

Implementing a mixed-methods approach to understand students' self-efficacy: A pilot study

Rachel Henderson

*Department of Physics & Astronomy, Michigan State University, East Lansing, MI, 48824 and
CREATE for STEM Institute, Michigan State University, East Lansing, MI, 48824*

Vashti Sawtelle

*Lyman Briggs College, Michigan State University, East Lansing, MI, 48824 and
Department of Physics & Astronomy, Michigan State University, East Lansing, MI, 48824*

Self-efficacy (SE) – the confidence in one’s own ability to complete a task – has been shown to be an important predictor of student success in the sciences. Traditionally, SE is measured using pre/post instruction surveys or retrospective interviews. In this study, we present a pilot of a mixed-methods approach to studying SE. This approach combines a fine-grained quantitative measure with close to the event qualitative reflections. The results presented from this initial study include five students who transferred credit from a two-year college into a bachelor’s program to study physics or astronomy at a large mid-western research university. We report on the incentive structure of the methodology and threats to the integrity of the data. In general, the findings indicate that our methodology yields interesting results and is ready to scale to a larger study.

I. INTRODUCTION

Students' self-efficacy (SE), or the confidence in one's own ability to complete the actions necessary to perform a task [1], has been shown to be a predictor of students' achievement in science courses [2–4], their persistence in science majors [5–9], and a critical aspect with respect to the science career choices that students make [10–12]. Historically, students' SE in science courses has been studied using surveys [13–16] that assess students' confidence in their own ability at the beginning and end of a course. For example, in physics, Sawtelle *et al.* [17] and Marshman *et al.* [18] found that, for physics students, their overall SE decreased from pretest to post-test. However, for students who major in chemistry, mathematics and biology, post-instruction SE is higher than pre-instruction SE, demonstrating an overall increase in SE over the duration of a course. [6, 19, 20].

While pre/post survey measures have shown shifts in students' SE, we also know that SE is variable that fluctuates over time [21]. Most measures of SE rely on participants reporting on events that happened a long time after the event occurred (see review in Sawtelle *et al.* [22]). While there have been studies investigating how an individual's experiences may influence a student's SE [23–25], the time between the experiences and the measurement means that researchers miss important features of these experiences. Bandura [1] posited four kinds of experiences to influence SE, but in order to know more precisely *how* those events impact SE we need measures to examine SE in-the-moment.

In this study, we begin the process of building a mixed-methods approach that blends a fine-grained quantitative measure with close to the event qualitative reflections. This approach explores students' changing SE to understand what critical experiences may impact the longitudinal development of SE. Since this design is a novel approach, here we present our results from a pilot study which investigated two research questions: (1) *Is there evidence that the incentives were sufficient for students to participate fully in the research study?* and (2) *Is there evidence of any threats to the integrity of the data such as attrition or sampling fatigue?*

As we recognize that the number of students in our pilot study is relatively small, we do not intend to make any generalizations about a student's SE. Rather, by answering these research questions, we can begin to understand the benefits and limitations of this methodology – critical components to the success of future larger-scale research studies. We will discuss our future plans for implementing this novel approach to further understand the development of students' SE.

II. METHODS

The sample was collected at a large, research-intensive, university during the Spring 2020 academic semester. The students in this study were physics majors who transferred at least one credit from a two-year college and were currently

enrolled in an upper-division physics course. In the future, we would like to implement this novel approach in the two-year college setting, which is why we chose to pilot this methodology with the students who have experience in that environment. All students meeting these criteria were invited to participate in the study ($n = 9$), and 5 students volunteered. The mixed-methods approach follows an explanatory design in which the quantitative and qualitative data provide equal weights of evidence [26]. Data was drawn from two sources: (1) the Experience Sampling Method (ESM), and (2) retrospective interviews; below we describe the details of each of these methodologies.

During our second week of scheduled data collection, the university moved to remote learning due to the COVID-19 pandemic. Our data collection was already running during this transition and, therefore, four students continued to respond without any direction from the research team, while one did not. As a result, our team modified the structure to give the entire incentive (described below) to the participants who were actively responding to the ESM notifications before the university moved to remote learning. Though in these circumstances, only 2 of the 5 students elected to participate in the reflective interviews; the insight these conversations gave in understanding the dynamic nature of student SE was a valuable part of research process.

