Conceptual Gains with the Synthetic Groupwork-based Newton's Third Law Open Source Tutorial

Michael M. Hull and Keshab R. Pokharel
University of Alaska Fairbanks, College of Natural Science & Mathematics, Fairbanks, Alaska 99775, USA

The University of Maryland's Open Source Tutorial (OST) on Newton's Third Law (N3) has demonstrated a significant impact on boosting student conceptual understanding of N3, even in comparison to other research-based curricular material. Although OSTs were originally designed as group-based guided worksheets, the first author has developed and implemented an online asynchronous alternative: the N3 Synthetic Groupwork-Based OST (SynG-OST). Here, we report data from a pilot study with the module that suggest it may be comparably effective to the in-person OST. Furthermore, we compare learning gains with those obtained with an interactive video-based lecture guiding students through the worksheet. Preliminary data indicate that student learning is greater with the N3 SynG-OST than with the lecture-based version. This strengthens the argument that synthetic groupwork-based materials may be a viable option to replace in-person groupwork for asynchronous learners. It can also be assigned—as we have done—as homework for in-person classes.

2025 PERC Proceedings edited by Pawl, Zwolak, and Leak; Peer-reviewed, doi.org/10.1119/perc.2025.pr.Hull

I. INTRODUCTION

With advances in AI technology, synthetic tutors are playing an increasing role in education [1–4]. There are several benefits found in groupwork, however, that should not be forgotten. From a social constructivist perspective, students may learn best within a Zone of Proximal Development [5]. Groupwork allows weaker students to learn from someone who is not so advanced, and so the explanations of the stronger students are more likely to resonate. Brown *et al.* [6] argued that groupwork results in insights and solutions that might not arise without the group since groupwork enables students to explore multiple viewpoints, teasing them apart from each other, and discerning which make sense and are worth pursuing. Furthermore, whereas direct instruction and Socratic questioning can constrain search for novel solutions, groupwork offers a greater freedom which can stimulate inventiveness and creativity [7].

Many large-enrollment physics classes in the USA have students work in groups, for example using Tutorials in Introductory Physics (TIPs) [8], Open Source Tutorials (OSTs) [9], or Investigative Science Learning Environment (ISLE) [10]. There are, however, students who learn physics completely online, and there is a need to equitably enable their access to the benefits of groupwork. At the University of Alaska Fairbanks (UAF), approximately 50 students enroll each fall semester in online asynchronous (*i.e.*, completed on one's own timing) algebra-based introductory mechanics (referred to in this paper as AlgPhys). Many of these students take the course as part of an asynchronous Biology and Wildlife BS program, and include students who are in the military, work full-time, or/ and are stay-at-home parents. To make the benefits of groupwork accessible to these learners, the first author has been developing videos of fictitious classmates in the form of cartoon-like avatars working in groups on a guided tutorial. The video pauses frequently to elicit an answer from the learner in response to the group discussion and continues once the learner has submitted an answer. The first author assigns these Synthetic Groupwork-based Tutorials (SynG-Tuts) as homework also to students in CalcPhys, the calculus-based version of the class (predominantly engineering majors) which meets in-person. Although students in CalcPhys complete group-based activities in-person as well (Tutorials, etc.), the use of SynG-Tuts provides additional group-based experiences to students. We have previously reported on SynG-IiR [11,12], a SynG-Tut which is based upon Inquiry into Radioactivity [13]. Here, we report on our second SynG-Tut, which is based upon the Open Source Tutorial on Newton's Third Law (N3 SynG-OST, pronounced "seen ghost").

