Promoting Intentions to Persist in Computing: An Examination of Six Years of the EarSketch Program

Dana Linnell Wanzer, Tom McKlin, Jason Freeman, Brian Magerko, Taneisha Lee

This is a postprint (accepted manuscript) of an article published by Taylor & Francis in Computer Science Education on January 2020, available online:

http://www.tandfonline.com/10.1080/08993408.2020.1714313

Abstract

Background and Context: EarSketch was developed as a program to foster persistence in computer science with diverse student populations.

Objective: To test the effectiveness of EarSketch in promoting intentions to persist, particularly among female students and under-represented minority students.

Method: Meta-analyses, structural equation modeling, multi-level modeling, and qualitative analyses were performed to examine how participation in EarSketch and other factors affect students' intentions to persist in computing.

Findings: Students significantly increased their intentions to persist in computing, g=.40[.25,54], but examination within just the five quasi-experimental studies did not result in a significant difference for students in EarSketch compared to students not in EarSketch, g=.08[-.07, .23]. Student attitudes towards computing and the perceived authenticity of the EarSketch environment significantly predicted intentions to persist in computing.

Implications: Participation in computer science education can increase students' intentions to persist in programming, and EarSketch is one such program that can aid in these intentions.

Keywords: Persistence, Computer Science, Computer Education Program

Promoting Intentions to Persist in Computing: An Examination of Six Years of the EarSketch Program

A significant goal in the STEM community, and with computer science education practitioners and researchers specifically, is to increase persistence in many STEM majors and careers, particularly among minority and female students who are historically underrepresented in STEM fields. Despite increased participation in computing courses at the high school level, students do not often persist in computer science through high school to college and beyond (Shaw & Barbuti, 2010). This suggests that participation and enrollment in high school computer science courses is not sufficient for creating a steady pipeline of students into computer science majors and careers.

The Theory of Planned Behavior (TPB) states that one of the best predictors of behavior is one's intentions towards performing that behavior (Ajzen, 2001; Montaño & Kasprzyk, 2008). There has been ample support for this model across a wide array of behaviors, including health and education (e.g., Armitage & Conner, 2000; Sutton, 1998). For instance, intent to persist has been shown to be positively related to college persistence (Bean, 1982; Cabrera, Castaneda, Nora, & Hengstler, 1992; Porter & Swing, 2006). Students who intend to declare a STEM major in college are also more likely to actually declare a STEM major once they are in college (Bottia, Stearns, Mickelson, Moller, & Parker, 2015; Wang, 2012). However, it may be beneficial to have these intentions to persist earlier in life, with research finding that strong intentions in middle or early high school predict intentions and actual persistence in college (Maltese & Tai, 2011; Sadler, Sonnert, Hazari, & Tai, 2012; Tai, Liu, Maltese, & Fan, 2006).

TPB also describes three predictors of intentions towards performing a behavior. First, attitudes towards the behavior are important precursors to intentions to persist; if students do not have positive attitudes towards computer science they are unlikely to persist in computer science.

Attitudes towards STEM is multi-faceted, comprised of constructs such as perceived confidence, competence, and success in STEM, as well as perceived importance of STEM (Unfried, Faber, Stanhope, & Wiebe, 2015). Ample research has examined how students' attitudes towards STEM subjects are related to their intent to persist in STEM education and careers (Beal & Crocket, 2010; Blinkenstaff, 2005; Eccles, Vida, & Barber, 2004; Wiebe, Unfried, & Faber, 2018). For instance, when students like STEM and the STEM experiences they have in high school they are more likely to declare a STEM major (Bottia, Stearns, Mickelson, Moller, & Valentino, 2015).

Second, people must also believe that those they care about (e.g., peers, family, teachers) approve of the behavior; students must believe that others think positively towards computer science as well. Research has supported the importance of parents (Bandura, Barbaranelli, Caprara, & Postorelli, 2001; Lee & Shute, 2010; Zeldin & Pajares, 2000), teachers (Barker, McDowell, & Kalahar, 2009; McInerney, 2008), and peers (Olitsky, Loman, Gardner, & Billiups, 2010; Papanastasiou & Zembylas, 2004; Robnett & Leaper, 2012), as well as combinations of multiple communities at home or at school (DuBow, Weidler-Lewis, & Kaminsky, 2019) on student persistence in STEM majors and careers.

Relatedly, stereotypes about who is and is not suited for a STEM career can impact students' perceived norms about STEM. For example, female and under-represented minority (URM) students may not believe they are the typical STEM student and therefore not pursue a STEM major (Cheryan, Master, & Meltzoff, 2015; Hyde & Mertz, 2009; Litzler et al., 2014; National Academies of Sciences, Engineering, and Medicine, 2016). Ample research has examined how gender and race/ethnicity, as well as other student background characteristics, are related to student persistence in computing (e.g., Barker et al., 2009; Blinkenstaff, 2005; Jagacinski, Lebold, & Salvendy, 1998).

Third, one's perceived behavioral control can affect their intentions to persist; when students believe they have volitional control over their performance or believe that they have the self-efficacy to do a particular behavior, they are more likely to want to persist in doing that behavior. Research in STEM fields supports the linkage between self-efficacy and intentions to persist in STEM education and careers (Chemers, Zurbriggen, Syed, Goza, & Bearman, 2011; Wang, 2013; Scott & Mallinckrodt, 2005; Simpkins, Davis-Kean, & Eccles, 2006). Together, these three factors—attitudes towards the behavior, subjective norms about the behavior, and perceived control over the behavior—have been found to be significant predictors of intentions and actual behavior in a wide range of contexts (Ajzen, 2001; Montaño & Kasprzyk, 2008).

Although there is rich literature on the TPB—and on aspects of TPB applied towards computer science—the replicability crisis permeating many fields suggests researchers need to avoid one-off studies which may provide inflated effects. Rather, researchers need to conduct multiple studies intended to replicate the findings achieved in studies by conducting similar studies with another group from the same population, from different populations, and using different measurements and analyses. The replication crisis has received ample attention from the field of psychology (Camerer et al., 2018; Ioannidis, 2005; Open Science Collaboration, 2015; Pashler & Wagenmakers, 2012), and computer science education has recently become concerned with the issue too (Ahadi, Hellas, Ihantola, Korhonen, & Petersen, 2016; Clear, 2006; Margulieux, Ketenci, & Decker, 2019). There have been many suggested practices for improving the quality and replicability of studies and findings (see John, Loewenstein, & Prelec, 2012; Nosek & Läkens, 2014; Simons, Nelson, & Simonsohn, 2011 for some suggestions), but one suggestion is to conduct more meta-analyses (Card, 2017).

A meta-analysis is a statistical procedure to combine the data—specifically the effect sizes—from multiple studies to determine a common effect size and reasons for variations in the common effect size across studies (Borenstein, Hedges, Higgins, & Rothstein, 2009). Individual studies may not have the power to detect a precise effect size and often vary by factors not considered related to the outcome of interest (e.g., sample characteristics). By combining individual studies through a meta-analysis, effect sizes are more precise and generalizable to a larger population, and inconsistencies across studies can be examined quantitatively. Meta-analyses are essentially an analysis of the replication of findings, lending itself well towards improving research and improving the replicability of scientific findings (Card, 2017).

Present Study

This study examines six years'—and 13 studies'—worth of data on EarSketch, a web-based learning environment and curriculum that engages students in introductory computing education within the context of music composition, production, and remixing. These studies ranged considerably by location (i.e. high school classrooms, summer camps, or college courses), mode of delivery (i.e., online MOOC, week-long camp, or multi-week teacher-delivered classroom module), study design (i.e., correlational or quasi-experimental), and participant demographics (e.g., gender, age, race/ethnicity), providing us the opportunity to examine the overall effect of EarSketch and for whom EarSketch works best.

The studies all sought to determine the extent to which an EarSketch-based learning intervention promoted student achievement and engagement across different student populations. The studies analyzed pre/post student content knowledge assessments and a retrospective pre/post engagement survey that measured intent to persist along with constructs such as perceived authenticity, belongingness, and motivation to succeed. Several studies also collected qualitative data through classroom observations, student focus groups, and interviews with students, teachers, and administrators. Two research questions guided this study:

1. How does participation in EarSketch affect students' intentions to persist in computing?

2. What factors influence students' intentions to persist in computing?

EarSketch

EarSketch is a computing education program which aims to engage diverse student populations in introductory computer science through music (Freeman et al., 2014; Freeman, Magerko, & Verdin, 2015; Mahadevan, Freeman, Magerko, & Martinez, 2015; Magerko et al., 2016; McCoid et al., 2013). Students learn the basic elements of computing (i.e., Python or JavaScript code for fundamental computing concepts such as loops and lists) to algorithmically create music in popular genres through sample-based music composition (i.e., composition using musical beats, samples, and effects). EarSketch is a web-based learning platform with a code editor that consists of both text and blocks-based modes; a multi-track digital audio workstation view that shows the musical results of code execution; an audio loop library with 4,000+ musical clips; and an inline curriculum, with both student- and teacher-facing components, that is closely aligned with Computer Science Principles (CSP; college board; Astrachan & Briggs, 2012).

