Measuring Affective Experience in the Midst of STEM Learning

Jayson Nissen, Jonathan Shemwell, University of Maine, 120 Bennett Hall, Orono, ME 04469,
Email: jayson.nissen@maine.edu, jonathan.shemwell@maine.edu

Abstract: The Experience Sampling Method, an at-the-moment survey technique, was used to measure university students' affective experiences within school and in their daily lives on four variables: activation, self-efficacy, motivation, and stress. Affect was compared for school vs. non-school, and within school, STEM coursework vs. non-STEM coursework. Within STEM, affect for a focal physics course was compared to affect for all other STEM courses. School was experienced with higher stress, lower intrinsic motivation, and lower self-efficacy than non-school. STEM and non-STEM courses were not experienced differently, but the physics course was experienced with higher stress and lower self-efficacy than other STEM courses. The results suggest that, broadly, university coursework may undermine intrinsic motivation and that the negative impact occurs in the midst of instruction. More tentatively, the process of engaging with challenging STEM content, such as that of the physics course, may tend to increase stress and undermine self-efficacy.

Introduction

Increasing the number and diversity of students who enter and remain in the STEM education pipeline is an important goal in the United States and in many developed countries (National Commission on Mathematics and Science Teaching for the 21st Century, 2000; Osborne & Dillon, 2008). Attainment of this goal will naturally depend on positive learning outcomes for diverse students in K-16 STEM education. One important and often overlooked class of outcome is positive affect toward STEM domains. Two examples of affect are feelings of self-efficacy and motivation. Declining STEM enrollments and increasingly negative attitudes toward STEM suggest that positive affect is in short supply within STEM instruction (Osborne, Simon, & Collins, 2003; Semsar, Knight, Birol & Smith 2011). Understanding the nature and sources of affect within the current STEM education system is therefore an important research goal.

Most studies of affect in STEM education have utilized survey techniques in which students report their general attitudes and beliefs for an overall experience, such as at the end of a course. Many studies have related affective variables measured by surveys to important outcomes such as achievement. For example survey-based studies in STEM have measured the possible mediating effect on student achievement of self-efficacy (Lee, 2009; Marra and Bogue, 2009), interest (Koller, Baumert & Schnabel, 2001), and motivation (Singh, Granville & Dika, 2002; Mujtaba and Reiss, 2013). Mujtaba and Reiss (2013) analyzed student end of course responses to a survey to show that the lower representation of women in physics was related to differences between male and female secondary student's affective experiences in physics. Lee (2009) used survey responses to investigate the factorial relationship of three affective traits within math, self-concept, self-efficacy, and anxiety, and how their relationship to student achievement varied between countries. Luzzo, Hasper, Albert, Bibby and Martinelli (1999) used pre/post surveys to investigate the impact of a self-efficacy-enhancing intervention on student's math and science self-efficacy and career interests. In a mixed methods approach Girod, Rau and Shepige (2002) used both surveys and interview case studies to demonstrate that elementary school students had a higher quality of experience in a course focused on aesthetic understanding as opposed to one focused on conceptual understanding.

A challenge in studying the impact of instruction on affect is that it is difficult to measure affective response in the midst of the experience. Holleran, Whitehead, Schmader and Mehl (2010) used analysis of random audio samples from day to day life to overcome this difficulty. They showed how stereotype threat affected how female STEM faculty interacted with other female faculty as opposed to male faculty. Stephens (2012) reproduced the classroom experience in the laboratory in order to measure stress caused by the mismatch between the culture of college and that of first generation students on students. Stephens measured stress by periodically collecting saliva (for cortisol analysis) while students gave speeches they had written.

Our approach to measuring affective response within instruction is to adapt a technique called the Experience Sampling Method (ESM) (Hektner, Schmidt & Csikszentmihalyi 2007). We used ESM to measure university students' affective experience in their courses and throughout their day-to-day lives on four distinct categories of affective experience: activation, self-efficacy, motivation, and stress. This method allowed comparing experience between the students' school and non-school activities, and between their STEM and non-STEM courses.