Our overall research question is "how useful is synthetic groupwork-based instruction?" In this paper, we compare scores to FCI [14] and FMCE [15] items targeting understanding of N3 administered pre and post-instruction with those reported in the literature [16,17] to address our first research question: 1) **how do learning gains with the N3 SynG-OST compare with the original in-person N3 OST?** We will compare student conceptual growth with the N3 SynG-OST to data obtained with an interactive video-based lecture of the first author guiding students through the N3 OST (without the synthetic groupwork) the preceding year. This will allow us to address our second research question: 2) **how important is the group-based aspect of the N3 SynG-OST?** Finally, we will analyze responses to a survey administered post instruction to address our third research question: 3) **to what extent do students enjoy learning with the N3 SynG-OST?**

II. METHODOLOGY

The N3 OST, like OSTs in general, is a guided worksheet that students complete in groups of four, with an ideal ratio of one TA per three student groups, such that the TA is not always immediately available but is able to intervene at critical points. The N3 OST has students consider a massive truck colliding with a stationary car. Students are asked to compare the force exerted by the truck on the car with that exerted by the car on the truck. Typically, students correctly intuit that the car will react more, but they incorrectly attribute the cause of this to be a larger force from the truck. Using force probes attached to carts of dramatically different weights on a low-friction track, students see that the forces are in fact equal. The Tutorial then has students consider the intuitive idea that the car reacts more in terms of speed change of the two vehicles (if the car is half the mass, it speeds up twice as much), how the accelerations compare (car's is twice the truck's), and, from Newton's 2nd Law, how the forces compare (they are equal). In this way, students refine their intuitions about force and acceleration with the ultimate aim of building the habit of pursuing such refinement in their physics studies [18].

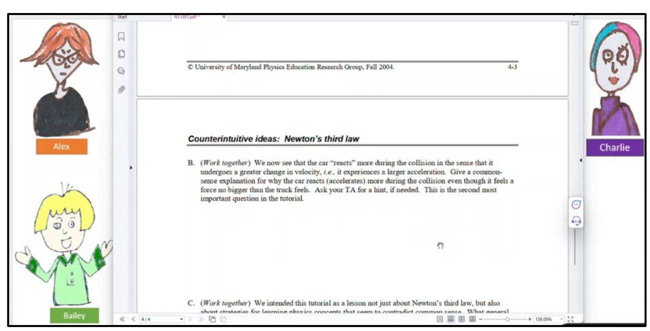

FIG. 1. Synthetic classmates Alex, Bailey, and Charlie discuss Newton's Third Law in the synthetic groupwork-based Open Source Tutorial.

In Fall 2023, students in AlgPhys and CalcPhys were invited to participate in an interactive lecture-based video walking the individual learner through the N3 OST. This video was administered through PlayPosit, which paused the video at each question on the worksheet until the respondent entered an answer to that question. In 2024, this activity was replaced with the N3 SynG-OST, which had Alex, Bailey, and Charlie discussing the worksheet and asking the learner for his or her input at each question (Fig. 1). Other than this difference, the two modalities of the asynchronous N3 OST were comparable.[1]

To assess the N3 SynG-OST, a survey was constructed that included the items assessing N3 from the FCI and FMCE. To address the first research question, student scores on these items were compared with [17] for the FCI post-test items and with [16] for the FMCE normalized gain, which is the gain divided by the maximum possible gain. To address the second research question in comparing the two versions of the asynchronous N3 OST, the questions developed by Sampaio-Kronister [19] were also added, such that the conceptual survey totaled 14 items. To address the third research question, the post-test survey in 2024 contained not only the 14 conceptual items, but also four items asking for student impressions of the N3 SynG-OST. Incentive to participate was extra credit worth ~1% of the overall course grade. However, only 27 students from AlgPhys (~50% of students enrolled at the time) and 24 students from CalcPhys (~25%) used the N3 SynG-OST. Only 10 of these students in AlgPhys and 16 students in CalcPhys also answered the N3 survey items prior to and after the N3 SynG-OST. One of these students in AlgPhys took the pretest survey after completing the Tutorial and was thus removed from the data pool. This left a total of $N = 25$ respondents from 2024 (and $N = 22$ respondents from 2023) in the analysis that follows. All methods were approved by UAF IRB.

III. RESULTS

A. Students can learn from the OST on their own…

Although this is only a pilot study with a small data sample ($N = 25$ total), our findings so far suggest that the N3 SynG-OST may be comparable to the original in-person Tutorial at improving student conceptual understanding. Regarding the four items on the FCI (Q.4, Q.15, Q.16, and Q.28), the average post-test score in AlgPhys was 91.7% and in CalcPhys was 92.2%. This can be compared with Fig. 6 on pg. 12 in [17], which shows post-test scores around 85% on these items. The normalized gain for these 25 students on these four items was 0.81. Since the FCI was administered online, a critic may question whether students used AI to procure answers to the FCI (although it was emphasized that the extra credit was not based upon correctness). However, the normalized gain for these students on the FCI overall was substantially smaller: 0.50, which is average for interactive engagement classes in Hake's study [20]), which makes such cheating unlikely.