The EarSketch program engages students by providing the opportunity to quickly begin coding and creating music in an environment perceived to be authentic by students (McKlin et al., 2018, Shaffer & Resnick, 1999). Through industry-relevant programming languages and popular music styles and content, students are able to become musically expressive while learning computing. When combined with teachers who have both content knowledge and pedagogical content knowledge in computing (McKlin et al., 2019) in classrooms with strong implementation of the EarSketch curriculum, student attitudes towards computing increase, thereby increasing both their content knowledge and intent to persist in computing. Across the six years of implementation of EarSketch, the program theory was refined. Table 1 describes the program as it was implemented in each evaluation of the program and Figure 1 displays the theory of change behind the EarSketch program.

Figure 1. Theory of Change Model for EarSketch

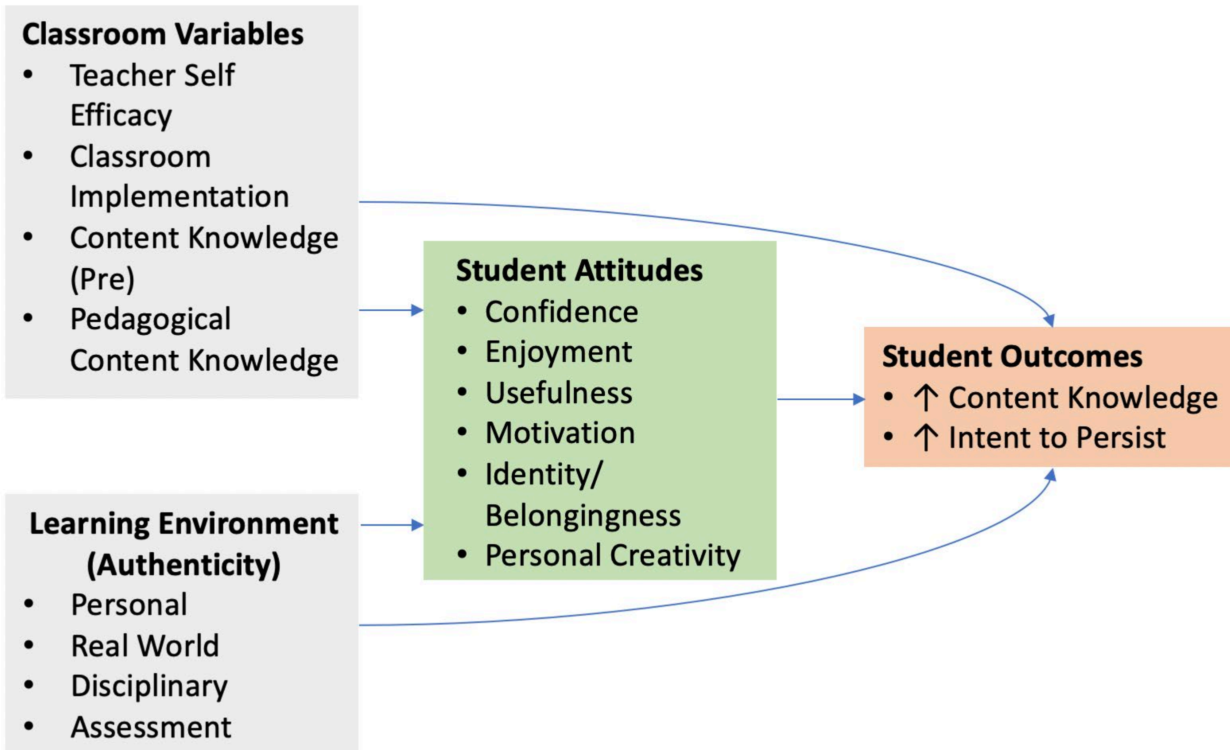


Table 1. Description of all EarSketch Studies

#	Project	Year	n (T)	n (C)	Age	Brief Description and Changes to Curriculum
1	Workshop	Spring 2012	17		HS	Five-day EarSketch workshop.
2	CE21	Spring 2013	69		HS	Eight-week instructional module within a course called "Computing and the Modern World" with little alignment to state and national standards as the software was an early prototype.
3	CE21	Fall 2013	29		HS	Full semester intro to music technology course with little alignment to state and national standards as the software was an early prototype.
4	Summer Camp	June 2013 Beginner I	17		HS	All of these were 1-week summer camps held at Georgia Tech. Early prototype version of both software and curriculum.
5	Summer Camp	June 2013 Advanced	6		HS	The advanced camp only admitted students who had done the 1-week beginner camp already. It covered advanced CS and music topics that have long since been removed from our curriculum b/c they were too advanced for intro CS contexts.
6	Summer Camp	July 2014 Beginner	22		HS	Minor tweaks to curriculum and software as compared to the prior beginner's camp.
7	MOOC	Fall 2014	118		MS through College	This was an intro to music technology MOOC on Coursera taught by one of the authors (Freeman). 2-week module within it provided a condensed intro to EarSketch, focused less on CS learning and more on the connection of CS to experimental music practices. Ages and backgrounds of students varied wildly since the course was open to anyone who wished to enroll online.
8	DRK12	2015-2016	68	30	HS	The DRK-12 and IUSE studies both used the "modern" version of the software (which has evolved and improved from year to year of these studies, but the core has remained the same). The DRK-12 studies used a newly developed curriculum aligned with Computer Science Principles (~10-12 weeks) and closely aligned with GA and AP standards. Implemented in multiple HS classes / schools.
9	DRK12	2016-2017	138	132	HS	CSP curriculum was substantially modified to add more targeted and frequent student assignments and projects.
10	DRK12	2017-2018	360		HS	Only minor tweaks to curriculum.
11	IUSE	Spring 2016	66	47	College	Study was with non-CS majors at an open-access four-year college fulfilling a requirement by taking an intro to programming course. Software was same as for DRK-12 studies, but curriculum was entirely new: full semester that integrated learning with EarSketch along with Python coding in other contexts.
12	IUSE	Fall 2016	58	66	College	Significant curriculum revisions in how EarSketch was integrated into the curriculum and assignments.
13	IUSE	Spring 2017	82	50	College	No significant changes.

Note: T = treatment, C = comparison group. Studies with a comparison group were all quasi-experimental studies; studies with no comparison group were descriptive/correlational studies. MS = middle school, HS = high school.

Methods

Evaluation Procedures

Every evaluation of the EarSketch program focused on how EarSketch affects intent to persist in computing, attitudes towards computing, and students' content knowledge of computing (see McKlin et al. [2019] for more details about the CKA outcomes of EarSketch). Demographic differences were examined across gender and under-represented minority (URM) status. Measurement of intent to persist and attitudes towards computing changed slightly throughout the evaluations as measures were refined given the particular implementation context of the EarSketch program (e.g., for college versus high school students). Later evaluations of the EarSketch program were implemented in high school and college classrooms and included assessments of how well teachers were implementing the EarSketch curriculum; many of these later evaluations also included comparison groups of students in computer science education courses that did not receive the EarSketch curriculum. Furthermore, some of the evaluations included observations, focus groups, and interviews with key personnel and participants to see how well the EarSketch program was implemented and to provide more context on the outcomes of EarSketch program.

Measures

Intent to Persist. The main construct of interest to this paper is Intent to Persist, which was measured on the retrospective pre/post survey administered at the end of each EarSketch project. The retrospective pre/post survey was used to account for response shift bias, which can occur when participants' level of self-knowledge changes as a result of an intervention (Howard et al., 1979; Pratt, McGuigan, & Katzev, 2000). Intent to Persist has been measured slightly differently over the time of EarSketch. See Table 2 for the scale used over the years.

Table 2. Changes to the Intent to Persist Scale across EarSketch Studies

#	Project	Intent to Persist Items
1	Workshop	1. I intend to get a college degree in computing.
2	CE21	2. Someday, I would like to have a career in computing.
3	CE21	3. I can see myself working in a computing field.
4	Summer Camp	4. I intend to take courses related to computing in the future.
5	Summer Camp	5. I intend to go to college.
6	Summer Camp	
7	MOOC	1. I will continue digitally creating music after the course is over.
		2. I will continue programming after the course is over.
		3. I will continue music programming after the course is over.
		4. I will use what I have learned about computer programming in this MOOC after the course is over.
		5. After this course is over, I would like to take a computer programming or computer science course.
		6. After this course is over, I would like to take another music technology course.
8	DRK12	1. I intend to get a college degree in computing.
9	DRK12	2. Someday, I would like to have a career in computing.
10	DRK12	3. I can see myself working in a computing field.
		4. I intend to take courses related to computing in the future.
11	IUSE	1. I intend to minor in computing.
12	IUSE	2. I intend to major in computing.
13	IUSE	3. Someday, I would like to have a career in computing.
		4. I can see myself working in a computing field.
		5. I intend to take courses related to computing in the future.