Methods

The purpose of the study was to investigate students' experiences of self-efficacy, stress, motivation, and activation in both non-school and school experiences, and within these school experiences to compare students experience in both STEM and non-STEM courses. A further purpose was to compare affect within a focal, reformed STEM course to affect within other STEM courses.

Context

The study took place at a flagship state university in United States. The focal STEM course was a 15-week large-enrollment calculus-based introductory physics course. The instructor, who had more than 20 years of teaching experience, was on his fifth year of teaching the course. In the previous four years, the instructor had implemented several research based teaching practices that are commonly referred to as interactive engagement (IE) practices (Hake, 1998; Kost-Smith, 2011). These included collaborative problem solving activities in the two one-hour recitation periods each week, and electronic response, i.e. “clicker”, questions embedded throughout both one-hour weekly lectures. There was also one two-hour lab each week that was not taught by interactive engagement. Data collected in this “IE Physics” course demonstrated learning outcomes that were similar or superior to those of IE courses at other institutions (Hake, 1998). Specifically, on Force and Motion Conceptual Evaluation (Thornton & Sokoloff, 1998), the IE Physics course yielded normalized gains for the previous four years ranging between 45% and 55%. Grade distributions in the course were approximately evenly distributed with 24% A’s, 23% B’s, 27% C’s and a 26% rate of drop, withdrawal, or failure (DWF). A report obtained from the university’s Director of Institutional Research showed that the grading distribution was not different from those of other STEM courses.

Participants

Participants were 33 of the 244 students enrolled in the IE physics course. Half of the participants were female (17). This proportion over-represented the population of females in the course, which was 18%. Participants were recruited by a brief presentation in the IE physics lecture and a follow-up email to those who provided their contact information. Participants had higher final grades than the average for students in the course, 82% versus 76%, but similar distributions in grades awarded. Participants were compensated with extra credit in the IE physics course for the first time they participated, and a check for 50 USD for the second time they participated.

Design

The design was a within-subject comparison of the affective experience within school (i.e., while in class or doing homework) and non-school activities spanning the range of students’ day-to-day lives. Experience within school was broken down according to three different types of courses participants were enrolled in: IE Physics, STEM courses other than IE Physics, and non-STEM courses.

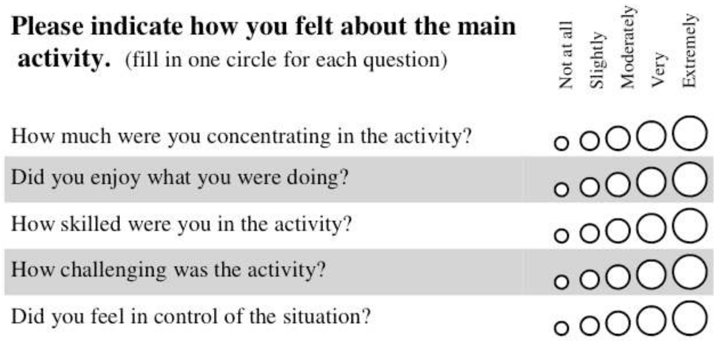
Procedures

Data collection occurred during two separate weeklong periods during the third and tenth week of the semester. Twelve of the 33 participants provided ESM data during both data collections. During the data collection weeks, no tests were taken or returned in the IE physics course. Participants completed a one-hour training the week prior to their first week of data collection. Training included a description of the ESM and practice doing the ESM.

During data collection participants in the ESM were semi-randomly signaled using a text message sent to their personal cell phone 5-8 times a day across each one-week period for a total of 50 signals for each week. Five of the signals were scheduled for random times during each of the scheduled IE physics course components (lab, lectures, and recitations) throughout the week. This allowed collecting a large enough number of surveys to provide a representative and diverse sample of experience within the IE physics course setting. Participants received the remaining 45 random text message signal during daily activities: making breakfast, driving, sitting in class, playing sports, etc.

