

SUBGOAL LEVEL FEEDBACK BENEFITS NOVEL PROBLEM SOLVING

A Dissertation
Presented to
The Academic Faculty

by

Laura May Schaeffer

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Psychology

Georgia Institute of Technology
December, 2021

COPYRIGHT © 2021 BY LAURA SCHAEFFER

SUBGOAL LEVEL FEEDBACK BENEFITS NOVEL PROBLEM SOLVING

Approved by:

Dr. Richard Catrambone, Advisor
School of Psychology
Georgia Institute of Technology

Dr. Jamie Gorman
School of Psychology
Georgia Institute of Technology

Dr. Bruce Walker
School of Psychology
Georgia Institute of Technology

Dr. Wendy C. Newstetter
College of Engineering
Georgia Institute of Technology

Dr. Anne McLaughlin
School of Psychology
North Carolina State University

Date Approved: [December 9, 2021]

ACKNOWLEDGEMENTS

Thank you to my committee members and especially my advisor, Richard Catrambone, for their invaluable advice and input. Thank you to Lauren Margulieux, whose work inspired the current study. I would also like to thank the Problem Solving and Educational Technology lab for their instrumental help, especially Nidhi Pai, Typhanie Hall, Zach Tidler, Adam Snoll, and Sara Ferez. This project would not have been possible without them.

Finally, I would also like to thank my family and friends for all of their support and encouragement, and a special thanks to Nathan Sharp for not divorcing me through this endeavor.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
SUMMARY	vii
CHAPTER 1. Introduction	1
1.1 Importance of Feedback	1
1.2 Feedback Best Practices	2
1.2.1 Feedback should be tailored to the task, not to the learner	2
1.2.2 Elaborated feedback is more effective than less informative types of feedback	2
1.2.3 Feedback should be presented to the learner in manageable pieces	3
1.2.4 Feedback should be specific and clear	3
1.2.5 Feedback should be as simple as possible while still including necessary information	3
1.2.6 Feedback should clarify the relationship between a learner's performance and their goals	4
1.2.7 Feedback should be unbiased	4
1.2.8 Feedback should encourage a goal orientation focused on learning instead of performance	5
1.2.9 Feedback should be provided only after the learner has attempted the task	5
1.2.10 Timing of the feedback should correspond with desired outcomes	5
1.2.11 Difficult or complex tasks should include immediate feedback	6
1.2.12 Easy or simple tasks should use delayed feedback	6
1.2.13 Feedback should avoid drawing attention to the "self"	6
1.2.14 Feedback should not interrupt the learner	7
1.2.15 Feedback should avoid using a progressive hint format that consistently terminates with the correct answer	7
1.3 Feedback Granularity	8
1.4 The Benefits of Learning Subgoals	9
1.4.1 Self-Explanation	9
1.4.2 Structural Recognition	10
1.4.3 Working Memory	11
1.5 Significance of the Problem	12
1.6 Problem Statement	13
1.7 Research Question and Hypotheses	13
1.7.1 Hypotheses	14
CHAPTER 2. Method	16
2.1 Participants	16
2.2 Experiment 1 Materials	16
2.2.1 Feedback	17
2.2.2 Assessments	19
2.3 Experiment One Procedure	21
2.3.1 Pre-test	21

2.3.2	Training	22
2.3.3	Assessments	22
2.4	Experiment Two	22
CHAPTER 3.	Results	24
3.1	Demographics	24
3.2	Pre-Test	26
3.3	Problem Solving Performance	27
3.3.1	Performance on Tasks with Feedback	27
3.3.2	Performance on Tasks without Feedback	30
3.4	Cognitive Load Assessment	33
3.5	Explanation Task	35
3.6	Other Measures	37
3.6.1	Attempted Assessments	37
CHAPTER 4.	Discussion	38
4.1	Conclusions	38
4.2	Limitations and Future Research	39
APPENDIX A.	EXCERPT OF INSTRUCTIONS	45
APPENDIX B.	ASSESSMENT ONE, STEP LEVEL FEEDBACK	48
APPENDIX C.	ASSESSMENT ONE, SUBGOAL FEEDBACK	49
APPENDIX D.	STEP FEEDBACK EXAMPLE FOR ASSESSMENT 2	51
APPENDIX E.	SUBGOAL FEEDBACK EXAMPLE FOR ASSESSMENT 2	53
APPENDIX F.	ASSESSMENT TWO	55
APPENDIX G.	ASSESSMENT THREE – NO FEEDBACK	56
APPENDIX H.	ASSESSMENT FOUR – EXPLANATION TASK	57
REFERENCES		59

LIST OF TABLES

Table 1	-- Demographic Averages for Lab Participants and Their Correlation with Problem Solving Performance in Experiment 1.	25
Table 2	-- Distribution of Selected Demographics Between Conditions	25
Table 3	-- Demographic Averages for Remote Participants and Their Correlation with Problem Solving Performance in Experiment 2.	26
Table 4	-- T-tests Comparing Groups for Correct Steps for the Problem Solving Tasks with Feedback	28
Table 5	-- T-tests Comparing Groups for Time on Task for the Problem Solving Tasks with Feedback	29
Table 6	-- Distribution of Time on Problem Solving Assessment and Correlation with Correct Steps Taken for the Problem Solving Assessment for Experiment One (Lab) and Two (Remote)	33
Table 7	-- T-tests for Cognitive Load Assessment	34

SUMMARY

The present study combined subgoal learning and feedback frameworks to further improve problem solving performance and demonstrated that subgoal level feedback resulted in better learning outcomes over step level feedback, first in a lab environment and then in an online-only environment. Feedback is an essential part of learning that tells the learners what they are doing correctly as well as where they can improve. Feedback can be provided at different levels such as the solution level, step level, and sub-step level. Previously, feedback at the step level had shown to be as effective as sub-step level feedback while requiring less time and fewer resources to create. Subgoals are components of a problem solution that transfer across problems in a given domain. Subgoal level feedback has several advantages over step level feedback that cause it to be more effective for learners. Subgoals help learners identify the structure of the problem, chunk steps together (thereby reducing extraneous load), and encourage self-explanations. The present study combined subgoal learning and feedback frameworks to further improve problem solving performance. Learners who received subgoal level feedback correctly completed more steps of novel problem solving tasks and were better able to explain problem solving solutions than learners who received step level feedback. The results suggest that subgoal level feedback leads to better transfer on novel tasks because the subgoal framework helps learners better understand and apply general procedures.

CHAPTER 1. INTRODUCTION

1.1 Importance of Feedback

Feedback plays a critical role in learning and has been argued to be one of the most significant influences on learner achievement (Hattie & Timperley, 2007). In the context of education, feedback is defined as external information provided to support learning (e.g., McLaughlin et al., 2006; Salmoni et al., 1984; Shute, 2008). The purpose of feedback is to bridge the gap between a given learning goal and the learner's current understanding (Hattie & Timperley, 2007). Feedback is meant to be an aid to help the learner improve their performance until the learner can produce internal feedback that accurately assesses their own progress (McLaughlin et al., 2006). Increasingly, learning occurs online without an instructor available to guide the student. Here, feedback is necessary to correct the student while they learn the necessary skills to direct themselves.

Intelligent tutoring systems (ITS) are increasingly used in all levels of education (Koedinger et al., 2013). Intelligent tutoring systems are a form of educational technology where the ITS tailors lessons based on the learner's performance (VanLehn, 2011). Feedback provided by an intelligent tutoring system presents its own distinctive challenges as well as the opportunity for unique benefits. Considerations that must be made when designing feedback from an ITS that are not necessarily present in traditional learning environments include unique factors affecting the learner's cognitive load and self-assessment. The following best practices apply to feedback provided by tutors or educational technology, such as an ITS.

1.2 Feedback Best Practices

Shute (2008) makes the following recommendations for designing feedback based on the literature:

1.2.1 Feedback should be tailored to the task, not to the learner

Feedback should focus on specific aspects of the learner's performance on a task instead of tailoring the feedback to characteristics of the learner (e.g., Butler, 1987; Corbett & Anderson, 2001; Kluger & DeNisi, 1996; Narciss & Huth, 2004; Shute, 2008). For example, Butler (1987) measured task performance and interest with 200 fifth and sixth grade students who were categorized as either high achievement (average grades were in the top 25% of their class) or low achievement (average grades were in the bottom 25% of their class). For all students, a single sentence of feedback that included a reinforcing statement and a goal-setting component for the task was more effective than praise or grades.