Originally, the scale consisted of five items and was used for the first six studies examining EarSketch ($\alpha = .87$). The MOOC project involved greatly different items for intent to persist because many of these students were college graduates for whom the original questions would not have been relevant. Revisions included asking whether students would continue digitally creating music, programming, or music programming after the course, use what they learned after the course, and take more CS courses in the future; this scale also had high internal consistency ($\alpha = .87$). For the DRK12 projects, the fifth item was dropped because more constructs were added to the survey and we wanted to keep the survey short. For the IUSE project, the questions were modified to reflect that students were already in college ($\alpha = .93$).

Demographics. Two demographic variables were analyzed in many of the EarSketch projects: gender (i.e., male or female) and under-represented minority (URM) status (i.e., URM students were not White or Asian race/ethnicity). The exceptions were a few projects with a very small number of participants or with only male students (i.e., Spring 2012 workshop, CE21 Fall 2013, the advanced summer camp) and a few projects in which URM status either was not collected or all students were not URM (i.e., Spring 2012 workshop, CE21 Fall 2013, both June 2013 workshops).

Attitudes Towards Computing and Creativity. The student survey measured students' attitudes towards computing and their perceptions that the EarSketch environment is authentic to computing and making music. Attitudes towards computing were measured with 19 items measuring computing confidence, enjoyment, importance and perceived usefulness, motivation to succeed, and identity and belongingness. Although five separate subscales were hypothesized, factor analyses revealed that a one-factor solution seems to fit best with a general "attitudes towards computing" construct (Wanzer, McKlin, Edwards, Freeman, & Magerko, 2019), with an internal consistency of .93.

Previous research suggests that creativity exists at both the level of the *person* and the *place* (Amabile, 1989; Mayer, 1999; Rogers, 1976). As such, the student survey also had a scale measuring both person-level creativity and place-level creativity. This assessment originally used a modified version of the Creativity Support Index (Cherry & Latulipe, 2014), but this scale was replaced by a researcher-created scale to better fit the evaluation studies of EarSketch. Person-level creativity consisted of six subscales measuring the traits, tendencies, and characteristics of the individual who creates something or engages in a creative endeavor (i.e., expressiveness, exploration, immersion/flow, originality/creativity, sharing, creative thinking skills; Engelman et al., 2017). *Place-level creativity* was measured as the extent to which the learning environment is *thickly authentic* (Shaffer & Resnick, 1999) which measures environmental factors that encourage creativity and learning activities that are simultaneously aligned with the interests of the learners, the structure of a domain of knowledge, valued practices, and the modes of assessment use (McKlin et al., 2018).

Teacher-Level Predictors. In the DRK12 projects, teachers were administered a 16-item self-efficacy survey that measures five self-efficacy subscales: computing pedagogical knowledge, instructional, engagement, disciplinary, and outcome expectancy. These subscales, however, were combined into one self-efficacy score. The instrument was administered upon entry into the program and again at the end of EarSketch; however, only post-test administrations were analyzed here. An enactment checklist was administered to teachers for three lessons over the course of the EarSketch program that focuses on curriculum development and implementation (see McKlin et al., 2019b for more information on the classroom implementation variable). Teachers were also administered the same CKA as students to understand their level of understanding of computer programming principles. Finally, teachers' pedagogical content knowledge (PCK) was assessed through an assessment adapted by Sadler and colleagues (2013); see McKlin et al., 2019a for more information on the PCK assessment.

Qualitative. Qualitative data were collected through interviews, focus groups, and observations in the DRK12 projects. Each year, teacher interviews were conducted to gather feedback on professional development, implementation, improvements, and impact at the student, classroom, and school level.

In both the second and third year of the DRK12 EarSketch project, three student focus groups, with between 8 and 12 students in each focus group and each focus group from a different school that participated in the EarSketch program, were conducted. The focus groups gathered students' perceptions of EarSketch and its impact on constructs such as engagement, creativity, collaboration, and communication. As an example, students were asked about their plans to pursue computing in their education or career: "Has this course affected your decision to take computing courses in high school? In college?"

In the fourth year of the DRK12 EarSketch project, teachers were specifically asked how various factors relate to students' intent to persist in computing, such as student attitudes, teacher and student characteristics, authenticity of the EarSketch environment, and classroom implementation. These interviews were conducted to confirm or refute the findings from the Structural Equation Model testing the EarSketch theory of change (see figure 1). A total of 20 teachers were interviewed, and interviews lasted between 30 and 60 minutes.

During Years 3 and 4 of the DRK12 projects, observations of teachers' classrooms were conducted to better understand teaching style, fidelity of implementation, and other classroom implementation factors.

Analytic Procedures

Meta-analysis. All EarSketch studies were included in the meta-analysis, though studies varied in terms of the type of study (i.e., correlational or quasi-experimental), the age of participants (i.e., high school, college-aged participants), and the items used to measure intent to persist. Given that all EarSketch studies were included, we assume there is no bias in terms of the availability of data. Means, standard deviations, and sample sizes were collected for all intention to persist measures across the EarSketch studies and converted into the standardized effect size Hedge's *g*. All analyses were performed in R using the metafor package (Viechtbauer, 2010).

There are two commonly used statistical models for meta-analysis: the fixed-effect model and random-effect model. The fixed-effect model assumes there is a common treatment effect across all studies whereas the random-effect model assumes there is variation in the treatment effect across studies. Typically, a fixed-effect model would be used if we believe all studies are essentially identical and our purpose is not to generalize to other populations; however, a fixed-effect model may also be desirable when the number of studies is small. We decided to present both fixed and random effect models; however, due to the small number of quasi-experimental studies, the meta-analyses examining these studies specifically are examined only using fixed effect models.

Furthermore, we assessed between-study heterogeneity, or the variability of outcomes across studies, using a variety of heterogeneity indices. First, we examined Cochran's Q which is based on the Chi-square distribution and determines whether the studies are drawn from a common population; a significant Q indicates heterogeneity (Hedges & Olkin, 1985). Second, we examined I^2 which is the percentage of total variation in effect sizes across studies due to heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003). Tau-squared (τ^2), the variance of the true effect size, and H^2, the ratio of total variability over the sampling variability, are also reported.

Structural equation modeling (SEM). Structural equation modeling (SEM) was performed using robust maximum likelihood estimation using the lavaan package in R (Rosseel, 2012; Rosseel et al., 2018). Model fit was measured on multiple indices including the robust Comparative Fit Index (CFI), robust Tucker-Lewis Index (TLI, also known as the NNFI), robust Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square

Residual (SRMR). Goal values for the CFI and TLI were greater than .90, with values above .95 preferred. Goal values for the RMSEA and SRMR were less than .08, with values below .06 preferred.

Multilevel modeling (MLM). Multilevel modeling (MLM) examines data that has a hierarchical structure. In the case of the DRK12 and IUSE projects, students (level 1) are nested within teachers/classrooms (level 2) which are nested in schools and districts (level 3; however, there were insufficient schools or districts to warrant a 3-level MLM). Due to the nested structure of the data, and the supported notion that students within classrooms/schools are more similar to each other than students across classrooms/schools, MLM analyses account for the violated assumption of independence of observations in regression analyses.

MLMs were performed using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015) in which students are nested within classrooms. Three separate models were performed. The first model includes only the teacher/classroom variable and is used to determine the intra-class correlation (ICC), which is the percentage of the outcome at the student-level that is predicted by membership in the classroom. The second model examined all student-level variables to determine the variance of the outcome variable accounted for by the student-level predictors. The third model added in the classroom-level variables to determine the added variance explained by the teacher-level predictors.