Upon receiving a signal, students would pause what they were doing to fill out very brief at-the-moment survey about their affective experience. Sometimes the signal would occur during a university course or an activity related to a course such as homework; sometimes it would occur during non-school recreational hours or while students were involved in quotidian tasks such as laundry or dishes. Participants’ responses began with writing brief statements about what they were doing, what they were thinking, and where they were. Participants would then check a circle for each of 20 Likert-scale affective questions to indicate their affective experience in the activity. Example items drawn from the survey are provided in Figure 1 and described more fully in Table 1. A pilot study showed that completing the survey took 1-3 minutes. Most participants in the pilot study indicated it had little or no impact on their activity.

Some activities (such as driving), prevented participants from providing their responses at the moment of the signal. In accordance with standard ESM procedures, participants were told to complete surveys as soon as possible after the signal and to complete them regardless of how much time had passed. The survey form included a space to indicate the delay between the signal and completion of the survey. Only surveys completed within 15 minutes of the signal were included in the data analysis. This is a standard technique of ESM to help ensure that the responses measure affect as close to the moment of the signal as practicable. Similarly, only participants who completed at least fifteen surveys were included in the data analysis because that is broadly considered to be a minimum threshold for measuring student's average experience when using the ESM (Hektner *et. al.,* 2007). The first author collected the surveys at the end of the data collection period and transcribed the Likert-scale and open-ended responses into a data file containing all responses for each student.


Please indicate how you felt about the main activity. (fill in one circle for each question)

	Not at all	Slightly	Moderately	Very	Extremely
How much were you concentrating in the activity?	o	o	o	o	o
Did you enjoy what you were doing?	o	o	o	o	o
How skilled were you in the activity?	o	o	o	o	o
How challenging was the activity?	o	o	o	o	o
Did you feel in control of the situation?	o	o	o	o	o

Figure 1: Examples of Likert-scale affective questions used in the ESM data collection.

Construct Definitions and Instrumentation

Table 1 defines each category of effective experience and shows several survey items used to measure it. The survey items were drawn from prior research using ESM (Hektner *et. al.,* 2007). As is done in many ESM studies, item groupings were confirmed in the present study using exploratory factor analysis. As can be seen from Table 1, the measures within the ESM access each of the affective categories broadly and cannot distinguish fine details within these constructs. For instance research on motivation includes concepts such as goal orientation (Belenky and Nokes-Malach, 2012), which are not represented in the present study's formulation. Similar examples of the complexity of the other affect categories defined in Table 1 exist throughout the scientific literature. The present study's coarser measurements were a trade-off for the ability to access multiple categories of affective experience across daily activities for a representative sample of participants.

Self-efficacy was the primary category of interest in designing the research due its obvious implications for attracting and retaining students in STEM. Bandura (1997) defined self-efficacy as the confidence in one's capability to perform the actions necessary to achieve a particular goal. The present study departed from Bandura in that it focused on self-efficacy experiences, as opposed to more stable dispositions or beliefs, by asking how skilled or successful people felt in the activity they were doing (Table 1). However, the present study aligned with Bandura in that the measure was specific to the activity at hand. The self-efficacy measure was based on items with unipolar scales, starting from a zero value and extending to a maximum.

Intrinsic motivation was comprised of enjoyment, excitement, freedom (as opposed to constraint), and importance to future goals. Several of the motivation items had bipolar scales, making the intrinsic motivation measurement either negative (extrinsic) or positive (intrinsic). Intrinsic motivation was measured on the assumption that it is necessary, at some level, for students to opt themselves into a learning experience. Additionally, Deci, Koestner & Ryan (1999) showed the importance of intrinsic motivation by demonstrating the long-term negative impact of extrinsic motivation on persistence and learning.

Stress and activation were utilized as complementary measures to further inform self-efficacy and motivation. For instance, we wanted to see if feelings of low self-efficacy were generally accompanied by stress. Lazarus & Folkman (1984) defined stress as a negative feeling resulting from an individual's perception that they do not have the resources to cope with a perceived situation. Our measure was only loosely aligned with this definition: we asked directly about stress, worry and frustration. All of the stress items utilized unipolar scales. Activation was defined as the level of involvement in the task, consistent with Thayer (1996). In contrast to Thayer's formulation activation was a unipolar measure, which included alertness, attentiveness, and the degree to which the participant was concentrating on the activity.

