Motivational and cognitive challenges faced when debugging block-based code

Grant Title
Collaborative research: Scaffolding preservice early childhood teachers to debug during block-based programming
DUE grants # 1906059 and 1927595 to Brian Belland and ChanMin Kim, respectively

Team Members
Brian R. Belland, Associate Professor, Educational Psychology | bbelland@psu.edu
ChanMin Kim, Associate Professor, Learning Design and Technology and Educational Psychology | cmk604@psu.edu
Anna Zhang, Eunseo Lee, PhD students in Educational Psychology
Emre Dinç, Afaf Baabdullah, PhD students in Learning Design and Technology

The Pennsylvania State University

PennState NSF

SCAN & DOWNLOAD

STUDY 1

SCAN ME

Belland, B. R., Kim, C., Zhang, A. Y., Baabdullah, A. A., & Lee, E. (2021). **Using process and motivation data to predict the quality with which preservice teachers debugged higher and lower complexity programs**. *IEEE Transactions on Education*, *64*(4), 374-382. https://doi.org/10.1109/TE.2021.3059258

Context
Central to preparing ECE teachers to teach computer science is helping them learn to debug. Little is known about how ECE teachers' motivation and debugging process quality contributes to debugging outcome quality.

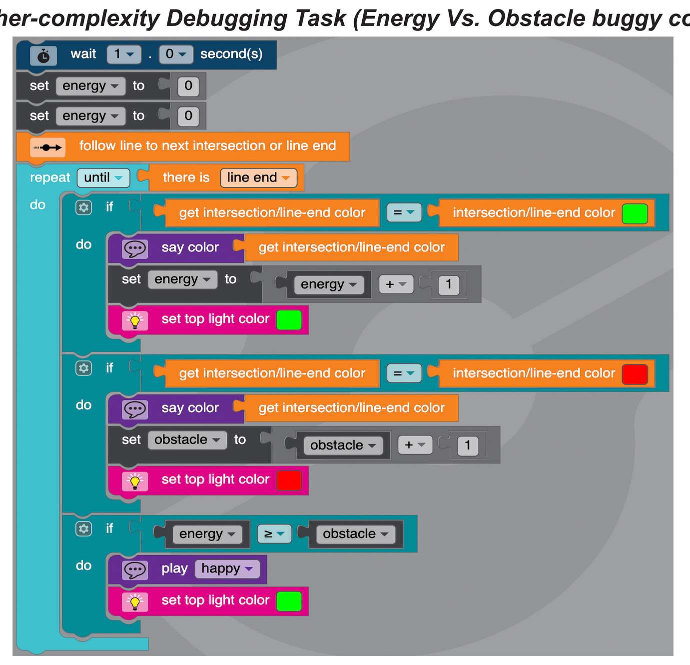
Research Questions
1) How do process and motivation variables predict the quality with which participants debug lower complexity programs?
2) How do process and motivation variables predict the quality with which participants debug higher complexity programs?

Method
Setting and Participants
Three sessions of 2.5-h each of a class on play-based activities in ECE in a large university in the eastern USA
Nineteen students (all female) participated.

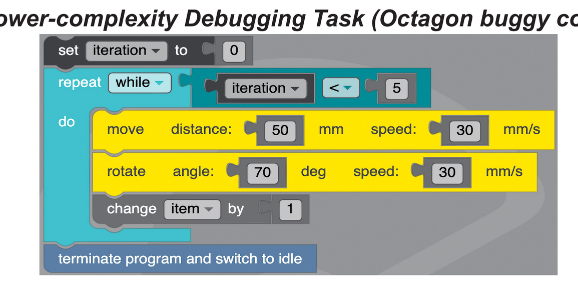
Materials
Ozobot Evo

Scaffolding Informed by
1) a two-step cluster analysis of an expanded coding data set from a metaanalysis
2) a literature review on debugging education
3) a synthesis of Kim et al.'s scaffolding recommendations for debugging in block-based programming

Higher-complexity Debugging Task (Energy Vs. Obstacle buggy code)



Lower-complexity Debugging Task (Octagon buggy code)



Data Collection
Presurvey Covering

STEM interest · Academic emotions · Self-regulated learning · View of coding · Domain identification with mathematics · Domain identification with English · Goal orientation · Perceptions of computer science and engineering

Debugging Process Quality Rubric

Debugging Outcome Quality Rubric

Data Analysis
- Bayesian multiple linear regression with prior distributions of inverse gamma for sigma2 and uniform prior for Beta values were used
- MCMC sampling (11 000 iterations with 1000 burn in) was run in MCMCpack R package to determine the posterior distribution

Results
How do process and motivation variables predict the quality with which participants debug higher complexity programs?