1.2.2 Elaborated feedback is more effective than less informative types of feedback

Elaborated feedback has a wide range of meanings found within the literature. Elaborated feedback takes many forms such as hints, strategy information, explanation of the correct answer, or additional study materials (Shute, 2008). Feedback should include enough information to make it clear to the learner what is incorrect, why it is incorrect, and how it should be fixed. Simply pointing out the correctness of an answer does not have as much of an effect as feedback that includes information on how to improve (e.g., Bangert-Drowns et al., 1991; Gilman, 1969; Mason & Bruning, 2001; Narciss & Huth,

2004; Shute, 2008).

1.2.3 Feedback should be presented to the learner in manageable pieces

Too much information can cause the learner to become overwhelmed, and lead to the learner not using the feedback (Bransford et al., 2000; Sweller et al., 1998). Excessive feedback overloads the available cognitive resources needed to process the information (e.g., Mayer & Moreno, 2002; Phye & Bender, 1989). Providing the information in manageable pieces allows the learners the opportunity and necessary information to address their errors.

1.2.4 Feedback should be specific and clear

Learners can be frustrated by feedback that does not have obvious connections to their performance and goals or is otherwise unclear (e.g., Moreno, 2004; Hoska, 1993; Song & Keller, 2001; Williams, 1997). Therefore, feedback should be clearly linked to the learner's goals and performance (Hoska, 1993; Song & Keller, 2001).

1.2.5 Feedback should be as simple as possible while still including necessary information

Complex feedback does not necessarily lead to better learning compared to simpler feedback (Kulhavy et al., 1985). Kulhavy et al. (1985) provided college students four types of feedback that differed in complexity. The simplest level of feedback gave the learner just the correct answer. As each level became more complex, it included the lower level of feedback with the addition of more information. The most complex feedback gave the learner the correct answer, explanations of why the incorrect alternative answers

were wrong, and an indication of where the correct answer is found in the passage. Kulhavy et al. (1985) found little benefit from the more complex feedback, and that the simpler feedback resulted in greater learning efficiency (measured as the ratio of feedback study time to posttest score).

1.2.6 Feedback should clarify the relationship between a learner's performance and their goals

Feedback should make clear what the learner is doing well and what needs to be done to reach their goal (e.g., Ashford et al., 2003; Bangert-Drowns et al., 1991). Novices often have trouble assessing how well they are performing a task (e.g., Kornell & Bjork, 2009). Feedback can indicate where a learner is and is not meeting their performance goals, thereby reducing uncertainty about their performance (Ashford, 1986; Ashford et al., 2003). Resolving the gap between a learner's performance and their goal level of performance can drive higher levels of effort from the learner (Locke & Latham, 1990; Song & Keller, 2001). Additionally, uncertainty in regards to performance can distract attention away from the task; reducing uncertainty about their performance may lead to higher motivation and more efficient task strategies (Kanfer & Ackerman, 1989).

1.2.7 Feedback should be unbiased

Feedback that is perceived to be unbiased is more likely to be used by a learner (Kluger & DeNisi, 1991). Learners perceive oral feedback to have more bias than written feedback (Kluger & DeNisi, 1991). Computer-based feedback is generally considered to be less biased than human-delivered feedback (Kluger & DeNisi, 1996). Therefore, written or computer-based feedback is generally recommended over oral or in-person

feedback.

1.2.8 Feedback should encourage a goal orientation focused on learning instead of performance

Goal orientation is a motivator for human behavior (VandeWalle, 1997). People who hold learning-oriented goals focus on acquiring new skills and seek feedback on their past performance to inform their future performance (Dweck & Leggett, 1988). People who are learning-oriented believe ability is a malleable trait. Performance oriented goals focus on obtaining approval from other people and are typically held by people who believe ability is a fixed attribute (Dweck & Leggett, 1988). Goal orientations can be changed from performance oriented to learning oriented by providing feedback that is specific and goal directed (Davis et al., 2005), normalizes making mistakes as a part of the learning process, and emphasizes that effort leads to increased learning and performance (Hoska, 1993; Dweck, 1986).

1.2.9 Feedback should be provided only after the learner has attempted the task

Providing feedback after a learner attempts a task generally results in increased performance and learning measures, but the literature shows inconsistent results when feedback is available to the learner before they have attempted the task (Bangert-Drowns et al., 199; Shute 2008).

1.2.10 Timing of the feedback should correspond with desired outcomes

Immediate feedback often leads to immediate increases in performance (Azevedo & Bernard, 1995; Corbett & Anderson, 1989, 2001; Dihoff et al., 2003; Mason & Bruning,

2001; Phye & Andre, 1989), but delayed feedback often leads to better retention and transfer (e.g., Kulhavy et al., 1985; Schroth, 1992). Therefore, the timing of when feedback is given to the learner is dependent on the goals of the training.

1.2.11 Difficult or complex tasks should include immediate feedback

When a task is difficult for a learner, immediate feedback is better early in the learning process (Clariana, 1990) as the additional aid helps guide students and avoids unproductive frustration (Knoblauch & Brannon, 1981).

1.2.12 Easy or simple tasks should use delayed feedback

When a task is simple or easy for a learner, immediate feedback can be perceived as intrusive or annoying (Clariana, 1990; Corno, & Snow, 1986). Immediate feedback is often necessary for difficult tasks to prevent the learner from floundering or becoming discouraged, but this does not apply to easy tasks. Feedback that is given after a delay is perceived as less intrusive and therefore preferable for simple tasks (Clariana, 1990; Corno & Snow, 1986).

1.2.13 Feedback should avoid drawing attention to the “self”

The content of the feedback also impacts the effectiveness of the feedback on learning and performance (Butler, 1987; McColskey & Leary, 1985). Feedback should focus on how the learner can improve their performance and should not include normative comparisons (Kluger & DeNisi, 1996; Shute, 2008; Wiliam, 2007). Wiliam (2007) found that students who received only comments demonstrated large learning gains but students who received only grades or grades with comments demonstrated

limited, if any, learning gains. Williams (2007) believed this was due to students focusing on the grade instead of focusing on the comments and how to improve. Similarly, praise as well as critical feedback have also both been shown to direct attention away from the task and learning to instead focus on the “self” (e.g., Baumeister et al., 1990; Baron, 1993; Fedor et al., 2001; Kluger & DeNisi, 1996; Butler, 1987). Therefore, it is generally advised to avoid praise, normative comparisons, or otherwise drawing attention to the learner’s “self” during learning (Kluger & DeNisi, 1996; Shute, 2008; Wiliam, 2007).

1.2.14 Feedback should not interrupt the learner

Interruptions often negatively impact performance (e.g., Bailey & Konstan 2006). Interruptions during learning, such as when a student is actively engaged and working to solve a problem on their own, should be avoided as they can disrupt the learner and hinder learning (Corno & Snow, 1986).

1.2.15 Feedback should avoid using a progressive hint format that consistently terminates with the correct answer

If feedback predictably ends with the correct answer, the learner may learn to rely on the feedback as a crutch instead of as a learning aid. Although hints can be helpful, steps should be taken to prevent their abuse, such as by using more specific types of hints, such as prompts or cues (e.g., Aleven & Koedinger, 2000; Shute, Woltz, & Regian, 1989).

Applying these best practices to feedback design is a good starting point, but there are still other feedback variables that have different effects on learning (e.g., McLaughlin et al., 2006; Shute, 2008). One of these variables is granularity. This study examines how two different types of feedback granularity affect problem solving performance.

1.3 Feedback Granularity

VanLehn (2011) defines feedback granularity as “the amount of reasoning required of participants between opportunities to interact” (p. 202). VanLehn (2011) categorizes feedback into three broad classes based on their granularity: answer-based, step-based, and substep-based.

Answer-based feedback refers to feedback given after the student has submitted a final answer to the problem. For example, a student given the algebra problem $3x+2=13$ would receive answer-based feedback after submitting their answer for the x value. Answer based is considered large granularity because the student must engage in a large amount of reasoning (the student’s reasoning, the feedback, and the interactions with the user interface) between the problem and the answer. Step based feedback provides feedback on each of the individual steps necessary to solve the problem. For the previous algebra problem, a student might receive feedback after isolating $3x$ on the left side of the equation. This is smaller-grained feedback than answer-based because a smaller amount of reasoning is required (i.e., carrying out a step) before feedback. Substep based feedback provides feedback on the mental inferences necessary to complete a step. If a student is struggling to take the first step with the algebra equation, substep based feedback might be given in the form of explaining how to isolate a variable. This is fine-grained feedback because very little reasoning is required before the learner receives feedback.