Table 1: Details for each of the affective categories measured by the ESM

Affect Category	Number of items	Example questions[1]	Definition
Activation	7	How alert did you feel? How attentive did you feel? How much were you concentrating?	The level of involvement in the task at hand.
Self-Efficacy	5	How skilled were you? Did you feel in control? Were you succeeding?	Sense of personal success and capability in accomplishing the task at hand.
Stress	3	How stressed did you feel? How worried did you feel? How frustrated did you feel?	Emotional strain or tension resulting from adverse or very demanding circumstances.
Intrinsic Motivation	4	Did you feel free or constrained? Did you enjoy what you were doing? How important was the activity to your future goals?	The reason for action is drawn from a sense of enjoyment rather than an external reward.

1 Actual questions used in the ESM followed a briefer format.

Methods of Analysis

Completed surveys were coded as either non-school activities or by the type of course associated with the school activity. School activities were then reduced to three approximately equal sized categories: (1) Non-STEM courses which included a diverse range of courses such as English, anthropology, and art; (2) STEM courses excluding IE physics, primarily consisted of chemistry, calculus and introductory engineering courses; and (3) IE physics. Analysis of variance showed that, within both the STEM and non-STEM categories, there were no statistically significant differences in affective responses between courses composing both categories, i.e. calculus and chemistry ($p > 0.20$).

For an activity type, each participant's set of responses to each of the Likert-scale affect questions was converted to Z-scores based on the mean and standard deviation of the participants' responses to that question. A Z-score is calculated for any score by subtracting from it the mean for the set of scores from which it is drawn, then dividing the result by the standard deviation of the set. Calculating Z-scores is a standard ESM procedure that allows comparing responses between students who use the Likert-scale differently. Essentially each person's score for a question is scaled in relation to their own average response for that question. Exploratory factor analysis on the responses confirmed that individual items grouped into four factors corresponding to the four constructs the survey was designed to measure: activation, self-efficacy, stress, and motivation. These categories of affect were calculated by averaging the Z-scores for each item within the construct.

Results

Differences in mean Z-scores were tested with two multivariate analysis of variance tests (MANOVA). The first test used two independent variables (non-school and school), and four dependent variables (activation, self-efficacy, motivation, and stress). It showed a statistically significant effect of activity type $F(4,1440) = 334.0$, $p < 0.001$. Subsequent one-way ANOVA tests contrasted activity type for each affect variable. They showed that there were statistically significant differences in mean Z-scores for school and non-school activities for all four variables, activation $F(1,1440) = 13.1$, $p < 0.001$, self-efficacy $F(1,1440) = 387.8$, $p < 0.001$, stress $F(1,1440) = 130.2$, $p < 0.001$ and intrinsic motivation $F(1,1440) = 1019.4$, $p < 0.001$. The second MANOVA used the same dependent variables but separated the school independent variable into non-STEM courses, STEM courses and IE physics. Non-school activities comprised the fourth independent variable. The MANOVA showed a statistically significant effect of activity type $F(12,1440) = 75.5$, $p < 0.001$. Subsequent one-way ANOVAs comparing the four activity types for each affect variable showed that all four affect variables had at least one statistically significant difference in mean Z-scores, activation $F(3,1440) = 9.15$, $p < 0.001$, self-efficacy $F(3,1440) = 140.5$, $p < 0.001$, stress $F(3,1440) = 57.8$, $p < 0.001$ and intrinsic motivation $F(3,1440) = 341.2$, $p < 0.001$. Post-hoc analysis was conducted using Tukey's HSD.

Results of the ESM measurements are summarized in Figure 2. Table 2 shows statistically significant differences and effect sizes in Z-score units. Both representations show that school and non-school activities were experienced very differently. School, compared to non-school, produced slightly more activation, much lower self-efficacy, much higher stress, and much lower intrinsic motivation (i.e., more extrinsic motivation). Within school, STEM courses were experienced with slightly more activation and slightly more self-efficacy than non-STEM, but these differences were not statistically significant. IE physics was experienced, as compared to other STEM courses, with similar activation, but lower self-efficacy and higher stress. Intrinsic motivation was not different across school categories.