A. Experience Sampling Method

The quantitative data was collected using the Experience Sampling Method (ESM) [27] via the mobile phone application, LifeData [28]. The ESM has the ability to collect high-frequency, self-reported data while students are engaged in a variety of tasks and experiences throughout their day, and thus enables researchers to study individuals' SE throughout their daily lives. The ESM is largely used in the field of psychology for clinical research studies such as treatments for addiction, job satisfaction, and internet usage [29–31]. More recently, Nissen and Shemwell introduced this method within the context of various STEM courses to show that women's SE on tasks within physics courses were significantly lower than in their other STEM courses [21]. Assessment error is greatly reduced due to using a repeated measures methodology that takes place in one's natural environment. [27, 32].

In this study, participants were notified via the LifeData application randomly four times throughout the day (9:00am-6:00pm) over two work weeks (Monday-Friday). The participants received daily notifications for one week, followed by a week break, and then another week of daily notifications; resulting in a total of 40 notifications for each participant. We incentivized participation in this high-frequency data collection with \$50 for responding to 80% of the notifications.

For each of the notifications, a short, 9-question survey was given to the participants. The larger goal of our study is to examine the interaction between task-specific and more general domains of SE. The small pilot study presented here,

TABLE I. The order of the ESM survey questions and items given to the research participants via the LifeData mobile phone application. Q1 and Q2 are open-response questions, Q3 is yes/no question, and Q4 through Q9 are 5-point Likert scale items.

Survey Block	ESM Question/Item
Task	(Q1) What is the main thing you are doing? (Q2) Where are you doing this activity? (Q3) Is the activity you are doing for a course you are currently enrolled in?
Task-Level SE	(Q4) How skilled are you in the activity? (Q5) Do you feel in control of the situation? (Q6) Are you succeeding at what you are doing?
Course-Level SE <i>or</i>	(Q7a) Considering the difficulty of this course, the instructor, and my skills, I think I will do well in this course. (Q8a) I'm confident I can do an excellent job on assignments and tests in this course. (Q9a) I'm certain I can master the skills being taught in this course.
Career-Level SE	(Q7b) Considering the difficulty of my planned profession and my skills, I think I will do well once I begin working in the profession. (Q8b) I'm confident I can do an excellent job on professional assignments once I am working in my chosen occupation. (Q9b) Once I am working in my chosen career, I am confident that I will master the skills that will make me successful.

however, had an emphasis on the integrity of the methodology. We provide an overview of the questions here to give the reader an understanding of the experience for the participants.

Table I outlines the questions and items used throughout this study. Each notification prompted a participant to answer Q1 through Q6. These questions asked participants about what activity or task they were doing (Q1), where they were doing the activity (Q2), and if the activity was for a course (Q3). The participants were then asked three, 5-point Likert scale questions about their SE toward the task that they were engaged in (Task-level SE; Q4-Q6). In addition to these six questions, a participant was prompted with three additional 5-point Likert scale items intended to measure their SE toward *either* their course (Course-level SE; Q7a-Q9a), *or* their intended career or profession (Career-level SE; Q7b-Q9b).

The questions to measure a student's SE toward a task, their courses, and their intended career were previously used and validated. The Task-level SE survey block (see Table I) included questions about the amount of skill, control, and success that a student experienced during a specific activity. These questions have been shown to form a valid and reliable measure of task-specific SE [21, 33, 34]. The Course-level SE items and the Career-level SE items were adapted from an existing instrument that was validated in the context of a calculus-based introductory physics course [35]. These items were developed using a modified version of the "Self-Efficacy for Learning and Performance" sub-scale of the Motivated Strategies for Learning Questionnaire (MSLQ) [13, 36].

B. Retrospective Interviews

In this pilot study, we also conducted one set of retrospective interviews after students had participated in the ESM

data collection. Due to the COVID-19 pandemic, these interviews did not occur *immediately* following the completion of the ESM data, but in the month following. These interviews were designed as episodic, semi-structured interviews to reflect upon student experiences. The larger study has the goal of structuring interviews to reflect on events that were also measured with the ESM tool. In this initial study, we did collect some of these data, but we primarily explored students' experience with the ESM methodology.

The interview protocols had three primary sections: the first unpacked students' experience with the ESM data collection tool, the second explored the incentive structure for participating in the study, and the third asked about the student's specific responses to the ESM prompts and events surrounding those responses. Table II provides example prompts used in the three sections of the interviews. The latter of these sections represents the style and type of interview prompts we plan to scale beyond the pilot study.

Each interview was conducted via videoconferencing software and recorded. In preparation for the interview, the research team had pulled together plots similar to that shown in Fig. 1 (described in Sec. III). This allowed the interviewer to reference specific responses and patterns in responses.

III. RESULTS

In this pilot study, we explored the advantages and limitations to the methodology, with an emphasis on the research incentive and the integrity of the data that was collected. Due to the turbulent transition to remote learning from the COVID-19 pandemic, only the results from the first week of data collection are presented below. Also due to the pandemic, only two respondents (Respondent 980 and Respon-

TABLE II. Example interview prompts used in the retrospective interviews.