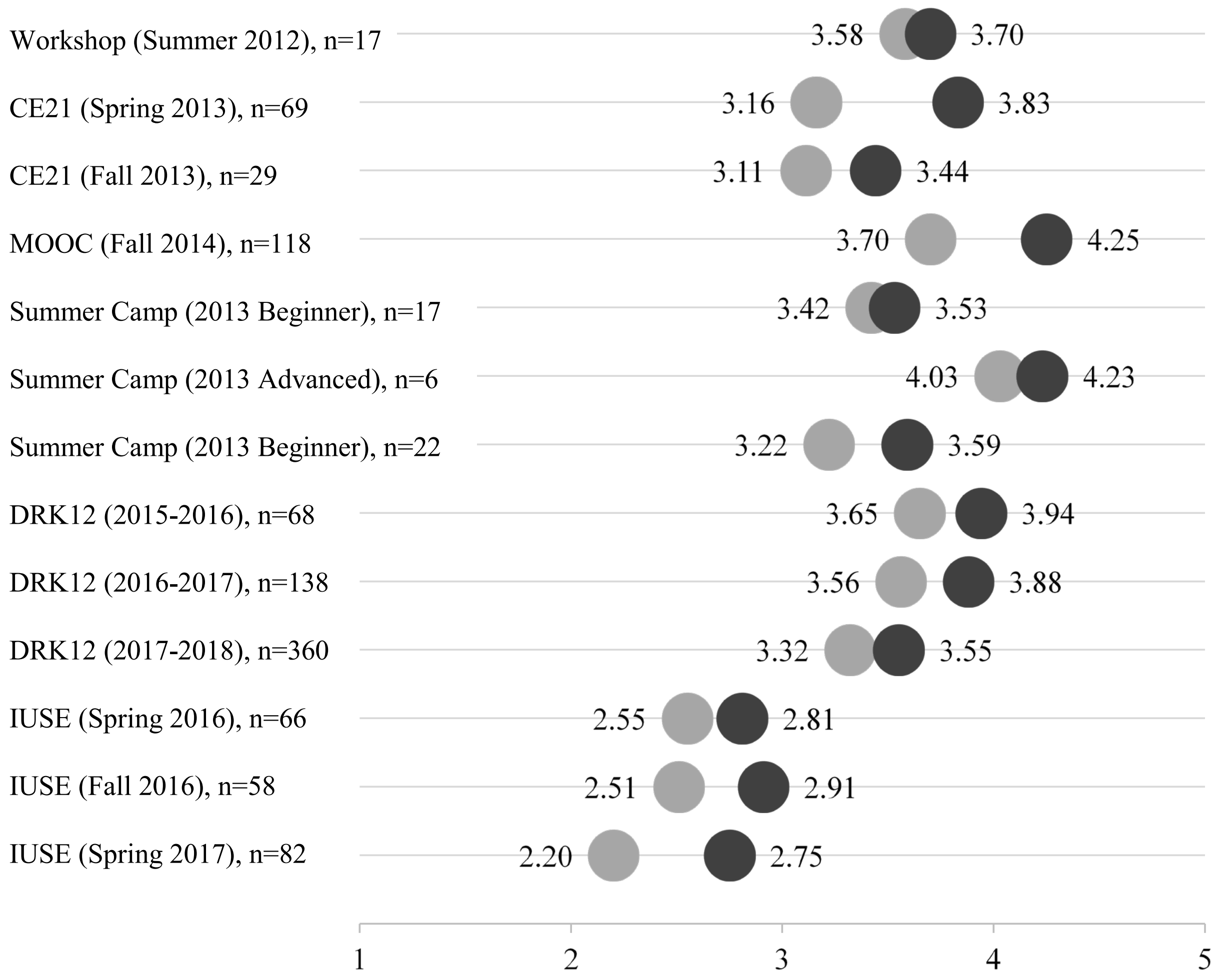
Qualitative. All interviews and focus group audio were transcribed verbatim. The coding schemas and data analysis techniques are informed by grounded theory methodology (Glaser & Strauss, 1967). All focus group and interview data were analyzed during two cycles. The first cycle of coding involved creating attribute and structural codes (Saldaña, 2013). Attribute codes indicate characteristics such as school district name, teachers' number of years teaching CSP and number of years using EarSketch, students' school, and gender. Structural codes identify content-based phrases related to the purposes for the interviews and focus groups. These codes were created based on the topics of the interview guide including differing interest, engagement, intent to persist, factors that enhance teaching, etc. During the second cycle of coding, pattern coding was employed to develop major themes in response to the research questions.

Results

How does participation in EarSketch affect students' intentions to persist in computing?

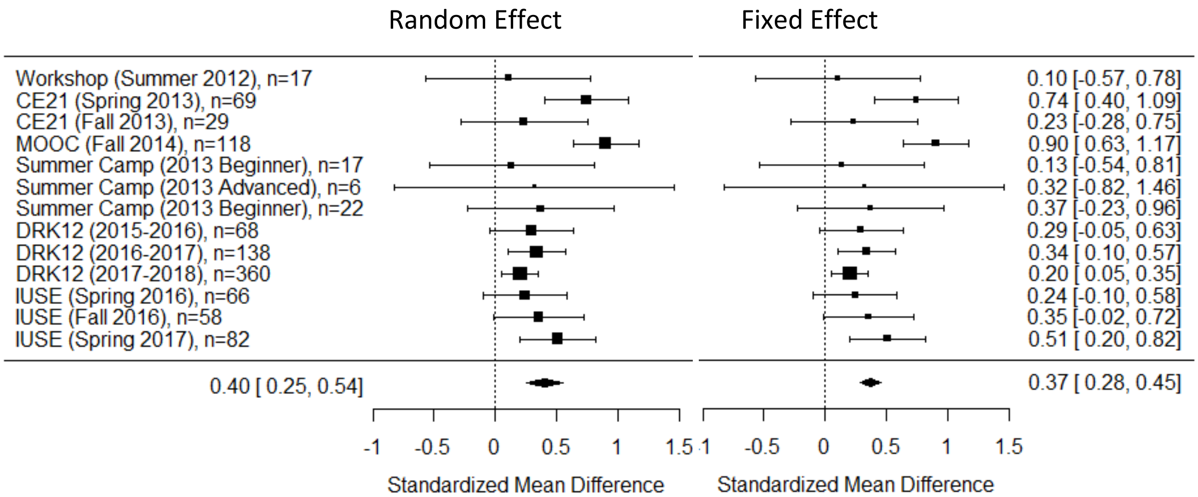
Across all studies of EarSketch, students' intentions to persist in computing increased, although changes of ratings from the retrospective pretest scores to posttest scores varied across studies (see Figure 2). Furthermore, college students' intentions to persist (i.e., the IUSE studies) were markedly lower than intentions to persist in other studies, most likely due to slight modifications in the scale for use in a college setting (see Table 2).

Figure 2. Average intent to persist across all EarSketch studies



Multiple heterogeneity indices indicated that studies were not homogenous: Q (12) = 27.46, p= .007; I^2 = 56% (CI = 9%, 79%); τ^2 = .036 (.003, .108); H^2 = 2.27 (1.09, 4.82). This indicates that there is variation between the studies, which may be due to changes in the implementation of EarSketch over time, the variability in study designs, and the variability of participants. However, we proceeded with both random and fixed effect meta-analytic models given both the heterogeneity of studies but also the small number of studies included and because all studies had a commonality of studying the EarSketch program. Under a random effect model, the overall Hedge's g effect size for intent to persist for all 14 studies was .40 (.25, .54); under a fixed effect model, it was g = .37 (.28, .45).

Figure 3. Forest plot for EarSketch Intentions to Persist, Random Effect and Fixed Effect Models



Note: Forest plots include the line of null effect (i.e., an effect size of zero) and the overall meta-analytic effect size (i.e., the diamond at the bottom of the plot), with the width of the effect size indicated by the width of the diamond. Each individual study's effect size is also plotted, along with the 95% confidence interval of the effect size. The size of the effect size square indicates the weight of that effect size in determining the overall meta-analytic effect size.

Focus groups with students provide further evidence of the effect that EarSketch had on their intent to persist. Students shared how EarSketch helped change their perceptions of computing and increase their desire to pursue computing as a career:

> *I thought coding was going to be boring and kind of just make me super-mad. It was going to be like tragic. But now that I've taken this class and I've seen all the things I can do with EarSketch and how that can be applied, like the same general concepts can be applied and expanded on to all these other aspects and different fields. It kind of opened up and made me kind of rethink my career choices like, "Oh, maybe I actually want to pursue something in like IT or computer science." Normally, you have like a one-sided opinion or view of coding. You don't really see it as being something creative and so personable…. It just kind of opened up your world, like broadened your horizons in seeing all the career fields that actually use coding and how that plays a role it, versus like this stereotypical view of what coding is.*

> *But EarSketch showed me coding cannot just be used for IT, but for anything else; making software, making music. So, that really opened up a different world, you could say, for coding. Now, I feel like I have a chance for something else. Like it's not just IT. I'll do something cool.*

One graduating senior shared, ”That’s what I’ll major in, computer science, because of this course...Before I didn’t know what I was going to do. But then I did this and it was really fun.” While some students expressed a desire to work as traditional programmers, other students realized that computing might be combined or useful with other careers:

> *I feel like it’s been useful for the future, because I know somebody. They’re like a police officer. They’re like an investigator. But she never took law enforcement. She just took computing...but she was able to go into that career path because computing kind of leaped into all of it. I was looking up different careers. For example, law enforcement stuff, they want people that actually have a background in computer science.*
>
> *I want to be a computer programmer. First of all, I was really interested in how you can create a website. It looks really different for each and every one. Then I also want to major in environmental science. So, I was thinking about combining environmental science with computer science and create something.*