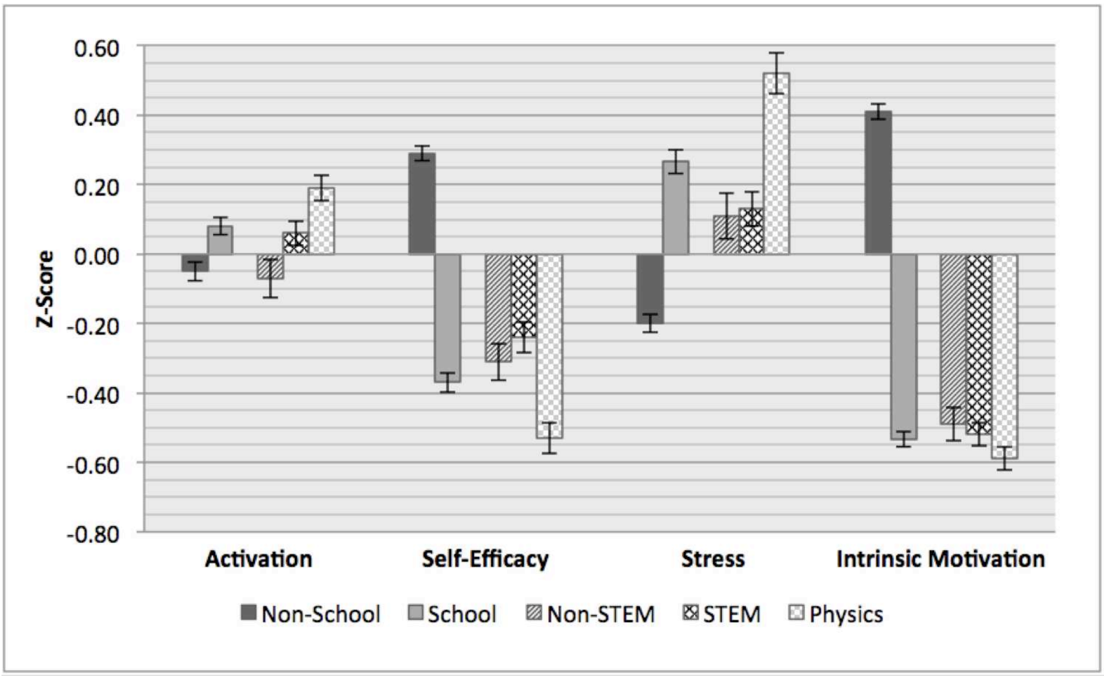


Figure 2. ESM data for all participants across the four activity types and four affective experiences. The vertical axis is in Z-score comparing the average experience in each activity to the overall average experience for each of the four affective experiences.

Table 2: Average experience, in Z-scores, across the four activity types with calculated differences between School V Non-School, Stem V Non-STEM, and STEM V IE physics. ***p<0.001.

	Non-School	School	School minus Non-School	Non-STEM	STEM	STEM minus Non-STEM	IE physics	IE physics minus STEM
N (surveys)	816	624		161	233		230	
Activation	-0.05	0.08	0.13***	-0.07	0.06	0.13	0.19	0.06
Self-Efficacy	0.29	-0.37	-0.66***	-0.31	-0.24	0.7	-0.53	-0.29***
Stress	-0.20	0.27	0.47***	0.11	0.13	0.02	0.52	0.39***
Intrinsic Motivation	0.41	-0.54	-0.95***	-0.49	-0.52	-0.03	-0.59	-0.1

Although the Z-score comparison is very useful and represents a common analytical approach to ESM, the fact that all comparisons are relative can be misleading. For example, the large negative Z-score for intrinsic motivation in school does not necessarily mean that intrinsic motivation was actually negative in school. Rather, students may have had a slightly positive intrinsic motivation in school that was nevertheless lower than a larger positive value for non-school experiences. For this reason, we examined the raw values for each of the affect categories to better interpret students' affective states within activity types. The following analysis is presented for motivation. A similar analysis was done for each of the other variables to ensure that the Z-score values in Figure 2 and Table 2 did not misrepresent students' affective states.

