**, 66 (1982).
- <sup>23</sup>Since we are later interested in changes in numbers of facets used and are not considering percentages, the difference between six or seven possible total facets used is not relevant to our analysis.
- <sup>24</sup><http://perlnet.umaine.edu/materials/>
- <sup>25</sup>Ron Thornton (private communication).
- <sup>26</sup>R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, *Am. J. Phys.* **66**, 64 (1998).