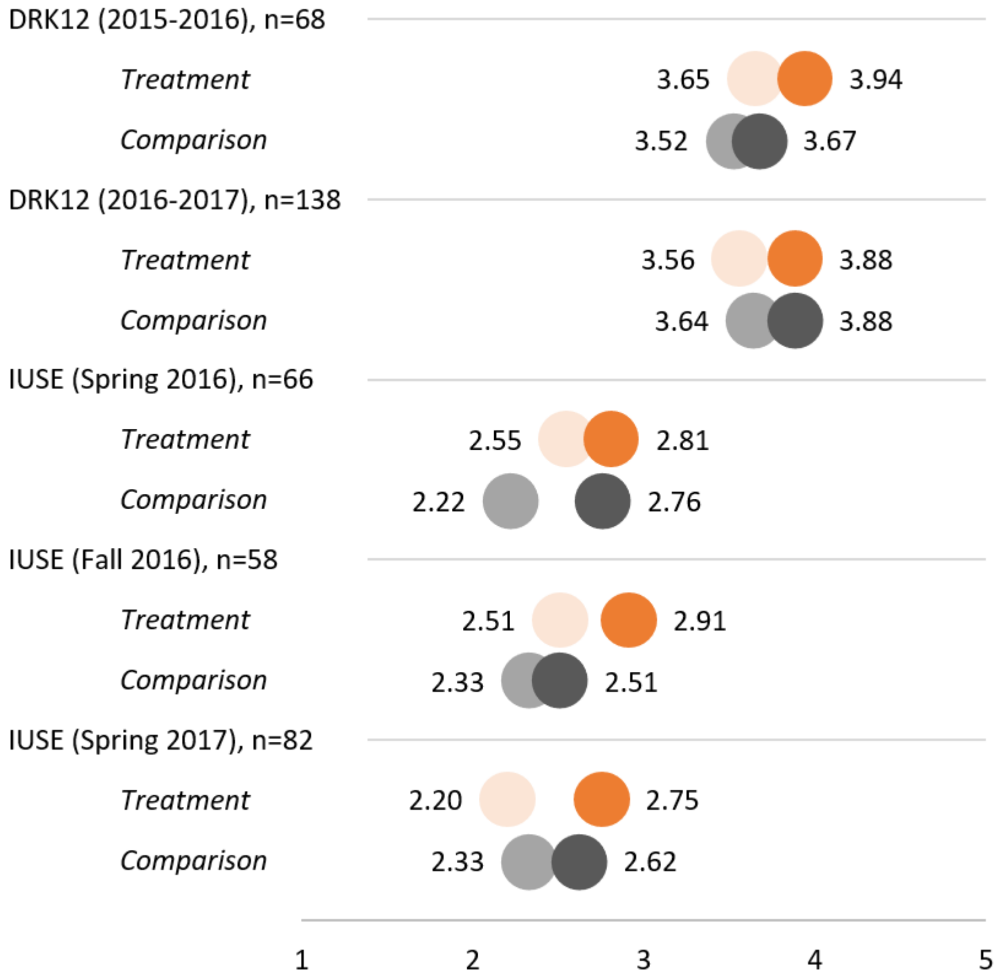
What factors influence students’ intentions to persist in computing?

Two-way ANOVA: First, mean differences by gender and URM status were explored using a two-way ANOVA both before and after the program. There was a significant main effect of gender both before ($F[1, 1549] = 94.25$, $p < .001$) and after ($F[1, 1531] = 74.04$, $p < .001$) the program. Male students rated intentions to persist higher than female students both before (M = 3.27, SD = 1.06 vs M = 2.73, SD = 1.21) and after (M = 3.63, SD = 1.09 vs M = 3.11, SD = 1.24) the program. There was also a significant main effect of URM status both before ($F[1, 1549] = 3.94$, $p = .047$) and after ($F[1, 1531] = 10.06$, $p = .002$) the program. Non-URM students rated intentions to persist higher than URM students both before (M = 3.18, SD = 1.09 vs M = 3.03, SD = 1.16) and after (M = 3.58, SD = 1.09 vs M = 3.36, SD = 1.21) the program. However, there was no significant interaction between gender and URM status, either before ($F[1, 1549] = 1.18$, $p = .277$) or after ($F[1, 1531] = .79$, $p = .375$) the program, suggesting that, for instance, female URM students did not have significantly lower intent to persist than female non-URM students.

Meta-analysis moderators: Second, three separate moderators—treatment vs comparison groups, gender, and URM status—were examined using meta-analysis. For the treatment versus comparison moderation analysis, only the five studies with a comparison group were included; given the small number of studies, only a fixed effect model was analyzed. Figure 4 shows the retrospective pretest and post-test scores for the five quasi-experimental studies across treatment and comparison groups. Students in the treatment had a greater increase in intentions to persist across all studies except for the first IUSE study where the comparison group had a larger increase in intent to persist[1]. However, in the meta-analysis results, participants in the treatment group did not have significantly greater intent to persist compared to the comparison group (i.e., students receiving a computer science education with a curriculum other than EarSketch), *g* = .08 (-.07, .23); see Figure 5 for the forest plot of the fixed effect model.

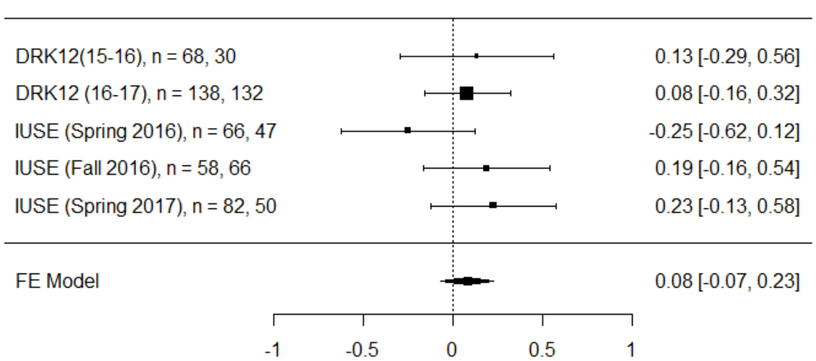
Figure 4. Average intentions to persist for treatment (orange) and comparison (grey) groups across all five quasi-experimental studies

[1] However, this is likely due to implementation issues. The first semester of the IUSE project implemented an early version of the revised EarSketch curriculum geared for college students. Very little of the classroom content was focused on EarSketch and most of it was based on teachers' previous non-EarSketch approach. These issues were resolved in the subsequent IUSE studies with a new e-book that was developed which more significantly integrated EarSketch into the course curriculum. When examining the meta-analysis with this study removed, the treatment group had slightly higher intent to persist compared to the comparison group, *g* = .14 (-.02, .30), but this was not statistically significant.



Note: Light orange dots reflect the retrospective pretest scores of the treatment group and dark orange dots reflect the posttest scores of the treatment group. Light grey dots reflect the retrospective pretest scores of the comparison group and dark grey dots reflect the posttest scores of the comparison group.