Interview Target	Protocol prompt
ESM Experiences	(1) Take me back to the week when you were answering the LifePak [ESM] questions for the first time. Walk me through that experience. What stands out to you? (2) How did you feel about the questions? (3) How long did it generally take you to respond to a notification? (4) Did you ever wait to answer the questions because of what you were in the middle of doing during the notification? If yes, what task did you end up thinking about when answering the questions?
Incentives	(1) How was the incentive? Did it feel like the right level in terms of how much time you spent on the surveys? (2) For that same amount of money, would you do it again? (3) Was there ever a point where you got tired of answering the questions, but kept doing so for the incentive? (4) Would you participate again or for longer for the same incentive?
Responses to Notifications	(1) You answered differently between the <i>in control</i> question (Q5) from the ways you are <i>succeeding</i> (Q6). How were you deciding how to answer that question? How did you decide to say I don't feel in control, or I do? (2) Could you tell me in your own words how you were interpreting question four, how skilled are you in the activity?

dent 987) participated in the reflective interviews, the ESM data from these two respondents are presented below.

Figure 1 shows each student's ESM Task-level SE over the time period in which data was collected; each plot represents the responses over time for an individual student. The solid red lines, the dotted green lines, and the dashed blue lines are the 5-point Likert scale results of Q4, Q5, and Q6, respectively. The black squares indicate if a participant did not respond to that particular notification (NR=No Response).

A. Research Incentive

In our design of this pilot study, we were concerned about how burdensome the participants might find the ESM data collection. The second author has experience with incentivizing long-term student participation in studies, but never with the frequency of data collection involved in the ESM. Our participants reported that we struck the right balance of research incentives to the amount of data requested from each participant. On average, the five students completed 85% of notification requests in the first week of data collection. Specifically in Fig. 1, the two students who opted in for a reflective interview completed 95% and 80%, respectively. In the interviews, these students reported the incentive level was commensurate with the intensity of data collection.

"The way I thought of it was just kind of in terms of how much time I spent on the surveys and how much money I got from the incentives and it was plenty of money. I didn't do an exact calculation but it was, you know, I probably spent less than, you know, half an hour total each day and I guess I would do it again for the same incentive over the same amount of time. I didn't have any problems with it so I don't think I would change anything." (Res. 987)

B. Integrity of ESM Data

Our second research question focused on threats to the integrity of the data. In this pilot study, we were primarily concerned with students understanding of the specific questions we were asking via the ESM. We also wanted to understand if the frequency of data collection led to any sampling fatigue for the participants. In this section we address both of these threats to integrity of the data.

In our pilot data, there is evidence that students are appropriately interpreting the questions as measures of SE and answering based on the task that they are currently engaged in. As shown by the circles around the Notifications in Fig 1, when Res. 980 answered the Task-level SE questions about course-related activities, their responses indicate that they mostly respond with 3's and 4's. In the interview, we probed around why Res. 980 selected 4's as compared to 3's,

"My thought process, especially when I said I was doing... homework... I'm way better in computational numerical stuff than analytical physics I guess. So I guess my confidence interval was basically in how easily I could solve the question that was put in front of me." (Res. 980)

Res. 980 describes not only thinking specifically about their confidence to succeed in the task, but also describes thinking specifically about the task right in front of them at the moment. This response aligns ideally with what the ESM is intended to probe. Similarly Res. 987 describes responding to the task specific questions by, "[tying] it to my understanding and how competent I felt in whatever I was doing."

The other area of integrity we explored in this pilot study was sampling fatigue. In Fig. 1 we see that Res. 980 primarily engaged in course related activities (indicated with the circles). We see variation in Res. 980 responses from answering 1's to 5's on the Task-Level SE questions. However, for

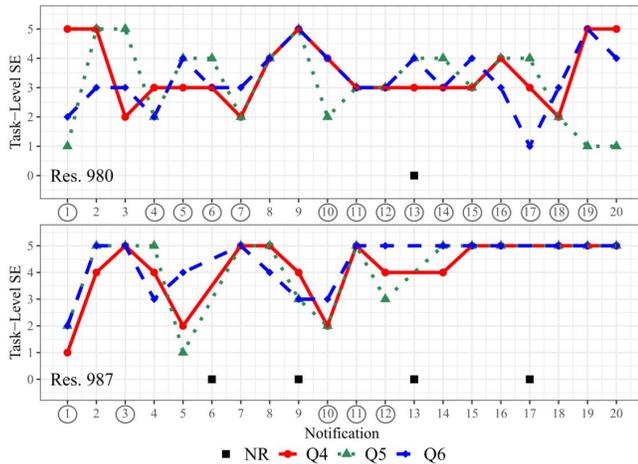


FIG. 1. Task-level SE over time for Res. 980 (top) and Res. 987 (bottom). Solid red lines indicate responses to Q4, the dotted green lines indicated responses to Q5, and the dashed blue lines indicate responses to Q6. The black squares indicate when a participant had no response (NR) to the survey notification. The circles on the Notification numbers indicate the in-the-moment task was course related.