VanLehn’s (2011) meta-analysis shows that effectiveness on learning gains increases as granularity becomes finer grained from answer-based to step based. However, no additional benefit is shown from further decreasing the granularity to the

even finer-grained substep based feedback. Given that more resources are needed to create the substep based feedback, providing feedback at the step level is considered the most efficient way to provide feedback to increase learning.

1.4 The Benefits of Learning Subgoals

A problem solving procedure can be broken down into necessary structural components. These components are referred to as subgoals (Catrambone, 1995). In order to solve a problem, subgoals are often solved along the way as part of the process to solve the main goal. Subgoals are smaller tasks that are necessary to completing the larger task.

For example, a student who is learning how to solve algebra problems might be given the equation $6x-4=4x+10$. The goal of this problem is to solve for x . The first subgoal would be “isolate the variable.” The steps in this subgoal would be, “subtract $4x$ from each side,” and also “add 4 to each side.” The second subgoal would be “simplify terms.” The step under this subgoal would be, “divide both sides of the equation by 2.” These subgoals provide a structure for completing the larger task of solving for x . When solving a new problem, the individual steps need to be modified to suit that particular algebra problem. However, the subgoals remain the same when solving for different problems in this domain. When solving a new algebra problem, such as $8x+18=36-x$, the same subgoals are used.

1.4.1 Self-Explanation

Incorporating subgoal labels in instructions has been shown to be an effective method for subgoal learning (Catrambone, 1995; Catrambone, 1998; Margulieux & Catrambone, 2017). Subgoal labels are function-based descriptions of a group of steps

that highlight their purpose (Catrambone, 1995). Including subgoal labels in novice instructional materials has led to improved problem solving performance and better self-explanations (e.g., Margulieux & Catrambone, 2017; Catrambone, 1995). Self-explanation is a learning strategy where students generate their own explanations of connections and conceptual relationships (Bisra et al., 2018). Self-explanation has several learning benefits, including the improvement of problem solving skill acquisition (Chi et al., 1994.) Subgoal labels have also been shown to reduce the number of incorrect explanations (Renkl & Atkinson, 2002). Self-explanation helps learners understand procedures instead of just memorizing steps (Chi et al., 1994; Margulieux & Catrambone, 2017; Paas & van Gog, 2009). When learners better understand the procedure, they are then able to better transfer the procedure to novel problems (e.g., Catrambone, 1998).

1.4.2 Structural Recognition

Novices often focus on and remember superficial, instead of structural, features of the problem. For example, Chi et al., (1989) found that novice physics students were more likely to remember that a problem involved a ramp rather than that it used Newton's Second Law. Subgoal learning has been shown to be effective for novices because subgoals help students recognize the structure of the problem solving solution which makes it easier for students to apply the structure to novel problems, again leading to better solving of novel problems (Atkinson et al., 2003; Catrambone, 1998; Catrambone & Holyoak, 1990).

1.4.3 Working Memory

Working memory restricts the learner's cognitive capacity to accommodate demands imposed by learning. Because of this, several researchers argue that a central principle in the design of instructional materials should be the consideration of cognitive load (e.g., Mayer & Moreno, 2003; Plass et al., 2003; Sweller, 1994). Extraneous cognitive load is the working memory load resulting from the way information is presented to the learner. Worked examples impose extraneous cognitive load because they contain context-specific information that the learner must process even though it is not relevant to the underlying procedure (Sweller, 2010). Subgoal labels highlight the structural features of the procedure, reducing the learner's focus on these superficial context-specific features, and thus reducing extraneous cognitive load (Renkl & Atkinson, 2002).

This is especially important in a computer-based learning environment because research suggests that computer-based instructional materials result in greater extraneous cognitive load than paper-based materials. For example, scrolling through text on a screen has been shown to be more mentally demanding than turning through pages in a book because moving text creates additional extraneous load (Wästlund, 2007). Scrolling is one example of extraneous workload inherent to screens because the workload does not contribute to understanding the material but instead is just a function of the presentation form (Chandler & Sweller, 1991). Wästlund et al., (2005) also found that screen reading resulted in reports of higher stress and tiredness levels than paper reading. Another study showed that reading an article on the computer resulted in similar comprehension but higher NASA-TLX workload scores compared to reading a paper-based article (Noyes et

al., 2004). Keeping these studies in mind, special design considerations should be used for multimedia instruction to reduce extraneous cognitive load when the material has high intrinsic cognitive load (Mayer & Moreno, 2003; Moreno & Valdez, 2005; Sweller, 2010).

Subgoal learning has been shown to be most effective when used in both worked examples and procedural instructions (Margulieux et al., 2012). Worked examples show a learner how to solve a specific problem (Eiriksdottir & Catrambone, 2011). A worked example has a problem statement and the steps used to arrive at the final solution (Eiriksdottir & Catrambone, 2011). Procedural instructions explain the steps or procedure used to solve problems in a given domain (Eiriksdottir & Catrambone, 2011). Students tend to rely on worked examples more than procedural instructions (Eiriksdottir & Catrambone, 2011; LeFevre & Dixon, 1986), but learners who receive subgoal labels throughout the instructional materials learn better than when subgoal labels are in only the worked example (Margulieux et al., 2012). Subgoal labels in instructional materials improve performance without increasing study time (e.g., Margulieux et al., 2016). Subgoal learning has shown to be an effective tool for learning, but has not been applied to feedback design.

1.5 Significance of the Problem

Providing feedback at the subgoal level could be more effective than at the step level for helping learners understand and apply the problem solving structure to novel problems. This is because subgoals highlight the structure of the problem solving procedure which is difficult for novices to do on their own. Additionally, subgoals

encourage learners to participate in activities that enhance learning, such as self-explanation.

1.6 Problem Statement

The development of instructional materials uses many resources, including the time of the instructor or domain expert. Knowing what level to provide feedback can help focus resources to help learners learn most effectively. This study compared subgoal level feedback and step level feedback on student learning outcomes in computer programming.

Programming is the process of creating instructions for a computer to act upon. Introductory programming courses have low levels of completion and programming is a domain in need of successful learning interventions (Hundhausen et al., 2009). Previous research has shown that subgoal learning and self-explanation facilitate learning programming (e.g., Margulieux & Catrambone, 2017; Pirolli & Recker, 1994; Soloway, 1986), making this an appropriate domain for the subgoal level feedback intervention.

1.7 Research Question and Hypotheses

Research question: Does feedback at the subgoal level lead to improved learning compared to step level feedback? The current literature has shown step level feedback to be the most effective level of feedback regarding both learning and the resources needed to create it. Providing feedback at the subgoal level rather than at the step level could be more effective for helping learners understand and apply the problem solving structure to novel problems. This is because subgoals highlight the structure of the problem solving

procedure, which is difficult for novices to do on their own. Additionally, subgoal learning encourages learners to participate in activities that enhance learning, such as self-explanation.

1.7.1 Hypotheses

1. Step level feedback and subgoal level feedback will have similar performance on near-transfer problem solving tasks.
2. There will be an interaction with transfer level and feedback level. As the need to understand the structure of the test problem increases (i.e., as transfer distance increases), the benefits of subgoal learning increases.
3. Subgoal level feedback will lead to better self-explanation than step level feedback.

1.7.1.1 Near transfer.

These are problems that use the same steps (and therefore subgoals) as the ones in the training materials, but the context-specific details differ. The subgoal level feedback group was not expected to outperform the step level feedback group when solving problems using the same steps but substituting different context-specific details. Previous research has not found differences between subgoal and non-subgoal learning conditions on near transfer (Catrambone, 1998; Margulieux & Catrambone, 2017). Learners are typically able to reproduce a procedure with different superficial details (Catrambone, 1998).

During the instructional period of this study, participants were guided through an example of the subgoal “Set Properties” as they were taught the steps to italicize the font for an app image label of a Magic 8-Ball. For this study, an example of near transfer for the subgoal “Set Properties” was bolding the font of a label. The participant would follow the same procedure for the subgoal “Set Properties” that they had learned and needed to change only the context-specific details of checking the italicize text box to checking the bold text box.