There are six FMCE items that were used in the analysis of [16]: Q.30, Q.31, Q.32, Q.34, Q.36, and Q.38. On these six items, the normalized gain with the N3 OST was about 0.52 for AlgPhys and 0.85 for CalcPhys (0.74 combined). This can be compared with Fig. 3 on pg. 6 of [16], showing the normalized gain with the in-person N3 OST to be about 0.75. In summary, the 25 students who learned with N3 SynG-OST improved their learning comparably well to students taking the in-person OST. Although this is certainly too few students to answer our first research question definitively, these preliminary results warrant further data collection.

B. … especially if there is simulated groupwork.

It is possible that the learning gains obtained with the N3 SynG-OST are the result of the very convincing experiment of a truck crashing into a car after eliciting the student common sense idea that the force from the truck should be greater. It is also possible that the learning gains are due to the demonstration of how the intuitive idea about the car gaining more speed leads to the correct prediction of the forces being equal. If these factors are exclusively the cause of the learning gain, then it is feasible that the same learning gains could be obtained in a synthetic tutor modality, without the fictitious classmates. This leads us to our second research question, "how important is the group-based aspect of the N3 SynG-OST?"

To address this question, we draw upon the 2023 data, where an interactive video-based lecture of the first author guiding students through the N3 OST (without the synthetic classmates) was used. Our original intent was to use Rasch analysis [21,22] of survey results from the 14-item survey administered pre- and post-Tutorial both years. Unlike raw score or normalized gain (which uses raw score), Rasch gain does not assume survey items to be of equal difficulty and thus provides a more accurate measure of ability [23]. According to Fisher [24], a rating scale instrument is poor if there are more than 5% ceiling or floor responses. With the introduction of the N3 survey items from Sampaio-Kronister [19], we were able to prevent scores of 0 points. However, perhaps because the correct answer to all 14 items was "the forces are equal", there were more than 5% ceiling responses even on the pretest. Unsurprisingly, we found that the model was not a good fit when analyzed in R Studio (p-value of Chi-square is $0.04 < 0.05$).

Despite the limitations of the approach, as a first step, we compared normalized gains on the four FCI items and on the six FMCE items pertaining to N3. Although the sample sizes

[1] The authors agreed that the next greatest difference is that the N3 SynG-OST has students predict the results of the collision experiment before it happens. However, only one student incorrectly predicted that N3 would not apply to the collision (and that student's FCI score went from 3/4 to 4/4 and FMCE did not increase).

are small, we do find that the learning gains were greater with the synthetic groupwork format. With the lecture-based interactive OST, on the six FMCE items, normalized gain was only 0.46; on the four FCI items, normalized gain was only 0.58 (post-test was 69%). These values were statistically significantly lower than the N3 SynG-OST values discussed above, as shown in a two-sample one-tailed t-test (p value is less than 0.05, as indicated in Table I). Cohen's d is medium ($d > 0.5$) or large ($d > 0.8$).

TABLE I. Learning gains with the N3 OST. Reference values from the original in-person OST are on the left, followed by the interactive video-based lecture (2023) and the N3 SynG-OST (2024), with sample standard deviations.

Metric	Reference value	Interactive lecture (N = 22)	SynG-OST (N = 25)	p	d
FCI post-test	85% [17]	69% ±29%	92% ±19%	0.001	0.86
FMCE normalized gain	0.75 [16]	0.46 ±0.5	0.74 ±0.4	0.035	0.58