Res. 987, most of the tasks are unrelated to course work and we find less variation toward the end of the week. Looking at the tasks that Res. 987 was doing, we see activities such as “Sleeping” and “Going downtown on a bus”; these tasks are ones that we would expect a high level of Task-Level SE and not much variation in the data.

Pairing these observations with the interviews, we find that while the respondents started to be able to predict the questions that would be asked, they indicate they still thought carefully about their responses before answering,

Interviewer: *So do you feel like you sort of starting skipping questions and going straight to the answer?*

Res. 980: *Yes. I know what the question is going to be like. Except a couple of times when I was like, “I’m studying for this quiz” or “I’m staring at this book.” You can see a sequence of mine – I’m studying for a quiz, I’m hopelessly studying for a quiz, I [give up].”*

This sequence that Res. 980 is reflecting upon is Notifications 17-20 in Fig. 1. For Notification 17 and 18, they respond “study for an EM2 quiz” and “staring at my EM2 notes” (variation between 1’s and 4’s). Then in Notification 19, they are “giving up studying for my EM2 quiz, what happens, happens” (mostly 4’s) and in Notification 20, they’ve “finished EM2 quiz” (4’s and 5’s). This alignment between reflective variation and the variation in the ESM data suggests that Res. 980 is not skipping the questions without thinking carefully, but is instead *anticipating* the questions, and considering the task at hand in the response.

IV. DISCUSSION & FUTURE DIRECTIONS

This pilot study sought to explore a novel mixed-methods approach to studying students’ SE. Through this design, we provided evidence that the incentives were sufficient for the students who participated in this study. In the future, our goal is for students to participate in the ESM data collection for three, two-week periods throughout the semester. Given the results presented above, it seems likely that students would participate in a larger study for the incentive of \$50 per week for completing at least 80% of the notifications. In our future research, we intend to use the justification from Res. 987 along with the average time to respond to transparently explain to the participants how our research incentive was generated. While the students who volunteered to participate in the study answered 85% of the notifications, in the future we plan to refine our recruitment strategies by offering participation to a broader audience and to further validate our methodology.

In addition, there was evidence that this methodology will allow us to probe into student thinking around self-efficacy. In a larger-scale study, the ESM data is intended to be used to identify critical points where a student’s SE varies significantly. These critical points would then be probed in an interview to dive deeper into the student’s experiences at that critical point. The alignment between the reflective interview data and the ESM data for Res. 980, strongly supports that this could be done. Specifically, Res. 980 remembered the experience they had around studying for a quiz and the Task-Level SE quantitative data was in line with how they described that experience. In the future, the interview protocol will then probe questions to further explore the student’s SE around that particular event.

In general, the results indicated that our study design is promising and yields rich results. We note that the evidence presented here was centered around the Task-Level SE scale. In the future, we plan to explore students’ SE with respect to the specific tasks that the students are engaging in. In addition, we will further validate the other levels of specificity, specifically, Course-Level SE and Career-Level SE; our pilot data indicate that there is less variability in those scales. Once validated, we intend to investigate the relationships of students’ SE across these levels of specificity. Overall, the pilot study provided evidence of the benefits and limitations to this novel approach to understanding the development of students’ SE.

ACKNOWLEDGMENTS

We would like to thank Laura Wood for conducting the interviews in the study and Alyssa Waterson for discussion around the plots. The funding for this work is provided in part by the MSU Foundation, administered by the MSU Office of Research & Innovation.