BAYESIAN MULTIPLE REGRESSION MODEL PREDICTING DEBUGGING OUTCOME QUALITY FOR THE HIGHER COMPLEXITY DEBUGGING TASK (HCDOQ)

Coefficients	Estimate (95% CrI)	Naïve SE
(Intercept)	-19.17 (-32.75, -5.18)	$6.96e^{-2}$
Higher-complexity debugging process quality (HCDPQ)	0.45 (0.22, 0.67)	$1.14e^{-3}$
Performance-approach goal orientation (PApGO)	0.41 (-0.05, 0.86)	$2.28e^{-3}$
Word count (HCWC)	-0.005 (-0.014, 0.003)	$4.28e^{-5}$

Note. MCMC iterations = 10,000; 95% CrI = 95% credible interval

How do process and motivation variables predict the quality with which participants debug lower complexity programs?

BAYESIAN MULTIPLE REGRESSION MODEL PREDICTING LCDOQ

Coefficients	Estimate (95% CrI)	Naïve SE
(Intercept)	-2.97 (-19.66, 14.17)	$8.47e^{-2}$
Lower-complexity debugging process quality (LCDPQ)	0.3 (0.1, 0.5)	$9.91e^{-4}$
Performance-avoid goal orientation (PAvGO)	-0.19 (-0.34, 0.49)	$7.53e^{-4}$
Lower-complexity Word count (LCWC)	-0.003 (-0.007, 0.002)	$2.38e^{-5}$
Lower-complexity sentiment analysis (LCSA)	-9.3 (-18.85, 0.34)	$4.82e^{-2}$

Note. MCMC iterations = 10,000; 95% CrI = 95% credible interval

Discussion
Scaffolding works, because debugging process quality was consistently a strong predictor of debugging outcome quality

Writing more or less in response to scaffolding prompts did not make a large difference, which goes against the literature

Sentiment analysis was associated with a large negative Beta for the lower complexity debugging task. This may be because to engage effectively with the scaffolding, one needs to be self-critical, which could be interepreted as negatively valenced

Mastery goal orientations were not significant predictors, which goes against the literature

Implications
Scaffolding should challenge learners to engage in constructive criticism of their work. Greater challenge may cause writing sentiment to become more negatively valenced, but it can, in turn, lead to stronger debugging outcome quality.

STUDY 2

SCAN ME

Belland, B. R., Kim, C., Zhang, A. Y., Lee, E., & Dinç, E. (2022). **Classifying the quality of robotics-enhanced lesson plans using motivation variables, word count, and sentiment analysis of reflections**. *Contemporary Educational Psychology, 69*, 1-11. https://doi.org/10.1016/j.cedpsych.2022.102058

Context
Play has long been key to early childhood education (ECE). Play can involve manipulables, including educational robots. To use robots in ECE, ECE teachers need to learn to program and debug. Keyt to understanding this are the perspectives of teachers as learners and teachers as designers.

Research Questions
1. How can prospective teachers' lesson plan quality be classifiedusing motivation and process variables?
2. How do motivation and process variables predict prospective teachers' lesson plan quality?