Intrinsic motivation was constructed from four questions on the ESM. Two of the questions, "Did you enjoy what you were doing?" and "How important was this activity to your future goals?" were measured on five-point unipolar scales. The other two questions, "How free or constrained did you feel?" and "How excited or bored did you feel?", were measured on seven-point bi-polar scales. The distributions of student responses on these four questions in school and non-school activities are shown in Figure 3. They show that intrinsic motivation in school was negative. School experiences tended to not be very enjoyable, and they were often boring or unexciting. Feelings of freedom or constraint were more mixed but trended toward constraint. Finally, school activities were important to students' future goals. All of these results were in opposition to experiences outside of school, which tended to be enjoyable, free, and exciting while being less important to students' future goals. Based on these results, we concluded that the negative Z-score for students' experiences of intrinsic motivation in school activities was representative of motivation that was extrinsic, compared to intrinsic non-school motivation.

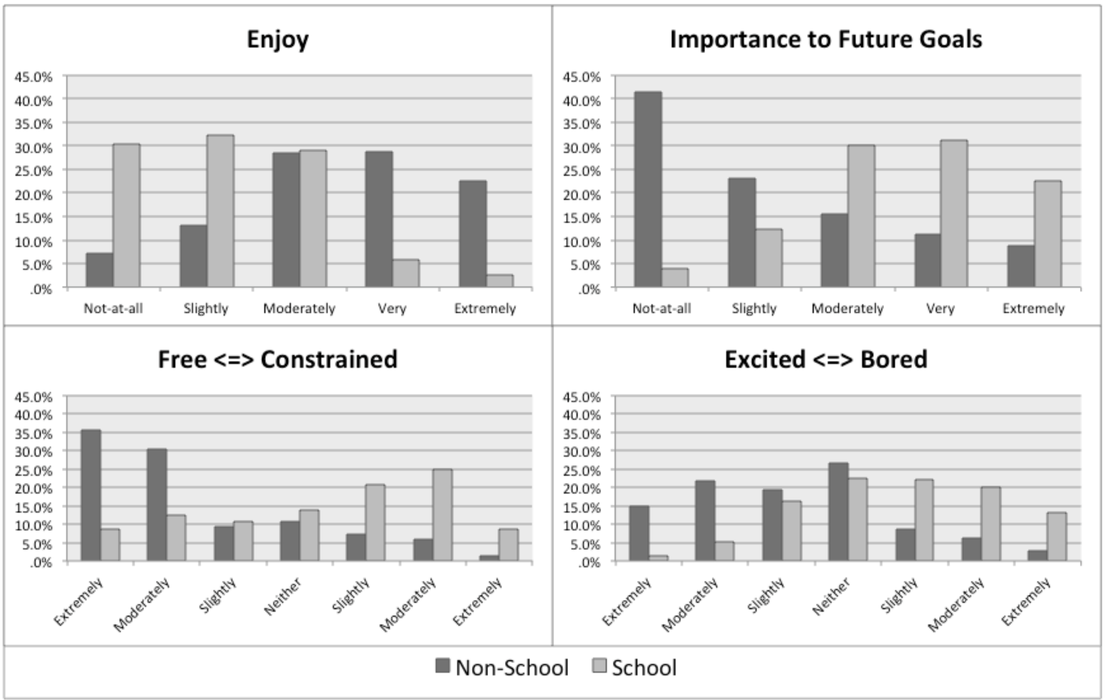


Figure 3: Distributions of student experience across the questions that made up intrinsic motivation.