-
- [1] A. Bandura, *Self-efficacy: The exercise of control* (1997).
- [2] S. L. Britner, *J. Res. Sci. Teach.* **45**, 955 (2008).
- [3] L. M. Larson, K. M. Pesch, S. Surapaneni, V. S. Bonitz, T. Wu, and J. D. Werbel, *J. Car. Assess.* **23**, 399 (2015).
- [4] S. Lau and R. W. Roeser, *Educ. Assess.* **8**, 139 (2002).
- [5] R. R. Bryan, S. M. Glynn, and J. M. Kittleston, *Sci. Educ.* **95**, 1049 (2011).
- [6] J. Dalgety and R. K. Coll, *Assess. Eval. High. Educ.* **31**, 303 (2006).
- [7] S. M. Glynn, G. Taasobshirazi, and P. Brickman, *J. Res. Sci. Teach.* **46**, 127 (2009).
- [8] R. W. Lent, S. D. Brown, and K. C. Larkin, *J. Counsel. Psych.* **33**, 265 (1986).
- [9] R. W. Lent, S. D. Brown, and K. C. Larkin, *J. Counsel. Psych.* **34**, 293 (1987).
- [10] P. R. Aschbacher, E. Li, and E. J. Roth, *J. Res. Sci. Teach.* **47**, 564 (2010).
- [11] D. A. Luzzo, P. Hasper, K. A. Albert, M. A. Bibby, and E. A. Martinelli Jr, *J. Counsel. Psych.* **46**, 233 (1999).
- [12] R. M. Marra, K. A. Rodgers, D. Shen, and B. Bogue, *J. Eng. Educ.* **98**, 27 (2009).
- [13] P. R. Pintrich, D. A. Smith, T. Garcia, and W. J. McKeachie (1991).
- [14] L. E. Kost-Smith, *Characterizing, modeling, and addressing gender disparities in introductory college physics* (Dissertation, 2011).
- [15] M. D. Lindstrøm, C. and Sharma, *Int. J. Innov. Sci. Math. Educ.* **19** (2011).
- [16] H. Fencl and K. Scheel, *J. Coll. Sci. Teach.* **35**, 20 (2005).
- [17] V. Sawtelle, E. Brewe, and L. Kramer, in *Physics Education Research Conference 2010* (Portland, Oregon, 2010), vol. 1289 of *PER Conference*, pp. 289–292.
- [18] E. M. Marshman, Z. Y. Kalender, T. Nokes-Malach, C. Schunn, and C. Singh, *Phys. Rev. Phys. Educ. Res.* **14**, 020123 (2018).
- [19] D. S. Brewer, *The effects of online homework on achievement and self-efficacy of college algebra students* (Dissertation, 2009).
- [20] L. Ainscough, E. Foulis, K. Colthorpe, K. Zimbardi, M. Robertson-Dean, P. Chunduri, and L. Lluka, *CBE Life Sci. Educ.* **15**, ar19 (2016).
- [21] J. M. Nissen and J. T. Shemwell, *Phys. Rev. Phys. Educ. Res.* **12**, 020105 (2016).
- [22] V. Sawtelle, E. Brewe, R. M. Goertzen, and L. H. Kramer, *Phys. Rev. Phys. Educ. Res.* **8**, 020111 (2012).
- [23] A. L. Zeldin and F. Pajares, *Am. Educ. Res. J.* **37**, 215 (2000).
- [24] A. L. Zeldin, S. L. Britner, and F. Pajares, *J. Res. Sci. Teach.* **45**, 1036 (2008).
- [25] M. Hutchison-Green, D. K. Follman, and G. M. Bodner, *J. Eng. Educ.* **97**, 177 (2008).
- [26] J. W. Creswell and J. D. Creswell, *Research design: Qualitative, quantitative, and mixed methods approaches* (Sage publications, 2017).
- [27] J. M. Hektner, J. A. Schmidt, and M. Csikszentmihalyi, *Experience sampling method: Measuring the quality of everyday life* (Sage, 2007).
- [28] *Lifedata*, <https://www.lifedatacorp.com/>. Accessed 1/2020.
- [29] S. Shiffman, *Psych. Assess.* **21**, 486 (2009).
- [30] A. Bishay, *J. Under. Sci.* **3**, 147 (1996).
- [31] M. A. Moreno, L. Jelenchick, R. Koff, J. Eikoff, C. Diermyer, and D. A. Christakis, *Comp. Hum. Behav.* **28**, 1097 (2012).
- [32] A. Stone, S. Shiffman, A. Atienza, and L. Nebeling, *The science of real-time data capture: Self-reports in health research* (Oxford University Press, 2007).
- [33] J. A. Schmidt and L. Shumow, *Psych. Self-Eff.* Hauppauge, NY: Nova Science Publishers (2012).
- [34] J. M. Nissen, *Phys. Rev. Phys. Educ. Res.* **15**, 013102 (2019).
- [35] R. Henderson, J. Stewart, L. Michaluk, J. Deshler, E. Fuller, and K. Rambo-Hernandez, *Accepted to J. Sci. Ed. Technol.* (2020).
- [36] P. R. Pintrich, D. A. Smith, T. Garcia, and W. J. McKeachie, *Educ. Psychol. Meas.* **53**, 801 (1993).