Method
Setting and Participants
An early childhood education course on integrating performing and visual arts to enhance communication, inquiry, and engagement in P-5 education, which was offered by a large public university in the United States
Included a field experience component at local preschools
46 prospective ECE teachers participated

Materials
Ozobot Bit

Lesson Design Template
contained the following sections: lesson goals, objectives, considerations (e.g., materials, prior knowledge), and details of class activities

Data Collection
Presurvey
Principal component analysis indicated a ten component solution

STEM emotions · Perceptions of mathematics · Views of coding · Science interest · Computer science and engineering emphasis in STEM career · Perceptions of English · Perceptions of computer science and technology · Mastery goal orientation · Performance goal orientation · Perceptions of computer science

Reflection Cards

Lesson Plans

Data Analysis
Preprocessing
+ Lesson design quality rubric + Sentiment analysis + Word count

Linear Discriminant Analysis
We used the MASS (Ripley et al., 2020) and KlaR (Roever et al., 2020) packages for R to conduct linear discriminant analysis (Lachenbruch & Goldstein, 1979) to predict lesson plan quality

Investigation of Classification Error Rate
We used the linear discriminant functions and support vector machines to predict lesson plan quality category, and compared that to the actual rating

Results
- For front-end analysis quality, the radial kernel resulted in the lowest misclassification rate of 12.821%.For STEM and programmingintegration quality, the radial kernel resulted in the lowest misclassification rate of 7.692%. For instructional activities quality, the polynomial kernel resulted in the lowest misclassification rate of 23.077%.

Strongest Contributors for

Front-end analysis quality
First linear discriminant a) sentiment analysis: coding task reflection, b) STEM emotions, and c) mastery goal orientation
Second linear discriminant a) sentiment analysis: coding task reflection, b) sentiment analysis: field experience reflection, and c) word count: coding task reflection

STEM and programming integration quality
First linear discriminant a) sentiment analysis: coding task reflection, b) sentiment analysis: field experience reflection, and c) STEM emotions
Second linear discriminant a) sentiment analysis: coding task reflection, b) sentiment analysis: field experience reflection, and c) mastery goal orientation

Instructional activity quality
First linear discriminant a) sentiment analysis: coding task reflection, b) sentiment analysis: field experience reflection, and c) word count: field experience reflection
Second linear discriminant a) sentiment analysis: coding task reflection, b) sentiment analysis: field experience reflection, and c) word count: coding task reflection

Discussion
Provides a vision for dynamic assessment of prospective teachers learning to plan lessons
Front-end analysis quality of lesson plans can be classified using
- The extent to which prospective teachers display a mastery goal orientation
- The amount they write when reflecting on coding tasks
- The sentiment reflected in their re flections on coding tasks and field experience teaching

STEM and coding integration quality of lesson plans can be classified using
- The amount prospective teachers write when reflecting on coding tasks
- The sentiment reflected in their reflections on coding tasks and field experience teaching

Instructional activities quality of lesson plans can be classified using
- The amount prospective teachers write when reflecting on field experience teaching
- The sentiment reflected in their reflections on coding tasks and field experience teaching

Implications
Implications for Women in Computer Science
Helping early childhood prospective teachers to better design robot and coding enhanced lesson plans
- Increases confidence in teaching computer science concepts and skills among female prospective teachers
- Increases the number of adequate computer science role models for female students

Implications for Prospective Teachers as Learners and Designers
Discriminant functions can
- Indicate which prospective teachers are on track to produce a low-quality lesson plan, and thus are in need of additional support

Study 3

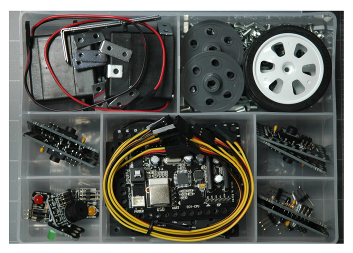
Belland, B. R., Kim, C., Zhang, A. Y., Lee, E., & Dinç, E. (In-progress). **Predicting early childhood teacher candidates' lesson plan quality using generalized estimating equations**.

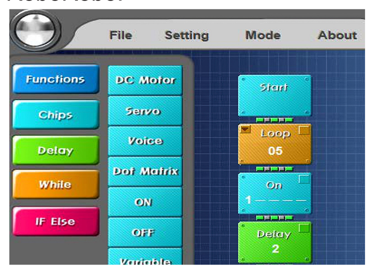
Context
To prepare ECE teachers to teach with robots, there is a need to help them learn about coding and its use with robots (Kay et al., 2014). But there is also a clear need to help them learn how to integrate robots and coding into their classrooms. To do so requires that they be able to plan flexible and adaptable lessons effectively and efficiently (Parsons et al., 2018)
.