1.7.1.2 Far transfer.

These are problems where the subgoals remained the same but the steps must be adapted to solve the problem. Recognizing the structure of the problem is crucial to solving these problems. The subgoal based feedback group was expected to do better than the step level feedback group. The subgoal based feedback group will be better able to adapt the procedures because they will be better able to identify the problem solution structure.

In this study, a far transfer task for the subgoal “Set Properties” was to change the speed of an animation component. Participants had set properties of other component types, such as changing the text properties of a label, but needed to adapt the procedure to successfully change the animation component’s speed.

CHAPTER 2. METHOD

2.1 Participants

This study focused on feedback for novices. Therefore, participants must not have had prior experience with App Inventor or have had more than one semester of programming experience. Previous studies have found no difference between those who have taken one semester of programming and those who have no programming experience (Margulieux et al., 2012). Participants were recruited through the SONA experimental system at the Georgia Institute of Technology. A total of 141 participants were recruited with 30 participants in each of the four conditions. Twenty-one participants were removed from the study because even after prompting from the ITS, they did not request any feedback and therefore were not exposed to the learning intervention. Eleven of those participants were part of Experiment 1 and ten were in Experiment 2.

2.2 Experiment 1 Materials

The materials used for this study were as follows: demographic questionnaire, pre-test, Bluejeans set-up instructions, “Magic 8-Ball” app instructional materials, tasks with feedback, and assessments (cognitive load assessment, problem solving tasks, and explanation task). The following demographic information was collected because Rountree, Rountree, Robins & Hannah (2004) found these characteristics to predict success in introductory programming: gender, comfort with computers, expected

difficulty of learning programming, high school GPA, college GPA, year in college, and field of study.

MIT's Android App Inventor is a drag-and-drop programming environment that allows users to build fully functional apps for smartphones and tablets. App Inventor is a useful tool for novices because the learner does not need to write any code. Instead, the learner drags components from a menu and pieces them together like puzzle pieces, which is more intuitive for the novice programmer (Hundhausen et al., 2009).

Qualtrics is a flexible web-based survey software that will be used to provide instructions and collect data. Bluejeans is a commonly used remote conferencing software that has screen sharing and chat features which were necessary for this study.

Initial instructional materials were the same for all groups. Subgoal labels were present in the initial instructional materials as the participant was introduced to the App Inventor programming language and given a worked example with step-by-step instructions to create an app of a Magic 8-Ball that presents a fortune when touched (see Appendix A). The procedure's subgoals were identified using the Task Analysis by Problem Solving (TAPS) procedure (Catrambone et al., 2012) that has been used in prior research (e.g., Margulieux & Catrambone, 2016; Margulieux & Catrambone, 2017; Margulieux, Catrambone, & Schaeffer, 2018.)

2.2.1 Feedback

The feedback during the problem solving phase followed Shute's (2008) best practices and was provided at different levels, either step level or subgoal level. For both feedback levels two types of feedback were intended to be given. The first feedback type was "knowledge of result" feedback. This informs the learner simply of whether they

have taken correct actions (Shute, 2008). This was intended to be delivered in two parts. The first line of feedback was phrased as an invitation for the learner to ask for help: “It looks like you are taking steps to [complete the given task]. Do you need help?” For example, if the participant was working on the task where they have been instructed to make the Magic 8-Ball text italic (see Appendix B), the first line of feedback was “It looks like you are taking steps to italicize the ‘Ask the Magic 8-Ball a Question’ text. Do you need help?”

The next line of feedback was dependent on the learner’s actions. If the actions taken were incorrect, the next feedback was intended to be: “Hmm, the current app doesn’t [complete the given task]. Please try again.” However, this situation did not occur during the study. Instead, the participant would explicitly answer “yes” or ask the ITS for specific help.

Alternatively, if the participant’s actions were correct, the feedback was either “You have made progress towards [completing the given task.] Keep going!” if the participant had taken correct steps, or “You have [completed given task]. Great job!” if the participant correctly finished the task. The knowledge of result feedback was the same for each feedback level for a given task.

If the participant again asked for feedback or made incorrect actions, the next feedback was formative feedback that instructs a learner on how to improve (Shute, 2008). First, feedback was given that addressed the area of the task where the participant needed guidance. For the step level feedback, this was identifying the step. For the subgoal group, the feedback identified the subgoal that needed to be completed (see

Appendix C). If the learner again made an incorrect action or asked for help, they were given an example of when they completed a similar step or subgoal earlier in the instructional phase. For example, if the participant was working on the task to italicize the ‘Ask the Magic 8 Ball a Question’ command, the feedback would be as follows:

Step feedback 1: “You need to change the button's text property.”

Step feedback 2: “Here is an example of when you changed a text property from earlier,”(see Appendix E).

Subgoal feedback 1: “You need to ‘Set Properties’.”

Subgoal feedback 2: “Here is an example of when you ‘Set Properties’ earlier,”
(see Appendix F).

The feedback script was cycled through as appropriate based on the participant’s actions in App Inventor or when they asked for help through the Bluejeans chat feature. During the study, the step level feedback group never asked or needed additional feedback for the same step. Only the subgoal group ever asked for additional help and received the follow up example.

2.2.2 Assessments

The first assessment evaluated problem solving performance while receiving either step or subgoal level feedback (see Appendices B and C). This assessment consisted of four tasks including modifying the Magic 8-Ball created during the initial instructional phase as well as creating a new app that moves around the screen, changes to random colors, and changes direction when touched. These tasks included nine subgoals.

The second assessment was a cognitive load survey (see Appendix F). The cognitive load survey was adapted from the materials created by Leppink et al. (2013) and has been used in several programming studies (e.g. Schaeffer et al., 2016; Margulieux & Catrambone, 2016). The original survey was used to measure cognitive load of materials presented in a statistics class, but Leppink et al. (2013) developed the survey with the intention to be used in other domains with modifications. The modifications for the present study included changing the word “formulas” to “procedures” in questions 2 and 9. In question 8, the word “statistics” was changed to “programming.” The 10-item questionnaire measures intrinsic load, which is demand on working memory due to the complexity of the task (questions 1, 2, and 3), extraneous load, which is unnecessary demand on working memory due to the design of the materials (questions 4, 5, and 6) and germane load, which is the demand on working memory necessary to learning the material (questions 7, 8, 9, and 10).

The third assessment consisted of two problem solving tasks (see Appendix G) where participants were asked to make a modification to their Magic 8-Ball app and then create a new app. This assessment contained tasks that required near and far transfer of the skills they learned during the instructional period. This assessment measured participants’ problem solving performance on novel tasks using App Inventor.

The fourth and final assessment was the explanation task (see Appendix H). Correct solutions to three of the first assessment’s problem solving tasks were provided to the participants. Participants were then asked to group steps of the problem solving task solution and to provide a meaningful label or explanation for their groups. This

assessment measured how well participants can group steps based on structural similarity and how well they can explain the solutions.

2.3 Experiment One Procedure

Participants were randomly assigned to either the step level or subgoal level feedback groups. To begin, the participant was seated at a lab computer with a Word document open that instructed them to open a Bluejeans link and to read the chat box for further instructions. When the participant joined the Bluejeans session, the experimenter, disguised as the ITS, gave the participant the link to the Qualtrics survey. The study's directions and assessments were all delivered through Qualtrics. Participants opened the Qualtrics link and were asked screening and demographic questions. Then, they were given step-by-step instructions on how to screen share with Bluejeans, the video conferencing software. They were instructed on how to login into App Inventor. They were also instructed to place the Bluejeans chat box on the right side of the screen so they were able to toggle between the Qualtrics instructions and App Inventor windows while having the Bluejeans chat visible at all times. Participants were told that the screen-sharing and chat feature were used by an intelligent tutoring system to provide them feedback. In reality, the experimenter gave them feedback as described previously.

2.3.1 Pre-test

The next screen in Qualtrics asked the participant to create a "Cat App" app that meows when tapped. They were given the cat image and sound clip needed to make the app. The purpose of this task was to assess the participant's current knowledge of app creation.

2.3.2 *Training*

All participants were given a short expository text explaining the procedure as well as a guided worked example with step-by-step instructions on how to create the Magic 8 Ball app.