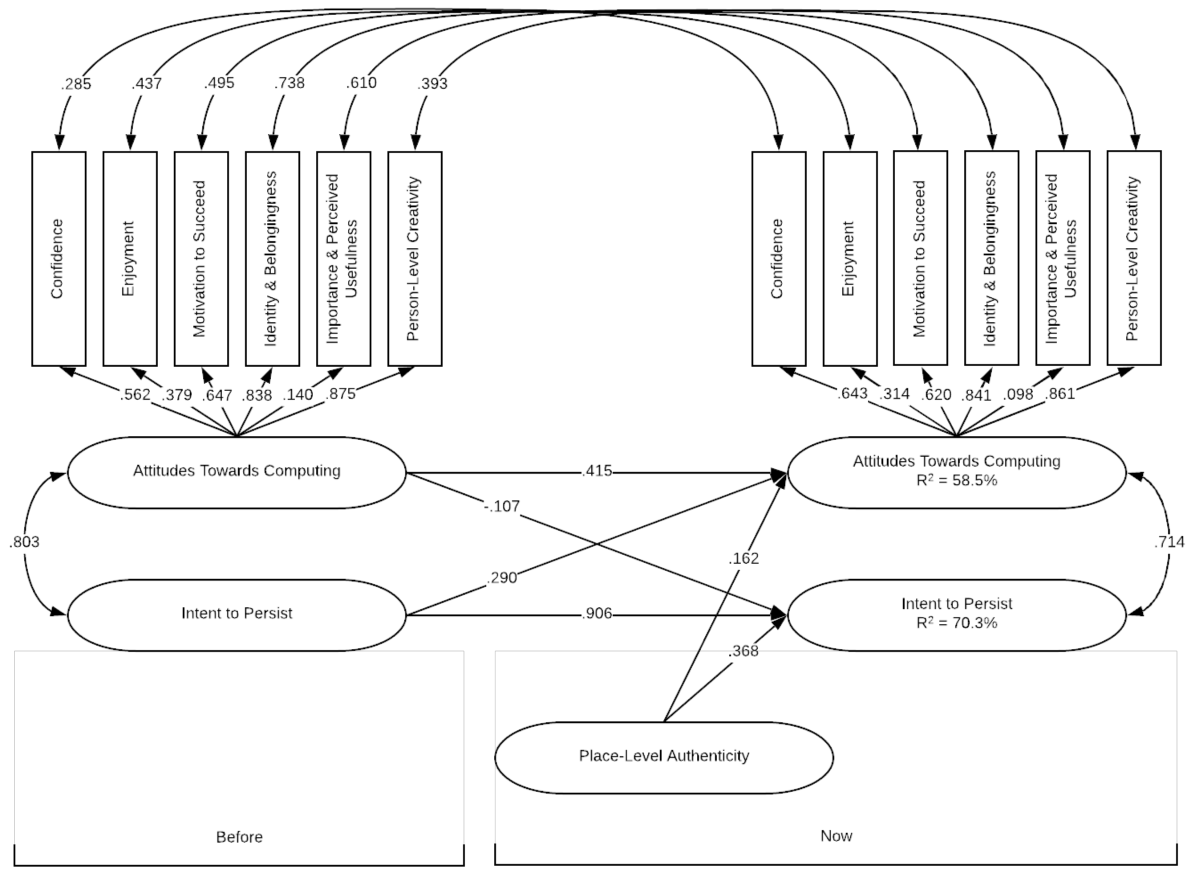
Figure 5. Fixed effect forest plot for intentions to persist for all five quasi-experimental studies



Furthermore, when examining all EarSketch studies (minus the few studies with no data on gender or URM status), there were no significant differences in intentions to persist by gender (n = 11 studies, g_{fixed} = .02 [-.16, .21], and g_{random} = .01 [-.23, .24]) or by URM status (n = 10 studies, g_{fixed} = -.04 [-.23, .14], and g_{random} = .01 [-.24, .26]). Furthermore, there were no gender or URM status effects when examining the five quasi-experimental studies. This suggests that EarSketch is having similar effects on students regardless of demographic. Given the heterogeneity of the effect sizes across studies, differences in effect sizes are likely attributable to other characteristics such as study design, implementation quality or design of EarSketch, or other personal characteristics, including the other aspects of TPB (i.e., perceived behavioral control or subjective norms about computer science).

Structural Equation Modeling: After the DRK12 and IUSE projects, the theory of change was tested using SEM to examine how authenticity and attitudes towards computing predicted intent to persist (see Figure 6). This was not tested with the other EarSketch projects because they did not measure the authenticity of the EarSketch environment. The model tested included each of the five attitudes towards computing subscale composites, as well as a composite of person-level creativity, loading onto the latent factor “Attitudes towards computing” at both retrospective pretest and posttest. Intent to persist was measured at both retrospective pretest and posttest. Furthermore, the authenticity of the environment was included in the model but was only measured at posttest. Measures at retrospective pretest were correlated, and measures at posttest were correlated. Both measures at retrospective pretest were regressed onto both measures at posttest; furthermore, authenticity of the environment was regressed onto both measures at posttest.

Figure 6. SEM model of predictors of Intent to Persist (All DRK12 and IUSE Studies, n = 1223)



Overall, the data fit the model well, χ^2 (79) = 709.80, $p < .001$, CFI = .945, TLI = .927, RMSEA = .079 [90% CI: .073, .086], SRMR = .075. Overall, attitudes towards computing now was best predicted by attitudes towards computing before the program ($\beta = .415$) and intent to persist now was best predicted by intent to persist before the program ($\beta = .906$). Attitudes towards computing now was also predicted by both intent to persist before ($\beta = .290$) and place-level authenticity ($\beta = .162$). However, intent to persist was also negatively predicted by attitudes towards computing ($\beta = -.107$)—likely due to a suppression effect since attitudes towards computing were positively correlated with intent to persist—and place-level authenticity ($\beta = .368$). This model suggests multiple factors related to intent to persist after EarSketch, including prior intentions to persist, authenticity of the EarSketch environment, and attitudes towards computing.

Multilevel Modeling: During the last year of the DRK12 project (2017-18), the theory of change was further tested by examining how the teacher-level variables of teacher self-efficacy, content knowledge, and classroom implementation further impacted intent to persist above and beyond authenticity and attitudes towards computing. First, the intra-class correlation scores for intentions to persist indicated that 11.2% of the student variation in intentions to persist were due to the classroom/teacher of the student. In the first step of the hierarchical MLM, only student-level variables were included; overall, authenticity and attitudes towards computing were significant predictors of intent to persist after EarSketch. When adding in teacher-level variables, none of the variables were statistically significant; however, the inclusion of these variables explained an additional 16% of the variance in students' intent to persist between classrooms.

Table 6. MLM Results for Intent to Persist as an Outcome

Model		**1**	**2**
AIC		971.2	729.1
Between-Teacher Variance		54%	70%
Within-Teacher Variance		33%	34%
Student-Level Variables			
	Authenticity	**.189 (.04) *****	**.149 (.05) ****
	Attitudes (Before)	**.484 (.04) *****	**.546 (.05) ****
Teacher-Level Variables			
	Self-efficacy		-.055 (.07)
	Implementation		.086 (.07)
	Pre CKA		.049 (.12)
	PCK		.052 (.12)

Note: for student-level and teacher-level predictors, the values shown are the Standardized Beta (standardized error). * indicates a statistically significant Beta at $p < .05$.

Interviews and focus groups: Using a sequential explanatory design (Creswell, 2014), qualitative data was used to aid in interpretation of the quantitative findings. Six focus groups with students ($n_{students} = 60$) and interviews with six teachers in the second and third years of the DRK12 EarSketch project provided further evidence that most students who entered EarSketch already had a strong desire to pursue computing. During interviews, many students confirmed that their intent to pursue computing was not impacted by their experiences with EarSketch; students who intended to pursue computing before possessed a desire to pursue computing after participating in EarSketch:

> *I wanted to do it [programming] when I came in here, and I still want to do it.*
>
> *I find [programming] interesting... I still would have continued on with computer science even without the encouragement from this class.*

Survey results support these findings. A total of 44% of students agreed (i.e., average score > 3) that they intended to persist in computing both before and after participating in EarSketch whereas 34% of students disagreed they intended to persist in computing both before and after participating in EarSketch. Only 20% of participants disagreed before participating in EarSketch they intended to persist in computing and then switched to agreeing after EarSketch they intended to persist in computing.

During interviews, teachers confirmed that EarSketch solidified the desire to pursue computer science for students with prior interest.

> *The students are hungry for knowledge. The students came in very much interested in gaming…The fact that they play games makes them want to learn how to develop a game. So, the CS Principles class was like an introduction for them into coding. They enjoyed it because of the musical aspect of it. So, I had to really let them know it's the beginning into coding for games.*
>
> *Well, a lot of my students, I'd probably say close to 30% of them or so have parents who work in a similar field, either in sales, or customer support, or actually being an engineer or a programmer themselves. So, they already see themselves as that being a real possibility…So, they already see themselves as inevitably being a part of that field. These are just stepping stones to get there…So, they don't need further motivation. I don't have to make long-winded speeches about them coming into the field. They're kind of just wanting to get through the coursework in order to get there already.*

Discussion

This study examined the effectiveness of the EarSketch learning environment and curriculum for promoting students' intent to persist in computing. A total of 13 studies from 2012 through 2018 were analyzed using meta-analyses to determine whether students in EarSketch had greater intent to persist after participating in the program. Furthermore, structural equation modeling and qualitative interviews and focus groups were analyzed to examine individual and contextual factors that also play a role in promoting intent to persist.

Overall, this study found that students' ratings of their intentions to persist in computing increased from before to after participating in EarSketch ($g = .40$ [.28, .54]). However, examination of the five quasi-experimental studies suggested that students in EarSketch did not experience greater increases in intentions to persist compared to students in other computer science education classes using a different curriculum ($g = .08$ [-.07, .23]). Future research should examine why there were no significant differences between students in the EarSketch program and students not in the EarSketch program. Although the results of this analysis suggest EarSketch has no added effect on students' intent to persist, there was no randomization of students or even classrooms and there were many differences between treatment and comparison classrooms beyond the presence or absence of the EarSketch curriculum. There may be other factors that can explain why no treatment effect was found. Furthermore, the comparison groups were receiving a computer science education curriculum similar to EarSketch without the added music component. Thus, these quasi-experimental studies compared the effects of EarSketch to other similar programs rather than EarSketch to no program.