To complement these quantitative findings, the first author coded responses to two prompts for whether he deemed intervention would be needed should a group write such an answer during an in-person Tutorial. These prompts were chosen based upon the N3 OST Instructor's Guide [9] and personal experience teaching with the N3 OST for over a decade. The first prompt coded reads "Does the intuitive guess about the car's gain in speed agree or disagree with Newton's third law? How do you know?" Here, instructors are to "ensure that the calculations are correct and that the students see agreement between N3 and their intuition." The second prompt chosen was the second-to-last prompt asking for "a common-sense explanation for why the car reacts (accelerates) more during the collision even though it feels a force no bigger than the truck feels." Each response was coded out of 2 points. For the latter prompt, for example, a student who answered with "The reason would be because the car is starting from 0 velocity which means it has a bigger change of velocity over time" received 0 points, indicating that instructor intervention would be crucial. A student who responded with "less mass to spread the force over" received 2/2 points. On this prompt, most students answered with something like "mass is less for the car, so acceleration is more", which received 1/2 points for being correct, but not a self-contained "common-sense" explanation. Although differences were not statistically significant, scores were somewhat higher in 2024 (for example, 92% of respondents received 2/2 points on the latter prompt in comparison to 82% in 2023).

C. Most students like the lesson

At the beginning of the post-test survey in 2024, students were asked three Likert-scale questions regarding their impressions of the N3 SynG-OST:

- *How did you find the online tutorial on Newton's Third Law? … extra credit is based only upon completion of the survey, not on what you answer to the questions, so please be honest!* (results in Fig. 2)
- *How did you feel about the "synthetic groupwork", where pre-recorded students contributed their ideas as you went through the worksheets?* 1: "It was too artificial to be helpful" → 5: "I found it just as good as normal groupwork"
- *Would you like to have more of your learning in physics be in this "synthetic groupwork"-based worksheet style?* 1: "No, I strongly prefer the more traditional lecture style" → 5: "Yes, this is a great way to learn!"

Students were also given the free response prompt "*Any and all comments about the online tutorial are appreciated here. In particular, what would you change for next time, and what would you be sure to keep the same?*" Here, data was used from all 51 students who completed the N3 SynG-OST, whether they took the N3 conceptual survey or not.

Responses were generally favorable. Students reported finding the N3 SynG-OST to be enjoyable, the synthetic groupwork to be comparable to in-person groupwork, and that more instruction should take this format. Students in the face-to-face CalcPhys class were less enthusiastic than counterparts in AlgPhys, but this difference was not statistically significant, as found by a Mann-Whitney U Test on each question ($p > 0.05$).

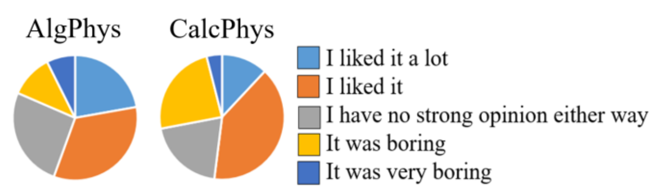

FIG. 2. More than half of the students who learned with the N3 SynG-OST ($N = 51$) reported enjoying it.

Feedback collected from the free response prompt included the following criticism:

- *I don't want to wait on people spouting out the wrong answers when I already have the right one. Let me skip 10 seconds at a time on the videos so I can pick what I hear instead of listening less overall*. (CalcPhys)
- *I like the idea of trying to replicate going through a worksheet but I think it would be better to have one person reading through it and explaining it trying to replicate a classroom expanse was just too distracting…* (CalcPhys)
- *It just felt really weird, I like the personal connection that comes with people, not ai.*(CalcPhys)

Evidently, although some of this criticism (*e.g.*, the third comment) we would expect to hold also with synthetic tutors, other statements reflected the views of students who would have likely preferred the interactive lecture-based Tutorial offered the previous year in 2023.