2.3.2.1 Problem Solving tasks with feedback.

After creating the Magic 8-Ball app, participants were asked to make some modifications to the Magic 8-Ball app, as well as to create another app. Participants were able to ask for help using the chat feature. The experimenter, disguised as the ITS, provided feedback using the previously described system.

2.3.3 *Assessments*

Participants were then asked to complete the cognitive load survey. Next, they were asked to complete several problem solving tasks without feedback that included near and far transfer tasks. Participants were asked to do tasks such as make modifications to the “Magic 8-Ball” app as well as make the Cat app, the same task they were given in the pre-test. Finally, participants were asked to complete the Explanation Task.

2.4 **Experiment Two**

The second experiment determined if the same results from experiment one are replicated when learners are not in a lab environment. Online learning continues to increase in popularity. These learners need support from their educational technology because a live instructor is not present. This study examined the efficacy of subgoal

feedback over step-based feedback in an environment that more closely aligns to real world learning scenarios.

The second experiment followed the same procedure as Experiment 1 but the participants were in an environment of their choice, although they were asked to choose a quiet and calm environment as though they were taking an online exam. When participants signed up for the study, they included their email address. At the time of the study, the participants were emailed a Bluejeans link where the experimenter, again disguised as the ITS, would then send them the Quatrics link at the meeting time.

CHAPTER 3. RESULTS

Experiment one examined how step level feedback and subgoal level feedback effect learning in a controlled lab environment. Experiment two examined this same effect but in an online learning environment. Results were hypothesized to be similar in both experiments, and are therefore discussed concurrently in the following sections.

3.1 Demographics

A variety of demographic factors were collected as possible performance predictors. Only one factor was correlated with performance. For experiment one, expected difficulty was negatively correlated with performance on the problem solving tasks without feedback (see Table 1). This predictor was not expected to confound the analyses of the performance metrics because there are no differences between the experimental conditions on these predictors (see Table 2). No significant correlations were present in experiment two (see Table 3).

Table 1 -- Demographic Averages for Lab Participants and Their Correlation with Problem Solving Performance in Experiment 1.

	Averages		Correlation	
	<i>M</i>	<i>SD</i>	<i>r</i>	<i>p</i>
Gender	47% male	-	.24	.07
Year in College	1.47	.9	-.09	.23
High School GPA	4.15	.34	-.09	.48
College GPA	3.72	.27	-.23	.11
Comfort with Computers (out of 7)	4.18	1.9	.12	.09
Expected Difficulty (out of 7)	4.67	1.7	-.29	.03
Previous Courses CS 66% taken 1 course		-	-.11	.412

Table 2 -- Distribution of Selected Demographics Between Conditions

Condition	Expected Difficulty			
	<i>M</i>	<i>SD</i>	<i>r</i>	<i>p</i>
Step	4.83	1.8	-.21	.27
Subgoal	4.50	1.8	-.32	.08

Table 3 -- Demographic Averages for Remote Participants and Their Correlation with Problem Solving Performance in Experiment 2.

	Averages		Correlation	
	<i>M</i>	<i>SD</i>	<i>r</i>	<i>p</i>
Gender	58% male	-	.02	.91
High School GPA	4.06	.39	-.07	.61
Year in College	1.45	1.3	.03	.69
College GPA	3.65	.31	.5	.28
Comfort with Computers (out of 7)	4.20	2.8	.12	.09
Expected Difficulty (out of 7)	5.65	2.4	.10	.44
Previous CS Courses	45% taken 1 course	-	-.25	.06

3.2 Pre-Test

Since participants have limited experience with programming, it was expected that most participants would not be able to complete this task. Two raters were used to score the assessments. ICC(A) for this assessment was .98. Participants earned one point for each correct step they took towards the problem solution. Participants could earn up to 7 points. Ninety-five percent of the participants did not receive any points. The remaining six participants earned two points and were evenly distributed between the two experiments.

3.3 Problem Solving Performance

Scoring followed the method developed by Margulieux et al., (2012) which has been shown to have high interrater reliability and high statistical power due to partial scoring methods discussed later. Two raters were used to score the assessments. ICC(A) for this assessment was .96. Participants earned one point for each correct step they took towards the problem solution. This scoring scheme provides more sensitivity than scoring only the solution.

3.3.1 *Performance on Tasks with Feedback*

The maximum score participants could earn for completing all four problem solving tasks correctly was 15. Independent samples t-test were used to compare the two groups.

In experiment one, lab participants who received step feedback group ($M = 10.6$, $SD = 4.2$) took more correct steps than the subgoal feedback group ($M = 8.1$, $SD = 4.0$), $t(58) = 2.35$, $p = .02$, $d = 0.61$ (see Table 4). Similarly in experiment two, scores on the problem solving tasks with feedback were higher for the step feedback group ($M = 8.3$, $SD = 5.2$) than for the subgoal feedback group ($M = 5.4$, $SD = 3.3$), $t(50) = 2.90$, $p = .01$, $d = 0.67$. Levene's test indicated unequal variances ($F = 3.56$, $p = .04$), so degrees of freedom were adjusted from 58 to 50. It was hypothesized that participants who received step level feedback might do better on this assessment because they are receiving the feedback with context-specific information to improve their performance.

Table 4 -- T-tests Comparing Groups for Correct Steps for the Problem Solving Tasks with Feedback

Condition	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t</i>	Std. error	<i>p</i>
Lab Step level	30	10.6	4.2	2.35	1.06	.02
Lab Subgoal level	30	8.1	4.0			
Remote Step level	30	8.3	5.2	2.90	1.12	.01
Remote Subgoal level	30	5.4	3.3			

Participants asked the ITS for feedback, so it could have been possible that one condition asked for more feedback than the other. However, this was not found to be the case. There was no significant difference between the amount of feedback received in the lab step feedback group ($M = 3.7$, $SD = 1.4$) than for the subgoal feedback group ($M = 3.7$, $SD = 1.4$), $t(58) = .09$, $p = .93$ (see Table 5). There was no significant difference between the amount of feedback received in the remote step feedback group ($M = 3.5$, $SD = 1.4$) than for the subgoal feedback group ($M = 3.5$, $SD = 1.6$), $t(58) = .18$, $p = .86$.

Table 5 -- T-tests Comparing Groups for Time on Task for the Problem Solving Tasks with Feedback

Condition	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t</i>	Std. error	<i>p</i>
Lab Step level	30	17.7	9.2	-1.44	1.88	.16
Lab Subgoal level	30	20.4	4.6			
Remote Step level	30	21.5	4.2	1.50	1.29	.14
Remote Subgoal level	30	19.5	5.7			

Participants who received subgoal level feedback might have taken more time on this portion as they might needed more time to determine the context-specific steps that are appropriate to complete the tasks. For experiment one, lab participants who received subgoal feedback ($M = 20.4$, $SD = 4.6$) did not spend significantly more time on the feedback tasks than those who received step feedback ($M = 17.7$, $SD = 9.2$), $t(58) = 1.44$, $p = 1.55$. Similarly for experiment two, remote participants who received subgoal feedback ($M = 19.5$, $SD = 5.7$) did not spend significantly more time on the feedback tasks than those who received step feedback ($M = 21.5$, $SD = 4.2$), $t(53) = -1.50$ $p = 1.40$. Levene's test indicated unequal variances ($F = 5.19$, $p = .03$), so degrees of freedom were adjusted from 58 to 53. For tasks where feedback was provided, the step feedback group

outperformed the subgoal feedback group without taking more time to complete the tasks.

3.3.2 *Performance on Tasks without Feedback*

The maximum score participants could earn for completing both problem solving tasks correctly was 12. Independent samples t-test were used to compare the two-groups. The same scoring scheme as described above was used for this assessment as well. These tasks included both near and far transfer tasks. I expected to see an interaction with transfer level and feedback level. As the need to understand the structure of the problem increases, the benefits of subgoal learning might increase.

For experiment one, lab participants who received subgoal feedback ($M = 6.6$, $SD = 3.5$) correctly completed more steps than those who received step feedback ($M = 4.2$, $SD = 2.6$), $t(54) = 3.00$, $p < .01$, $d = 0.78$. Levene's test indicated unequal variances ($F = 8.91$, $p < .01$), so degrees of freedom were adjusted from 58 to 54.