Discussion

Comparing results for school and non-school, the differences in affect were consistent with the pressures that university students would be expected to feel as they strive to learn and generally perform well as novices in a competitive, (i.e., graded), environment. From this perspective, the lower self-efficacy and higher stress of school activities compared to non-school at least make sense. However, it is difficult to interpret the values of self-efficacy and stress for school activities. Perhaps these values could be improved; or perhaps they are a natural and healthy part of a student's struggle to learn. By contrast, the negative experience of intrinsic motivation in school activities (i.e., extrinsic motivation) is more clearly a cause for concern. Of course, extrinsic motivation is commensurate with the compulsory nature of many coursework activities such as studying for tests and completing problem sets. Educational systems generally tend to rely on structures and procedures, such as performance-contingent rewards, known to negatively impact intrinsic motivation (e.g., Deci, Koestner & Ryan, 1997). While the motivation results are not surprising, they are nevertheless worth paying attention to, since any educational program ultimately depends on students' self-regulated attempts to learn from what they are doing (Brophy, 2010). What the results of the present study show, compared to more theory-driven studies in the mold of Deci Deci, Koestner & Ryan (1997), is that negative intrinsic motivation is a measureable response to the education system, and not only a probable response based on the system's design. This "direct" measure of motivation has the potential to more squarely confront administrators, instructors, and other stakeholders with the consequences of motivation-reducing structures and procedures.

The IE physics course had higher stress and lower self-efficacy compared to other STEM courses, despite similar levels of activation (i.e., attentiveness, alertness, and concentration). These results could have been due to idiosyncrasies of the instructor. However the instructor's relatively long experience teaching the course, high average achievement, and distribution of grades similar to those of other STEM courses all undermine this possibility. We think it more probable that students' affect resulted from the experience of a fairly rigorous STEM course focused on learning conceptually challenging STEM content. If so, the present study would confirm the results of studies utilizing pre-post survey measures showing negative impacts of IE physics on self-efficacy (Kost-Smith, 2011). In the case of the present study, however, the negative impact of IE Physics was shown to occur within instructional activities, instead of upon reflection after the fact. This finding rules out the possibility that the negative impacts of self-efficacy measured by Kost-Smith may have occurred primarily as a result of experiences outside instruction, such as when students received their grades. Rather, a negative effect on self-efficacy was located unequivocally within the process of instruction.

The lower self-efficacy for IE Physics might seem be inconsistent with the relatively high achievement in the course compared to other IE courses, since high achievement would be expected to encompass mastery experiences promoting self-efficacy (Bandura, 1997). However, while students in an IE course may learn more than they would in a traditional, lecture-based course (Hake, 1998), they are probably still far from mastery, which takes much longer to achieve than the brief experience of a semester-long course would allow (Chi, Feltovitch, & Glaser, 1981; diSessa, 2006; Ohlsson, 2009). Thus, it seems reasonable to assume that most students in a course would not be able to achieve anything near the levels of mastery that would lead to increasing in self-efficacy, even if they were learning at optimal levels. For this reason, it may be that educators should expect effective learning of challenging STEM content to be decoupled from self-efficacy on the time frame of one or two semesters during which most students experience this content. In essence, while students knowledge and skills are growing in the course, their awareness of what they don't yet understand and can't yet do are growing faster.

The present study has taken steps toward better understanding what is arguably a crucial and too long neglected aspect of STEM teaching and learning, affective experiences within instruction. The most important contribution of the study has been to measure the affective response in the midst of the instructional process. The results of this measurement, should they be replicated and extended in future studies, have the potential to raise the awareness of affective response to the process of instruction, and to motivate and inform the search for better instructional methods.

References

Adams, W., Perkins, K., Podolefsky, N., Dubson, M., Finkelstein, N., & Wieman, C. (2006). New instrument for measuring student beliefs about physics and learning physics: The Colorado Learning Attitudes about Science Survey. *Physical Review Special Topics- PER*, 2(1).

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

Belenky, D. & Nokes-Malach, T. (2012). Motivation and Transfer: The Role of Mastery Approach Goals in Preparation for Future Learning. *Journal of the Learning Sciences*, 21(3): 399–432,

Brophy, J. E. (2010). *Motivating students to learn*. New York: Routledge.

Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and Representation of Physics Problems by Experts and Novices. *Cognitive Science*, *5*(2), 121–152.

Deci, E., Koestner, R., & Ryan, R. (1999). A Meta-Analytic Review of Experiments Examining the Effects of Extrinsic Rewards on Intrinsic Motivation. Psychological Bulletin, 125(6): 627–668.

diSessa, A. A. (2006). A history of conceptual change research. In K. Sawyer (Ed.), *The Cambridge Handbook of the Learning Sciences* (pp. 265-281). New York: Cambridge University Press.