Feedback also included the following encouraging responses:

- *I think the dialogue and slow-paced reasoning through Newton's third law should give a great in-depth understanding to all viewers.*(CalcPhys)
- *… I liked how there was different opinions on how to solve the problems made it easier for me to learn.*(AlgPhys)
- *… Letting the group get the wrong answer and showing the right answer is one of the best ways to learn. … In this case, with a low impact of failure, allowing the group to fail and then demonstrate how to change it into a success leaves the group feeling positive about the knowledge they gained.* (AlgPhys)

Again, it should be noted that some of these comments (*e.g.*, the third) would likely hold also if the learner had used the video-based lecture instead of the N3 SynG-OST. Others, however, specifically showed value for the "dialogue" and "different opinions". It should also be noted that, although not explicitly asked, two respondents in AlgPhys offered the opinion that they prefer synthetic groupwork to in-person groupwork (which was not used in AlgPhys):

- *I prefer the video over in-person groupwork.*(AlgPhys)
- *Even though artificial, I enjoyed this more than actual group work as if I make a mistake I don't panic over what other's may think of me…*(AlgPhys)

IV. LIMITATIONS AND FUTURE WORK

There are several important limitations in our study that should be kept in mind. First and foremost, our sample size is disappointingly small. The majority of enrolled students opted *not* to take the N3 SynG-OST, even with the incentive of extra credit, which may reflect low student buy-in to the idea of synthetic groupwork. On the other hand, it is possible that students did not participate because they perceived the extra credit offered as insufficient to justify the time demands (extra credit was given only if pretest, Tutorial, and post-test were completed). In the future, the N3 SynG-OST will be assigned as homework, motivated by our findings.

Above, we discussed how it seems unlikely that the high FCI and FMCE scores of the students who *did* participate are due to cheating, since their overall FCI gains were significantly smaller. Nevertheless, those overall FCI gains were larger than the typical value obtained in the first author's classes, which may represent a selection bias. Future work is planned to collect additional data and compare gains on each FCI cluster between students who take the N3 SynG-OST and those who do not. This future work, which will compare student scores on the full FCI, will utilize Rasch analysis.

In qualitatively analyzing responses to the N3 SynG-OST, we did not analyze the "most important question" (the last one, which asks students "What general strategies are suggested by this tutorial—strategies you might be able to use with counterintuitive concepts appearing later in the course?" Future work should look at student responses to this question and changes in student views about the nature of physics knowledge and learning, as measured by the MPEX [25] and/or CLASS [26]. Although it is unlikely that a single Tutorial will make students more expert-like in this regard [18,19], it should be remembered that a primary goal of OSTs is to develop these attitudes [18].

The current version of the N3 SynG-OST does not employ jump points. Future versions will combine free response questions with follow-up multiple choice questions that will direct learners to different parts of the video, depending on which response they choose, as we have done with SynG-IiR [11]. In later versions, we expect to be able to use AI to jump students directly from their free response answers.

V. DISCUSSION AND CONCLUSION

In our study, we created and implemented an online asynchronous module utilizing synthetic groupwork for individual learners to go through the Open Source Tutorial on Newton's Third Law (N3 SynG-OST). Addressing our first research question, we found that the 25 students who learned with the N3 SynG-OST came to an understanding of N3 that has been reported in literature for students learning with the in-person OST. By comparing participant responses to FCI and FMCE items pertaining to N3, we found that learning was statistically significantly greater than the previous cohort of 22 students, who were guided through the N3 OST by an interactive video-based lecture. This finding is important and should inform future directions of AI-based instruction. Typically, students interact with educational AI in the form of a synthetic tutor. Although the tutor can be constrained such that it does not immediately give answers to students but rather engages in Socratic dialogue, this nevertheless removes the opportunity for the student to learn vicariously [27]. Regarding our third research question, we found that the majority of learners had positive reports about learning with the N3 SynG-OST, including statements indicating awareness of the benefits of learning from classmates (*i.e.*, within a Zone of Proximal Development [5]), who have "different opinions on how to solve the problems". Although synthetic groupwork is most appropriate for asynchronous classrooms, we have shown that it can be used to benefit also in-person physics learners when administered as homework.

ACKNOWLEDGEMENTS

The authors thank Shizuka Nakayama for graphic design.