For experiment two, remote participants who received subgoal feedback ($M = 6.3$, $SD = 4.4$) correctly completed more steps than those who received step feedback ($M = 3.6$, $SD = 3.7$), $t(56) = 2.55$, $p = .01$, $d = 0.66$. Levene's test indicated unequal variances ($F = 5.85$, $p = .02$), so degrees of freedom were adjusted from 58 to 56.

3.3.2.1 Near and Far Transfer Subgoals.

The subgoals in the assessment were classified as either “near” or “far” based on their relationship to the subgoals taught in the initial instructional materials. Near transfer subgoals were operationally defined as tasks that use the same steps (and therefore

subgoals) as the ones in the training materials but the context-specific details differ. Far transfer subgoals were defined as problems where the steps must be adapted to complete the task. For example, for the subgoal “Set Properties” participants were taught to adjust different properties of a label component. A near transfer task would be making different adjustments to a same or different app’s label. A far transfer task would be setting the properties of a component they had not been taught about, such as adjusting the properties of an animation component. The subgoal “Set Properties” remains the same, but the steps are either able to be reproduced with different superficial details (near transfer) or need to be adapted (far transfer).

Independent samples t-tests were used to compare the two groups. Scores were summed and participants could earn up to six points for near transfer subgoals and four points for far transfer subgoals.

In experiment one, lab participants who received subgoal feedback ($M = 3.9$, $SD = 2.1$) correctly completed more near transfer subgoals than those who received step feedback ($M = 2.6$, $SD = 1.7$), $t(55) = 2.634$, $p = .01$, $d = 0.68$. Levene’s test indicated unequal variances ($F = 5.7$, $p = .020$), so degrees of freedom were adjusted from 58 to 55. Similarly in experiment two, remote participants who received subgoal feedback ($M = 3.7$, $SD = 2.7$) correctly completed more near transfer subgoals than those who received step feedback ($M = 2.1$, $SD = 2.3$), $t(57) = 2.61$, $p = .01$, $d = 0.64$. Levene’s test indicated unequal variances ($F = 5.3$, $p = .03$), so degrees of freedom were adjusted from 58 to 57.

Lab participants who received subgoal feedback ($M = 1.7$, $SD = 1.3$) correctly completed more far transfer subgoals than those who received step feedback ($M = 0.8$, SD

= 1.2), $t(58) = 2.83$, $p = .01$, $d = 0.72$. Remote participants who received subgoal feedback ($M = 1.8$, $SD = 1.5$) also correctly completed more far transfer subgoals than those who received step feedback ($M = 0.7$, $SD = 1.4$), $t(58) = 2.84$, $p = .01$, $d = 0.76$.

3.3.2.2 Time on Task.

Time on task was measured in minutes. Lab participants who received subgoal feedback ($M = 8.4$, $SD = 3.0$) spent more time working on the assessment than those who received step feedback ($M = 5.4$, $SD = 3.7$), $t(58) = 3.54$, $p = .001$, $d = 0.89$. Similarly for experiment two, remote participants who received subgoal feedback ($M = 8.3$, $SD = 3.2$) spent more time on the assessment than those who received step feedback ($M = 5.2$, $SD = 1.5$), $t(42) = 3.07$, $p < .001$, $d = 1.24$. Levene's test indicated unequal variances ($F = 10.57$, $p = .002$), so degrees of freedom were adjusted from 58 to 42.

One statistically significant correlation was found for time on task and performance with the remote subgoal group $r = .43$, $p = .02$. The other groups did not have statistically significant correlations for time on task and performance measured as correctly completed steps (see Table 6).

Table 6 -- Distribution of Time on Problem Solving Assessment and Correlation with Correct Steps Taken for the Problem Solving Assessment for Experiment One (Lab) and Two (Remote)

Condition	<i>n</i>	<i>M</i>	<i>SD</i>	<i>r</i>	<i>p</i>
Lab Step level	30	5.4	3.7	.20	.286
Lab Subgoal level	30	8.4	3.0	.037	.845
Remote Step level	30	5.2	1.5	.015	.938
Remote Subgoal level	30	8.3	3.2	.426	.019*

3.4 Cognitive Load Assessment

The cognitive load questionnaire assessed the amount of mental effort that participants experienced during the instructional phase of the study. Intrinsic cognitive load is the complexity of the information that is innate to the material (Paas et al., 2003). This study did not manipulate intrinsic complexity, and therefore a significant difference among conditions was not expected. There were no significant differences between the groups in either experiment for intrinsic load (see Table 7).

Table 7 -- T-tests for Cognitive Load Assessment

Load Type	Condition	<i>n</i>	<i>M</i>	<i>SD</i>	<i>t</i>	<i>df</i>	Std. error	<i>p</i>
Intrinsic	Lab Step level	30	3.1	1.7	-.795	58	.433	.430
	Lab Subgoal level	30	3.5	1.6				
Intrinsic	Remote Step level	30	3.4	1.8	.468	58	.404	.642
	Remote Subgoal level	30	3.2	1.3				
Germane	Lab Step level	30	4.7	1.6	.510	58	.376	.612
	Lab Subgoal level	30	4.5	1.3				
Germane	Remote Step level	30	5.0	1.6	.769	58	.411	.445
	Remote Subgoal level	30	4.7	1.6				

Extraneous load refers to features of instructions that hinder learning (Paas et al., 2003). In the present study, instructions were the same with the exception of the type of feedback provided. Both step feedback and subgoal feedback were expected to benefit learning. The type of feedback was not expected to include differences in extraneous information, and therefore differences in extraneous cognitive load were not expected.

A significant but small effect was found in the lab condition. The subgoal group ($M = 1.8$, $SD = 1.4$) reported higher extraneous load than the step feedback group ($M = 1.2$, $SD = 1.1$) $t(54) = 2.15$, $p = .04$, $d = 0.52$. Levene's test indicated unequal variances ($F = 4.47$, $p = .04$), so degrees of freedom were adjusted from 58 to 54. The scale for this survey ranged from 0-10, so although this finding was significant, the means of 1.8 and 1.2 on a ten point scale demonstrate that both groups reported experiencing low levels of extraneous load. There were no significant differences between the remote step group ($M = 2.1$, $SD = 1.6$) or subgoal feedback group ($M = 2.2$, $SD = 1.9$) $t(58) = -.15$, $p = .89$.

Germane load refers to working memory resources that are being used to enhance learning (Paas et al., 2003). Subgoal feedback was hypothesized to focus working memory resources on understanding the structure of the problem solving solutions which could be reflected in the germane cognitive load score. However, this hypothesis was not supported. No significant differences were found between the groups in either experiment for germane load (see Table 7).

3.5 Explanation Task

The participants completed an explanation task to measure how well they could organize and explain problem solutions. To do well on this assessment, participants did not need to be able to solve problems, but they did need to recognize the steps of the solution that are structurally similar and explain why they are similar. Participants received two scores for this assessment: a grouping score for how well they organized the solution steps and an explanation score for how well they explained the groups. To score

the grouping portion of this task, participants received one point for each group that contained only structurally similar steps and could earn up to 9 points.

ICC(A) for this assessment was .98. For experiment one, the lab subgoal feedback group ($M = 5.1$, $SD = 2.7$) made more structurally similar groupings than the step level group ($M = 2.8$, $SD = 1.8$), $t(51) = 3.97$, $p < .001$, $d = 1.00$. Levene's test indicated unequal variances ($F = 7.46$, $p = .01$), so degrees of freedom were adjusted from 58 to 51. For experiment two, the remote subgoal feedback group ($M = 4.0$, $SD = 2.8$) also made more structurally similar groupings than the step level group ($M = 1.6$, $SD = 1.3$), $t(41) = -4.42$, $p < .001$, $d = 1.10$. Levene's test indicated unequal variances ($F = 16.24$, $p < .001$), so degrees of freedom were adjusted from 58 to 41.

For the explanation score, explanations were analyzed qualitatively to determine if participants correctly identified the purpose of the groups. In experiment one, over 60% of the labels created by those who received subgoal feedback correctly identified the purpose of the steps. Conversely, less than 20% of the labels created by step feedback participants demonstrated a structural understanding of the steps. In experiment two, similar results were found with 45% of the subgoal group's labels and 15% of the step feedback group's labels identifying the purpose of the steps.

These results suggest that subgoal feedback helps learners better understand the structure and purpose of solutions.