Furthermore, a two-way ANOVA suggested both male and non-URM students rated intentions to persist higher both before and after participation in EarSketch compared to female and URM students, respectively, but no demographic group had a significant increase in intent to persist over another. This suggests that EarSketch is not producing any additional benefit (or, conversely, detriment) for any one demographic group, but is rather increasing students' intent to persist equally among females and male students and URM and non-URM students. Using the program's theory of change as a guide, students' attitudes towards computing and perceived authenticity of the EarSketch environment influenced students' intent to persist. Although teachers' implementation of EarSketch and knowledge about computing and pedagogy explained an additional 16% of the variance in intent to persist between teachers, none of the teacher-related variables were statistically significant.

Implications

Although intent to persist in computing increased from participating in EarSketch, results suggested that students may have already had an intention to persist in computing prior to joining the program. For example, only 20% of students in the DRK12 project went from disagreeing they intended to persist to agreeing they intended to persist from before to after participating in EarSketch. This suggests that EarSketch--and perhaps other computing education programs—may perform better at solidifying students' intent to persist rather than producing their intent to persist. Rather, schools and programs may need to find methods to encourage students who have no interest in pursuing a degree or job in computing to take a computing course in the first place to build up their interest. For this reason, states that are requiring computing education courses as a high school graduation requirement may see an increased number of students interested in pursuing computing degrees or jobs. Other innovative approaches, ranging from museum-based programs and afterschool clubs to contests, are also an essential part of the ecosystem of educational offerings that can drive students towards taking their first computing course.

Strengths and Limitations

Cronbach argued that evaluations of programs should not conduct one large experimental or quasi-experimental study, but rather a "fleet of studies" (Cronbach & Associates, 1980, p. 7). This study epitomizes this view by conducting 13 evaluation studies across six years to examine how well EarSketch works, for whom EarSketch is working best, and how to improve EarSketch. Variations in the evaluation studies across the years has produced a version of the EarSketch curriculum that students and teachers like, teachers can implement well, and improves students' intent to persist in computing, as well as their computing knowledge and attitudes towards computing. Meta-analysis aligns itself well to this fleet of studies approach, allowing us to examine the effectiveness of EarSketch across the years as the curriculum and implementation was improved. Furthermore, this fleet of studies approach allows for examining specific components or processes of the EarSketch program, such as the role of classroom implementation (McKlin et al., 2019b) and teachers' pedagogical content knowledge (McKlin et al,, 2019a), the authenticity of the EarSketch environment (Engelman et al., 2017), and how EarSketch engages URM students (Freeman et al., 2014).

However, with the fleet of studies approach comes some limitations. The meta-analysis performed was not able to account for the changes to the implementation of the program over time. Effect sizes seem to be largely clustered towards the last three years of implementation of EarSketch, when the curriculum was mostly solidified, compared to the first three years of implementation when sample sizes were small, confidence intervals large, and the effect size estimates ranging from .10 to .90. Furthermore, the environments in which EarSketch were examined ranged considerably, from a MOOC and summer camps to implementation in high school and college classrooms; this study was unable to account for these differences in environments, nor did it attempt to examine age-related differences in implementation because implementation varied by age to account for developmental appropriateness.

Two other limitations warrant mentioning. First, all the studies used a retrospective pretest to account for potential response shift bias. This may lead to inflated effect sizes compared to a true pretest, but it also may be more appropriate if participants' understanding of a subject change throughout the program. Future research will examine to what extent there is a response shift bias with intent to persist and attitudes towards computing by comparing true pre-posttests to retrospective pre-post responses. Second, there are limitations regarding the five quasi-experimental studies, particularly for the DRK12 studies. Comparison classrooms consisted of either teachers who were not interested in using EarSketch because they were committed to another approach, or were teachers who were interested in EarSketch but, because of an agreement with their school district, were told they had to first teach something else for a year for purposes of our study. This may mean that students were not as similar across the treatment or comparison groups and that students in comparison classrooms could be using widely different non-EarSketch curricula. Furthermore, there was no true control group in any of the five quasi-experimental studies; rather, EarSketch was compared to a similar computing course using a curriculum EarSketch but without the added music component. Thus, we do not know the effectiveness of EarSketch compared to not being in any computing education program. Overall, given the limitations with the quasi-experimental studies, there are many other variables that could be affecting differences between treatment and comparison groups analyzed in this study.

Conclusions

This study demonstrates the importance of examining whether and to what extent computing education programs can promote students' intentions to persist in computing, both in college majors and future careers. Given the importance of increasing persistence in computing majors and careers--and STEM majors and careers more broadly--we need to understand what we can do as computer science educators to instill and foster this interest in computing. This study finds that EarSketch is one computing education program and curriculum that can help students continue to pursue computing into the future.

References

Ahadi, A., Hellas, A., Ihantola, P., Korhonen, A., & Petersen, A. (2016, November). Replication in computing education research: researcher attitudes and experiences. In *Proceedings of the 16th Koli Calling International Conference on Computing Education Research* (pp. 2-11).

Ajzen, I. (2001). Nature and operation of attitudes. *Annual Review of Psychology*, *52*(1), 27–58. https://doi.org/10.1146/annurev.psych.52.1.27

Amabile, T.M. (1989*). Growing Up Creative*. New York: Crown.

Armitage, C. J. & Conner, M. (2000). Social cognition models and health behavior: A structured review. *Psychology & Health, 15*(2), 173-189.

Astrachan, O., & Briggs, A. (2012). The CS principles project. *ACM Inroads*, *3*(2), 38-42.

Bandura, A., Barbaranelli, C., Caprara, G.V. and Pastorelli, C. (2001) Self-Efficacy Beliefs as Shapers of Children's Aspirations and Career Trajectories. *Child Development, 72*, 187-206. http://dx.doi.org/10.1111/1467-8624.00273

Barker, L. J., McDowell, C., & Kalahar, K. (2009). Exploring factors that influence computer science introductory course students to persist in the major. In *Proceedings of the 40th ACM Technical Symposium on Computer Science Education* (pp. 153-157). ACM.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effets models using lme4. *Journal of Statistical Software, 67*(1), 1-48.

Beal, S. J., & Crockett, E. J. (2010). Adolescents' occupational and educational aspirations and expectations: Links to high school activities and adult educational attainment. *Developmental Psychology, 46*(1), 258-265. https://doi.org/10.1037/a0017416

Bean, J. (1982). Student attrition, intentions, and confidence: Interactions effects in a path model.

Research in Higher Education, 17, 291-320.

Blinksenstaff, J. C. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and Education, 17*(4), 369-386.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: John Wiley & Sons, Ltd.

Bottia, M. C., Stearns, E., Mickelson, R. A., Moller, S., & Valentino, L. (2015). Growing the roots of STEM majors: Female math and science high school faculty and the participation of students in STEM. *Economics of Education Review, 45*, 14-27.

Bottia, M., Stearns, E., Mickelson, R., Moller, S., &Parker A. (2015). The Relationships among High School STEM Learning Experiences and Students Intent to Declare and Declaration of a STEM Major in College. *Teachers College Record, 117,* 1-46.

Cabrera, A.F., Castaneda, M.B., Nora, A., & Hengstler, D. (1992). The convergence between two theories of college persistence. *Journal of Higher Education, 63*, 143-164.

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, *2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Card, N. A. (2017). Replication and research accumulation in developmental science. In N. A. Card (Ed.), Developmental Methodology, *Monographs of the Society for Research in Child Development, 82*(2), 105-121.

Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues*, *67*(3), 469-491. https://doi.org/10.1111/j.1540-

4560.2011.01710.x

Cherry, E., & Latulipe, C. (2014). Quantifying the creativity support of digital tools through the creativity support index. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *21*(4), 21.

Cheryan, S., Master, A., & Meltzoff, A. N. (2015). Cultural stereotypes as gatekeepers: Increasing girls' interest in computer science and engineering by diversifying stereotypes. *Frontiers in Psychology, 6*. https://doi.org/10.3389/fpsyg.2015.00049

Clear, T. (2006). Valuing computer science education research? In *Proceedings of the 6th Baltic Sea Conference on Computing Education Research: Koli Caling,Baltic Sea* (pp. 8-18).

Cronbach, L. J., and Associates. 1980. *Toward reform of program evaluation*. San Francisco: Jossey-Bass.

DuBow, W., Weidler-Lewis, J., & Kaminsky, A. (2019). Multiple factors converge to influence women's persistence in computing: A qualitative analysis of persisters and nonpersisters. In *Proceedings of the 2016 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT).*

Eccles, J. S., Vida, M. N., & Barber, B. (2004). The relation of early adolescents' college plans and both academic ability and task-value beliefs to subsequent college enrollment. *Journal of Early Adolescence, 24*(1), 63-77.