Research Question
How can preservice teachers' lesson plan quality be predicted using collaboration status, motivation variables, and academic standing?

Method
Setting and Participants
ECE teacher education undergraduate courses offered in six different semesters between spring 2014 and spring 2017 at a large public university in the southeastern United States.
91 participants submitted a lesson plan.

Materials
RoboRobo - modules (e.g., body, arms, wheels) that needed to be assembled to create a robot
Rogic - a block-based coding platform for use with RoboRobo.





Lesson Design Template
contained the following sections: subjects (e.g., science, math), grade level, objectives, standards, considerations (e.g., materials, prior knowledge), and details of class activities

Study 3 continues ...

Data Collection
Presurvey
Self-reported credit hours completed and semesters completed as predictor variables.
Principal component analysis indicated a 6 component solution:

Mathematics interest · Perceived value of STEM · STEM self-regulation · Science and technology interest · Engineering and STEM career interests · Achievement emotions in STEM

Individual vs. collaborative lesson planning

Lesson Plans

Data Analysis
Preprocessing
Lesson design quality rubric
Dummy coding: Six studies were dummy-coded from 1 to 6. Collaborative and individual work to create lesson plan were dummy-coded 1 and 2, respectively. 54 (59.3%) were produced through collaborative work and 37 (40.7%) were written individually.

Generalized Estimating Equations
We used a Generalized Estimating Equations (GEE) approach (Halekoh et al., 2006) to model lesson plan quality.

Results
Strongest Predictors of Lesson Plan Quality for

Full score
- Achievement emotions in STEM is a positive predictor
- Mathematics interest is a negative predictor

Front-end analysis
- Achievement emotions in STEM is a positive predictor
- Mathematics interest is a negative predictor
- STEM self-regulation is a negative predictor

STEM & Robotics integration
- None of the predictors were significant

Instructional activities design
- Science and technology interest is a positive predictor
- When the students did **not** collaborate on their lesson plans, the teaching and learning activities design score was higher.

Discussion
- Participants perceived that they were more in control of engaging with STEM and the associated outcome, they likely experienced anticipated joy (Pekrun & Linnenbrink-Garcia, 2012), and this in turn predicted a higher overall lesson plan score and a higher front-end analysis score.
- It is puzzling why ***mathematics interest*** was a negative predictor of full lesson plan score and front-end analysis score. Participants with high mathematics interest may have seen a lesson using robots as being at odds with or taking time away from mathematics instruction.
- It is also puzzling why ***STEM self-regulation*** was a negative predictor of front-end analysis quality score. It is possible that students who were high in STEM self-regulation were very interested in integrating STEM and robotics within the lesson, and/or carefully detailing the teaching and learning activities, and as a result, they spent less time and effort at doing and documenting front-end analysis.
- ***Engaging in lesson planning alone*** led to better teaching and learning activities design than planning lessons in pairs. This may be because the design of the robot was always done in pairs, and the major direction of the lesson plan may have already been largely decided during the design of the robot.

Implications
- When teaching preservice teachers to create lesson plans incorporating robots and coding, teacher educators should

Controllability	highlight controllability of engaging with robots and the likelihood of positive results
Interest	work to enhance science and technology interest
Individual work	allow for some individual work in lesson planning

STUDY 4

Belland, B. R., Zhang, A. Y., Lee, E., & Kim, C. (in-progress). **Characterizing the most effective scaffolding approaches in engineering and technology education: A clustering approach**.

Context
Computer science education is most effective when students learn while engaging in authentic practice such as addressing ill-structured problems. But it is not enough to simply give such problems to students. Rather, one needs to provide scaffolding that can address students' learning and performance needs. Thus, it is important to consider what combination of scaffolding features are most effective for which students in which conditions.

Research Question
What combinations of scaffolding characteristics, contexts of use, and assessment levels lead to medium and large effect sizes among college- and graduate-level engineering and technology learners?