3.6 Other Measures

3.6.1 Attempted Assessments

Participants worked under timed conditions and might not have completed all the assessment tasks. To account for this, participants were given an “attempted” score for the number of assessment tasks they began. Any interaction with the AppInventor interface by the participant for a given task were counted as an attempted task (as opposed to just clicking in Qualtrics and advancing to the next assessment). The groups were originally going to be compared to see if one feedback style allowed more tasks to be attempted than the other, should accuracy between the two groups be similar. However, all participants except for five attempted all assessment tasks.

CHAPTER 4. DISCUSSION

4.1 Conclusions

Participants who received subgoal level feedback performed better than those who received step level feedback presumably because centering problem solving around subgoals helps the learner recognize the structure of the problem solving procedure. These results are consistent with previous research that shows that when a person learns subgoals, they can solve novel problems more effectively compared to someone who did not learn the subgoals (e.g., Catrambone, 1998; Margulieux et al., 2012; Margulieux & Catrambone, 2017). Subgoal labeled materials might help students apply knowledge to novel situations because subgoal labels facilitate the organizational aspect of learning the steps needed to solve problems in a domain, which in turn improves their ability to solve novel problems in the same domain (Bassok & Holyoak, 1989; Catrambone, 1998).

The results of this study support providing subgoal level feedback to novices. Learners who received subgoal feedback were not only better able to solve novel problems that had solution procedures that differed superficially from the instructional materials, but also to those problems where the solution steps needed to be adapted to solve the problem. These findings support the literature that shows that subgoal labels encourage learners to think about the problems in more abstract terms, additionally aiding them when solving new problems (e.g., Catrambone & Holyoak, 1990) and that subgoal learning encourages self-explanation which, in turn, aids transfer (Renkl & Atkinson, 2002).

Step level feedback can be time-consuming to create because the feedback differs for structurally similar problems due to superficial differences. Subgoal level feedback can be more quickly applied to novel problems because it transfers to new problems. For example, a student struggling with the algebra problems “ $9x=18$ ” or “ $4x=12$ ” could be given the subgoal level feedback of “Isolate x ” for either problem. Meanwhile, step level feedback might be “Divide both sides by 9” for the first problem, and “Divide both sides by 4” for the second problem. Customizing the feedback to include the superficial details of the specific problem is time-consuming for the feedback creator. Additionally, the step level feedback might focus the learner’s attention on superficial details, such as the 9 or 4, instead of the larger problem solving structure.

Effects demonstrated in lab settings have inherent value because they demonstrate what can happen, even if it rarely does (Mook, 1983). Being able to learn on one’s own, without a teacher present, is one of the greatest benefits of using an ITS. Learners are able to use an ITS in a variety of places such as their home that, compared to a lab environment, are less controlled. Therefore, demonstrating that a learning intervention is effective in settings that are similar to those an ITS user would naturally use demonstrates the power of the intervention to withstand environmental effects that can impede learning.

4.2 Limitations and Future Research

This study focuses on same-day learning of novices in computer programming. Subgoal learning research is often used with novices in science, technology, engineering, and math (STEM) domains. Future research should apply subgoal feedback in other

STEM and non-STEM domains. One reason subgoal learning is effective with novices is because they struggle to identify the structure of problem solving procedures. More advanced learners have a stronger problem solution framework than novices, but they are still less developed than an expert's. Applying subgoal learning and subgoal feedback to learners with more experience in a domain should also be explored.

In this study, learners were able to ask the ITS for assistance. For any given task, the ITS would first provide subgoal or step level feedback. If the ITS was again asked for more assistance to the same problem, the ITS would provide a worked example of a similar task from the instructional period. Subgoal level participants asked for the additional feedback that was given in the form of worked examples while step level participants did not. Future research is needed to explore how this difference in types of feedback impacted problem solving performance.

The cognitive load survey was chosen because it has been shown to not just detect differences in total cognitive load (e.g. the NASA TLX) but detects difference in the types of cognitive load: extraneous, intrinsic, and germane. The current study did not find any differences in intrinsic or germane load between the conditions but did find that the lab subgoal group reported higher extraneous load than the step feedback group. There were no significant differences between the remote step group and subgoal feedback group. If true extraneous load differences existed between the different types of feedback, I would have expected to find similar reports in both the lab and the remote group.

While the cognitive load survey was chosen because it had shown to be effective in other domains, a few issues might have impacted the accuracy of the survey in the

current study. Taking a closer look at the way learners use subgoal and step feedback might give some insight into the possible differences in cognitive load. When a learner receives the step feedback, they follow the steps with few needed inferences. They go to the correct menu and take the prescribed steps.

When a learner receives subgoal feedback, they think about how to apply that subgoal to the task at hand. During the study, participants would sometimes say aloud, “but how do I actually do that?” after receiving subgoal feedback. The learner then works to apply the subgoal and determine the steps needed for their specific situation. For the subgoal “create components,” the learner starts by identifying the menu where components are located. Then they search through the components and work to identify which component makes the most sense for this situation. This often involves some trial and error as the learner pulls different components, reads about their properties, refers back to the task at hand, and determines which component accomplishes the task.

The questions in the cognitive load survey related to extraneous load are as follows:

- [4] The instructions and/or explanations during the activity were very unclear.
- [5] The instructions and/or explanations were, in terms of learning, very ineffective.
- [6] The instructions and/or explanations were full of unclear language.

The participants came from a range of backgrounds and majors so it is possible that some were more familiar with the language used in the feedback than others. The questions above might not account for other factors that affect extraneous load such as

gaze patterns. For example, if learners in the subgoal group spent more effort directing their gaze from the Bluejeans chat where the feedback was found and back to the AppInventor screen, this survey would not have captured this extraneous load. Although these questions might have worked well in other studies, in hindsight they might not have been a good fit for the current study.

The questions in the cognitive load survey related to germane load are as follows:

[7] The activity really enhanced my understanding of the topic(s) covered.

[8] The activity really enhanced my knowledge and understanding of programming.

[9] The activity really enhanced my understanding of the procedures covered.

[10] The activity really enhanced my understanding of concepts and definitions.

The subgoal group might have been exerting more mental work and using more cognitive resources to apply the subgoal feedback but not have been able to express this using the given questions. Both groups reported similar ratings for the instructions enhancing their understanding even though one group might have been engaging in more mental work than the other.

Recent research supports a two-factor intrinsic-extraneous model of cognitive load as a sufficient and valid tool for assessing levels of cognitive load (Jiang & Kalyuga, 2020). Jiang and Kalyuga (2020) argue that germane load shares the identical theoretical ground with intrinsic load and is therefore redundant. This could also explain the lack of differences between the conditions for germane load because both step level and subgoal

level feedback provided instruction on the same tasks in the same domain. I would expect the intrinsic load to be similar because the instructional topic remained the same. Additionally, this raises more potential questions about the accuracy of the questionnaire in measuring germane load. These questions ask learners to report greater enhanced knowledge or understanding of the topic, which seems to be more aimed at perceived knowledge acquisition than workload related to the domain.

On a final note, when completing the cognitive load questionnaire, participants had completed both the initial instructional phase of creating the Magic 8-Ball app as well as the tasks with feedback. It is possible that they applied this scale to the overall instructions and activity, and not just the feedback portion. Redesigning the questionnaire page on the Qualtrics screen to make clear that participants should think about only the feedback given during the activity while they complete the following ratings could clear up any confusion that might have existed during the current study.

The cognitive load questionnaire is really asking one's perception of cognitive load. Physiological measures are much more accurate for accessing cognitive load than self-reports but this study was designed to be completed as similarly as possible in both a lab and remote environment. Follow up studies using physiological measure in a lab could provide further insight into the impact of subgoal feedback on cognitive load.

Using subgoal learning has several potential implications for improving communication among people with differing abilities and backgrounds because subgoals provide a common structure as well as a shared language for communicating. Subgoals could aid universal design by allowing people with differing abilities to communicate on

a mutually understood level. For example, blind and sighted people might not be able to communicate at the step level where a step such as "click on the text-type box" is not applicable to the blind person but using subgoals such as "set property" would apply to both blind and sighted people. Subgoals could aid teams where people with different backgrounds (e.g., psychology, design, physics) use different approaches when discussing and solving problems. Team members could still share subgoals even if the details about approaching them might be different. Additionally, they can use shared language at the subgoal level even when only one team member can really understand the steps for a particular subgoal while another member understands the steps for a different subgoal.