Girod, M., Rau, C., and Schepige, A. (2003). Appreciating the beauty of science ideas: Teaching for aesthetic understanding. *Science Education*, 87(4): 574–587.

Hake, R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66 (1): 64–74.

Hektner, J., Schmidt, J., & Csikszentmihalyi, M. (2007). *Experience Sampling Method: Measuring the Quality of Everyday Life.* Thousand Oaks, CA: SAGE Publications.

Holleran, S., Whitehead, J., Schmader, T., & Mehl, M. (2010). Talking Shop and Shooting the Breeze: A Study of Workplace Conversation and Job Disengagement Among STEM Faculty. *Social Psychological and Personality Science*, 2(1): 65–71.

Koller, O., Baumert, J., & Schnabel, K. (2001). Does Interest Matter? The Relationship Between Academic Interest and Achievement in Mathematics. *Journal of Research in Mathematics Education*, 32(5): 448–470.

Kost-Smith, L. Characterizing, Modeling, and Addressing Gender Disparities in Introductory College Physics. PhD Thesis, 2011.

Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learning and Individual Differences*, 19(3): 355–365.

Lazarus, R. & Folkman, S. (1984). *Stress, appraisal and coping.* New York: Springer.

Luzzo, D., Hasper, P., Albert, K., Bibby, M. & Martinelli, E. (1999). Effects of self-efficacy-enhancing interventions on the math/science self-efficacy and career interests, goals, and actions of career undecided college students. *Journal of Counseling Psychology*, 46(2): 233–243.

National Commission on Mathematics and Science Teaching for the 21st Century. (2000). Before It's Too Late. Washington: U.S. Department of Education.

Marra, R. & Bogue, B. (2009).Women Engineering Students and Self-Efficacy : A Multi-Year , Multi-Institution Study of Women Engineering Student Self-Efficacy. *Journal of Engineering Education,*

January.
Mujtaba, T. & Reiss, M. (2013). Inequality in Experiences of Physics Education: Secondary School Girls' and Boys' Perceptions of their Physics Education and Intentions to Continue with Physics After the Age of 16. *International Journal of Science Education*, 35(11): 1824–1845.
Ohlsson, S. (2009). Resubsumption: A possible mechanism for conceptual change and belief revision. *Educational Psychologist, 44*, 20–40.
Osborne, J., & Dillon, J. (2008). Science Education in Europe : Critical Reflections A Report to the Nuffield Foundation.
Sawtelle, V., Brewe, E., Goertzen, R. M., & Kramer, L. H. (2012). Identifying events that impact self-efficacy in physics learning. *Physical Review Special Topics - Physics Education Research*, 8(2).
Sawtelle, V., Brewe, E., Kramer, L., Singh, C., Sabella, M., & Rebello, S. (2010). Positive Impacts of Modeling Instruction on Self-Efficacy. PERC Proceedings, 289–292.
Semsar, K., Knight, J., Birol, G., & Smith, M. (2011) The Colorado Learning Attitudes about Science Survey (CLASS) for use in Biology. *CBE Life Sciences Education*, 10(3): 268–78.
Singh, K., Granville, M., & Dika, S. (2002). Mathematics and Science Achievement: Effects of Motivation, Interest, and Academic Engagement. *Journal of Educational Research*, 95(6): 323–332.
Stephens, N., Townsend, S., Rose, H., & Phillips, L. (2012). A cultural mismatch: Independent cultural norms produce greater increases in cortisol and more negative emotions among first-generation college students. *Journal of Experimental Social Psychology*, 48: 1389–1393.
Seymour, E. & Hewitt, N. (1997). *Talking About Leaving: Why Undergraduates Leave the Sciences,* Boulder, CO: Westview Press.
Thayer, R. E. (1996). *The origin of everyday moods: Managing energy, tension, and stress*. New York: Oxford University Press.
Thornton, R. & Sokoloff, D. (1998). Assessing student learning of Newton 's laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula. *American Journal of Physics*, 66(4): 338–352.