Engelman, S., Magerko, B., McKlin, T., Miller, M., Edwards, D., & Freeman, J. (2017). Creativity in Authentic STEAM Education with EarSketch. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (pp. 183-188). ACM.

Freeman, J., Magerko, B., & Verdin, R. (2015). Computer Science Principles With EarSketch. In

Proceedings of the 46th ACM Technical Symposium on Computer Science Education (pp. 710-710). ACM.

Freeman, J., Magerko, B., McKlin, T., Reilly, M., Permar, J., Summers, C., & Fruchter, E. (2014). Engaging underrepresented groups in high school introductory computing through computational remixing with EarSketch. In *Proceedings of the 45th ACM technical symposium on Computer science education* (pp. 85-90). ACM.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory strategies for qualitative research*. Mill Valley, CA Sociology Press.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal, 327*, 557-560.

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences, 106*(22), 8801-8807.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine, 2*(8), e124.

Jagacinski, C. M., Lebold, W. K., & Salvendy, G. (1988). Gender differences in persistence in computer-related fields. *Journal of Educational Computing Research, 4*(2), 185-202.

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological science*, *23*(5), 524-532.

Lee, J., & Shute, V. J. (2010). Personal and social-contextual factors in K-12 academic performance: An integrative perspective on student learning. *Educational Psychologist, 45*(3), 185-202. http://dx.doi.org/10.1080/00461520.2010.493471

Litzler, E., Samuelson, C., & Lorah, J. (2014). Breaking it down: Engineering student STEM confidence at the intersection of race/ethnicity and gender. *Research in Higher Education, 55*, 810–832.

Magerko, B., Freeman, J., Mcklin, T., Reilly, M., Livingston, E., Mccoid, S., & Crews-Brown, A. (2016). Earsketch: A steam-based approach for underrepresented populations in high school computer science education. *ACM Transactions on Computing Education (TOCE)*, *16*(4), 14.

Mahadevan, A., Freeman, J., Magerko, B., & Martinez, J. C. (2015). EarSketch: Teaching computational music remixing in an online Web Audio based learning environment. In *Web Audio Conference*.

Maltese, A. V., & Tai, R. H. (2011). Pipeline persistence: Examining the association of educational experiences with earned degrees in STEM among U.S. students. *Science Education, 95*, 877-907.

Margulieux, L., Ketenci, T., & Decker, A. (2019). Review of Measurements Used in Computing Education Research and Suggestions for Increasing Standardization. *Learning Technologies Division Faculty Publications, 20*, 1-30.

Mayer, R.E. (1999). Fifty years of Creativity Research. In Sternberg, R.J. (Ed.). *Handbook of Creativity* (pp. 449-60). Cambridge: Cambridge University Press.

McCoid, S., Freeman, J., Magerko, B., Michaud, C., Jenkins, T., Mcklin, T., & Kan, H. (2013). EarSketch: An integrated approach to teaching introductory computer music. *Organised Sound*, *18*(2), 146-160.

McInerney, D. (2008). Personal investment, culture and learning: Insights into school achievement across Anglo, Aboriginal, Asian and Lebanese students in Australia.

International Journal of Psychology, 43, 870–879.

McKlin, T., Lee, T., Wanzer, D., Magerko, B., Edwards, D., Grossman, S., ... & Freeman, J. (2019a). Exploring the Correlation Between Teacher Pedagogical Content Knowledge and Content Knowledge in Computer Science Classrooms. In *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education* (pp. 315-315). ACM.

McKlin, T., Wanzer, D., Lee, T., Magerko, B., Edwards, D., Grossman, S., & Freeman, J. (2019b). Implementing EarSketch: Connecting classroom implementation to student outcomes. In *Proceedings of the 50th ACM Technical Symposium on Computer Science Education (SIGCSE '19)*. ACM.

McKlin, T., Magerko, B., Lee, T., Wanzer, D., Edwards, D., & Freeman, J. (2018). Authenticity and personal creativity: How EarSketch affects student persistence. In *Proceedings of the 49th ACM Technical Symposium on Computer Science Education* (pp. 987-992). ACM.

Montaño, D. E., & Kasprzyk, D. (2008). Theory of reasoned action, theory of planned behavior, and the integrated behavioral model. In K. Glanz, B. K. Rimer, & K. Viswanath (Eds.), Health behavior and health education: Theory, research, and practice (pp. 67-96). San Francisco, CA, US: Jossey-Bass.

National Academy of Sciences, National Academy of Engineering, and Institute of Medicine. (2011). *Expanding underrepresented minority participation: America's science and technology talent at the crossroads.* Committee on Underrepresented Groups and the Expansion of the Science and Engineering Workforce Pipeline. Committee on Science, Engineering, and Public Policy and Policy and Global Affairs. Washington, DC: The National Academies Press.

Nosek, B. A., & Läkens, D. (2014). Registered reports. *Social Psychology, 45*, 137-141.

Olitsky, S., Loman, L. F., Gardner, J., & Billiups, M. (2010). Coherence, contradiction, and the development of school science identities. *Journal of Research in Science Teaching, 47*, 1209–1228.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716–aac4716. https://doi.org/10.1126/science.aac4716

Papanastasiou, E., & Zembylas, M. (2004). Differential effects of science attitudes and science achievement in Australia, Cyprus, and the USA. *International Journal of Science Education, 26*, 259–280.

Pashler, H., & Wagenmakers, E.-J. (2012). Editors' introduction to the special section on replicability in psychological science: A crisis of confidence? *Perspectives on Psychological Science, 7*, 528-530.

Porter, S.R., & Swing, R.L. (2006). Understanding how first-year seminars affect persistence. *Research in Higher Education, 47*(1), 89-109.

Robnett, R. D., & Leaper, C. (2012). Friendship groups, personal motivations, and gender in relation to high school students' STEM career interest. *Journal of Research on Adolescence, 23*(4), 652-664.

Rogers, C. (1976). Toward a Theory of Creativity. In Rothenberg, A. and Hausman. C.R. *The Creativity Question*. Durham, NC: Duke University Press.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Rosseel, Y., Oberski, D., Byrnes, J., Vanbrabant, L., Savalei, V., Merkle, E., … Jorgensen, T. D. (2018). lavaan: Latent Variable Analysis (Version 0.6-2).

Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The Influence of Teachers' Knowledge on Student Learning in Middle School Physical Science Classrooms. *American Educational Research Journal*, *50*(5), 1020–1049. https://doi.org/10.3102/0002831213477680

Sadler, P. M., Sonnert, G., Hazari, Z., & Tai, R. (2012). Stability and volatility of STEM career interest in high school: A gender study. *Science Education, 96*, 411-427.

Saldaña, J. (2013). *The coding manual for qualitative researchers*. Thousand Oaks, CA: Sage.

Scott, A. B., & Mallinckrodt, B. (2005). Parental emotional support, science self-efficacy, and choice of science major in undergraduate women. *The Career Development Quarterly*, *53*(3), 263-273.

Shaffer, D.W., & Resnick, M. (1999). "Thick" authenticity: New media and authentic learning. *Journal of Interactive Learning Research*, *10*(2), 195-215.

Shaw, E. J., & Barbuti, S. (2010). Patterns of persistence in intended college major with a focus on STEM majors. *NACADA Journal, 30*(2), 19-34.

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, *22*(11), 1359-1366.

Simpkins, S. D., Davis-Kean, P. E., & Eccles, J. S. (2006). Math and science motivation: A longitudinal examination of the links between choices and beliefs. *Developmental psychology*, *42*(1), 70.

Sutton, S. (1998). Predicting and explaining intentions and behavior: How well are we doing? *Journal of Applied Social Psychology, 28*(15), 1317-1338.

Tai, R. T., Liu, C. Q., Maltese, A. V., & Fan, X. T. (2006). Planning early for careers in science.

Science, 312(5777), 1143–1144.

Unfried, A., Faber, M., Stanhope, D., & Wiebe, E. (2015). The development and validation of a measure of student attitudes toward science, technology, mathematics, and engineering. *Journal of Psychoeducational Assessment, 33*(7), 622-639.

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of statistical software*, *36*(3), 1-48.

Wang, X. (2013). Modeling entrance into STEM fields of study among students beginning at community colleges and four-year institutions. *Research in Higher Education*, *54*(6), 664-692.

Wiebe, E., Unfried, A., Faber, M. (2018). The Relationship of STEM Attitudes and Career Interest. *Eurasia Journal of Mathematics, Science and Technology Education, 14*(10), em1580. https://doi.org/10.29333/ejmste/92286

Zeldin, A. L., & Pajares, F. (2000). Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers. *American Educational Research Journal, 37*(1), 215-246. http://dx.doi.org/10.2307/1163477