[1] S. Khan, How AI Could Save (Not Destroy) Education. TED (2023).
[2] aiPlato, Democratizing 1-on-1 education. Https://Aiplato.Ai/ (2023).
[3] R. P. dos. Santos, Enhancing physics learning with chatgpt, bing chat, and bard as agents-to-think-with: A comparative case study. ArXiv Preprint ArXiv:2306.00724 (2023).
[4] Pearson, Get instant homework help from our AI tutors. Https://Www.Pearson.Com/Channels/Ai-Tutor (2025).
[5] L. Vygotsky, Interaction Between Learning and Development. in *Readings on the Development of Children*, edited by Gauvain and Cole (Scientific American Books, New York, 1978), pp. 34–40.
[6] J. S. Brown, A. Collins, and P. Duguid, Situated cognition and the culture of learning. Educational Researcher **18**, 32 (1989).
[7] M. Kapur, Examining productive failure, productive success, unproductive failure, and unproductive success in learning. Educ Psychol **51**, 289 (2016).
[8] L. C. McDermott and P. S. Shaffer, Tutorials in Introductory Physics. *Tutorials in Introductory Physics*, 1st ed. (Prentice Hall, Upper Saddle River, NJ, 2002).
[9] A. Elby, R. Scherr, T. McCaskey, R. Hodges, E. F. Redish, D. Hammer, and T. Bing, Open Source Tutorials in Physics Sense-making. *Open Source Tutorials in Physics Sense-Making* (DVD, funded by NSF DUE-0341447, 2007).
[10] E. Etkina, D. T. Brookes, and G. Planinsic, Investigative Science Learning Environment: When learning physics mirrors doing physics. *Investigative Science Learning Environment: When Learning Physics Mirrors Doing Physics* (Morgan & Claypool Publishers, 2019).
[11] M. M. Hull and K. R. Pokharel, Learning about Radioactivity via Synthetic Groupwork in Introductory Physics. Physics Education Research Conference (2024).
[12] M. M. Hull, M. Jeidler, and E. S. Holzinger, Simulated Groupwork in an Asynchronous Course Learning about Radioactivity. in *Physics Education Research Conference Proceedings* (2023).
[13] A. Johnson, Radiation and Atomic Literacy for Nonscientists. Science IBI Series Winner **342**, (2013).
[14] D. Hestenes, M. Wells, and G. Swackhamer, Force Concept Inventory. Phys Teach **30**, 141 (1992).
[15] 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).
[16] T. I. Smith and M. C. Wittmann, Comparing three methods for teaching Newton's third law. Physical Review Special Topics-Physics Education Research **3**, 020105 (2007).
[17] E. Redish and D. Hammer, Reinventing college physics for biologists: Explicating an epistemological curriculum. (2009).
[18] A. Elby, Helping physics students learn how to learn. Am J Phys **69**, S54 (2001).
[19] I. N. Sampaio-Kronister, The conceptual and epistemological effects of Open Source Tutorials when teaching Newton's third law in Austrian high schools. The Conceptual and Epistemological Effects of Open Source Tutorials When Teaching Newton's Third Law in Austrian High Schools, MS Thesis, University of Vienna, 2020.
[20] 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).
[21] T. Bond and C. Fox, Applying the Rasch Model. Fundamental Measurements in the Human Sciences. *Applying the Rasch Model. Fundamental Measurements in the Human Sciences*, 2nd ed. (Erlbaum, Mahwah, NJ, 2007).
[22] W. J. Boone, J. R. Staver, and M. S. Yales, Rasch analysis in the human sciences. *Rasch Analysis in the Human Sciences* (Springer Science & Business Media, 2013).
[23] M. Planinic, W. J. Boone, A. Susac, and L. Ivanjek, Rasch analysis in physics education research: Why measurement matters. Phys Rev Phys Educ Res **15**, 020111 (2019).
[24] W. Fisher, Rating Scale Instrument Quality Criteria. Rasch Measurement Transactions: Https://Www.Rasch.Org/Rmt/Rmt211m.Htm **21**, 1095 (2007).
[25] E. F. Redish, J. M. Saul, and R. N. Steinberg, Student expectations in introductory physics. Am J Phys (1998).
[26] W. K. Adams, K. K. Perkins, N. S. Podolefsky, M. Dubson, N. D. Finkelstein, and C. E. Wieman, New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. Physical Review Special Topics - Physics Education Research **2**, 1 (2006).
[27] D. J. Rosen and A. M. Kelly, Working together or alone, near, or far: Social connections and communities of practice in in-person and remote physics laboratories. Phys Rev Phys Educ Res **18**, 010105 (2022).