Method
Data Source
1) A data set from a metaanalysis study on the effects of computer-based scaffolding on students' cognitive learning outcomes in STEM education
2) Hedge's g effect size was transformed into large, medium, small, and no effect categories
3) 1,726 cases were used whose results showed at least medium effect size in the technology or engineering disciplines

Variables
Nine variables were used which included scaffolding characteristics, study characteristics, student characteristics, and assessment characteristics and their sub-categories.

Attributes	Sub-categories
Scaffolding Characteristics	
Scaffolding Intervention	Conceptual, Metacognitive, Motivation, Strategic
Scaffolding Intended Outcomes	Enhance Motivation, Higher-order Thinking Skills, Knowledge Integration
Scaffolding Strategy	Specific, Generic
Scaffolding Change	Fading, Adding, Fading/Adding, No Change
Scaffolding Schedule	Performance-adapted, Fixed Time Interval, Self-selected, No Schedule
Context	
Education Level	College, Graduate
Context of Use	Case-based Learning, Inquiry-based Learning, Learning by Design, Modeling/Visualization Problem-based Learning, Project-based Learning, Problem-solving
Assessment	
Assessment Level	Principles, Concept, Application
Outcome	
Effect Size	Large, Medium

Analysis
A two-step cluster analysis with hierarchical and partitioning clustering
Cluster solution stability was assessed by checking replicability across different samples from the same dataset.

Results
Model Fit
The silhouette measure of cohesion and separation was 0.62, which indicates a good fit.

Number of Clusters
Two R packages were used to validate the number of clusters and cluster outcomes: the 'cluster' package was used to create a dissimilarity matrix and 'klaR' was used to run the k-modes clustering algorithm. The elbow method indicated that the optimal number of clusters was 8.

Profile of Cluster Outputs

Cluster 4: Large effect size problem solving with principles assessment
Cluster 6: Large effect size problem solving with concept assessment
Cluster 8: Medium effect size problem solving with concept assessment

Clusters 4, 6, and 8 have similar characteristics with the only difference being the composition of the assessment level and effect size variables

Cluster 4: Large effect size problem solving with principles assessment
Cluster 6: Large effect size problem solving with concept assessment
Cluster 8: Medium effect size problem solving with concept assessment

increasing scaffolding intensity (i.e., adding support) has at least medium size effects on improving their higher order thinking skills in problem-solving contexts.

This combination might be ***most effective*** in terms of learning outcomes if the assessment is made on the knowledge of relationships between facts instead of declarative knowledge.

Study 4 continues ...

Cluster 1: Large effect size modelling and visualization with no scaffolding change
Cluster 3: Large effect size inquiry-based learning
Cluster 5: Large effect size project-based learning

support and guidance on what to consider in learning or how to structure content knowledge (i.e., conceptual scaffolding intervention) can be ***effective for higher order thinking skills without any customization*** (i.e., no scaffolding change) in the various contexts.

Cluster 7: Large effect size problem solving with conceptual scaffolding

for ***the same intended outcome of scaffolding*** (i.e., higher order skills), a similar effect size can be achieved through
- targeting students' learning conceptual knowledge and making some variations
- giving flexibility to scaffolding change (i.e., none: 45%, fading and adding: 48%) and its schedule (i.e., none: 45%, self-selected: 38%)
- across more diverse contexts (i.e., problem-solving, case-based learning, and modeling/visualization context)

Cluster 2: Medium effect size modeling and visualization with fading & adding
- scaffolding interventions can be designed in *different ways* (e.g., none: 43%, fading: 14%, fading & adding: 43% in scaffolding change)
- while remaining ***highly effective*** for both knowledge integration and higher-order thinking skills

Discussion
- A **combination of fading and adding scaffolds** can be most effectively used when they deliver context-specific supports targeting students' conceptual learning on the basis of their performance level. This scaffolding customization can be effectively applied to enhance either higher-order thinking skills or knowledge integration.
- **Adding scaffolding** is particularly useful when the scaffolds provide context-specific procedural guidance from a problem solving strategy standpoint (i.e., strategic scaffolds) within the problem-solving context. Further, adding scaffolding on the basis of student performance is most effective.
- Compared to the finding that adding scaffolds is most useful when scaffolding is designed to provide process-related support, **scaffolds focusing on content-related support** (i.e., conceptual scaffolds), and which are design to enhance higher-order thinking skills, do **not** need to be customized to lead to medium or high effect sizes.
- Contrary to the suggestions of much scaffolding literature, **utilizing fading by itself** is little associated with medium or large effect sizes in college and graduate-level technology and engineering education. We found only two studies in which fading scaffolds was used effectively in a very limited circumstance.

Implications

Effectiveness	Scaffolding is most effective when it is (a) designed to ***enhance higher order thinking skills***, and (b) is either ***not customized*** or is ***customized on the basis of performance***
Large effect size	- modeling and visualization with no scaffolding change, - inquiry-based learning, - problem solving with principles-level assessment, - project-based learning, - problem solving with concept-level assessment, and - problem solving with conceptual scaffolding
CS and Engineering	scaffolding (a) is a highly flexible and powerful tool within computer science and engineering education at the college and graduate levels and (b) can be packaged in a variety of ways that can lead to large effect sizes

STUDY 5

Belland, B. R., Kim, C., Zhang, A. Y., Lee, E. (In-progress). **A generalized estimating equations approach to investigate predictors of teacher candidates' views of coding**.

Context
Field experience in which teacher candidates need to integrate technology is one of the best predictors of future technology integration. Preparing ECE teachers to teach computer science involves not only helping them learn the skills of coding and debugging, but also gain a belief that coding is important to ECE curricula. A key predictor of teachers' teaching quality and pedagogical choices is their motivation, which can be thought of from many perspectives, including domain identification, interest, goal orientation, and academic emotions.

Research Questions
- How do ECE teacher candidates' views of coding change as a result of learning to use coding in teaching?
- How can their views of coding be predicted using study, time point (pre-survey versus post-survey), prior programming knowledge and experience, ten latent survey factors, and the inclusion of lesson design/field experience?

Method
Participants and Setting
A total of 199 participants from five different preservice, early childhood education classes from spring 2018 to spring 2020 in two large public universities in the United States. (Female: 96%, n = 191; Male: 4%, n = 8). Most participants majored in Education (98.5%, n = 196; other majors: 1.5%, n=3)

Measures: Presurveys and Post-surveys
Principal component analysis of the pre-surveys indicated a ten-component solution:

Perceptions of mathematics · Perceptions of English · Self-determination in STEM + Computer science · Computer science and engineering in STEM career · Identification with computer science and engineering · Perception of computer science and technology · Mastery goal orientation · Science interest · Performance goal orientation · Perceived value of coding

Measures: Open-ended items
1) In your view, what is coding? What is its purpose?
2) In your view, how can (or cannot) coding be integrated in preschool classrooms?
3) In your view, how does coding relate to disciplines and fields other than computer science? Please provide an example.

Data Analysis
- The five studies were coded from study 1 to study 5.
- Having field experience or not and the time point were both dummy-coded.
- 199 participants' responses to the open-ended items on the pre- and post-survey were evaluated with a rubric.
- We used geepack package in R to conduct generalized estimating equation analysis

Results

Overview
- Overall views of coding scores increase significantly from pretest to posttest for all five study groups
- However, *study* was not a significant predictor for an increase in open-ended response scoring

Overall Views of Coding Scores
- Time variable was a positive predictor
- Robot programming experience was a positive predictor
- Perceived value of coding was a positive predictor
- Programming knowledge was a negative predictor

Views of Integration of Coding in Preschool
- Time variable was a positive predictor
- Perception of mathematics was a negative predictor

Views of Relation of Coding to Non-CS Disciplines
- Time variable was a positive predictor
- Perceived value of coding was a positive predictor

Discussion and Implications

Time	Pre-survey vs. post-survey is a significant predictor for all models
Novice programmers	The use of educational robotics and coding was more ***beneficial for novice programmers*** than participants who already had some level of knowledge
Perception of Mathematics	Students with a more positive perception of mathematics are more likely to have higher satisfaction with the way they learned mathematics and ***prefer teaching using traditional procedures*** than integrating new teaching tools or approaches.
Broad application	With their ***positive perception of the value of coding***, the participants seemed capable of detecting and recognizing the broad application of coding in other subject areas other than CS.

Poster presented at the 2022 NSF IUSE summit, Washington, DC