APPENDIX A. EXCERPT OF INSTRUCTIONS

Steps for creating components and setting properties

Magic 8-Ball will have a **Button** component that displays the image of the 8-Ball you downloaded earlier.

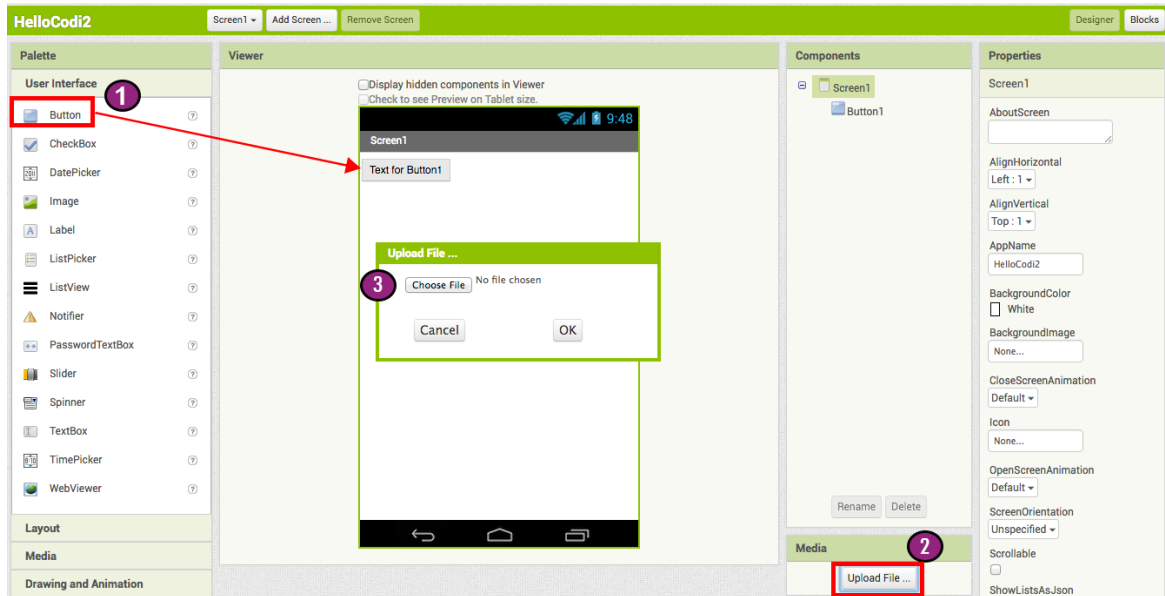
To accomplish this:

Create Components

From the **User Interface** palette, drag and drop the **Button** component to Screen1 (#1).

Set Properties

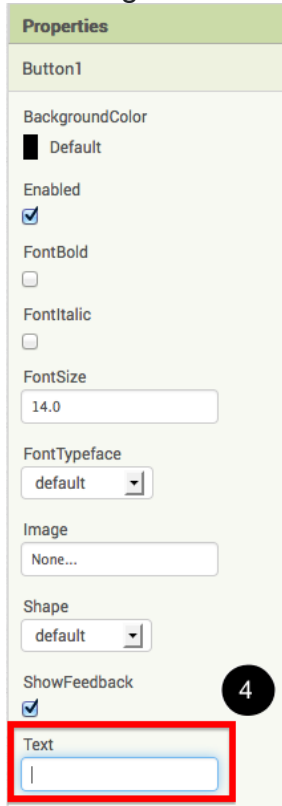
To give the button the image of the 8-Ball, in the **Properties** pane, under Image, click on the text "*None...*" and click "*Upload File...*" (#2). A window will pop up to let you choose the image file. Click "Browse" and then navigate to the location of the *magic8ball.jpg* file you downloaded earlier (#3). Click the *magic8ball.jpg* file, click "Open", and then click "OK".



Set properties

Change the Button's **Text** property:

Delete "Text for Button1", leaving the Button's text property blank so that there is no writing over the 8-Ball image (#4).



Properties

Button1

BackgroundColor
Default

Enabled
☒

FontBold
☐

FontItalic
☐

FontSize
14.0

FontTypeface
default

Image
None...

Shape
default

ShowFeedback
☒

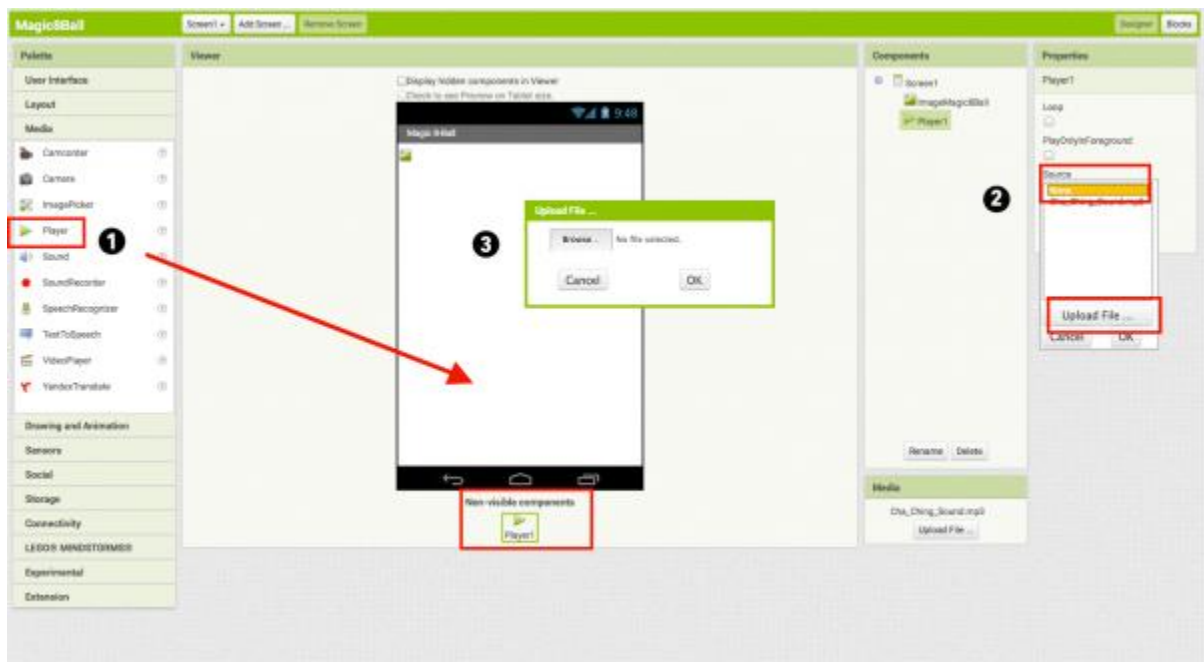
Text
|

Create Components

Under Palette, click on the **Media** drawer and drag out a **Player** component and place it in the Viewer (#1). Notice that since the **Player** will not be a visible part of the app, it appears at the bottom of the Viewer pane as a **Non-visible component**.

Set Properties

Set the **Player** component's source file:
Click on your newly added **Player** component to see its properties in the Properties pane on the right. Under Source click in the small box on the word "None..." and a small selection window will pop up (#2). Click the "Upload File" button and browse to where you saved the sound file. Select the sound file, then click OK to close the selection window. Click OK again on the properties pane to close the small popup window (#3).



APPENDIX B. ASSESSMENT ONE, STEP LEVEL FEEDBACK

Complete the following tasks in App Inventor.

1. Italicize the “Ask the Magic 8-Ball a question” command on your Magic 8 Ball app.

Feedback 1: You need to select the text label.

Feedback 2: You need to Click “FontItalic”

2. You can create a ball that moves around your screen at a set heading (in degrees, 0 degrees is towards the right, 90 degrees is towards the top), set interval (in milliseconds), and set speed (in pixels). Make a ball that moves at a rate of 5 pixels every 250 milliseconds towards the right of the screen (hint: animation components must be on a canvas).

Feedback 1: You need to drag out a ball.

Feedback 2: You need to set the heading.

Feedback 3: You need to set the interval.

Feedback 4: You need to set the speed.

3. Create a list of colors and make the ball change to a random color whenever it collided with something. (Hint: Use the blocks editor).

Feedback 1 : Drag out “when Ball1.CollidedWith”

Feedback 2 : Add “set Ball1.PaintColor” and put it in “do”

Feedback 3: Add “pick a random list item” and put it in “to”

Feedback 4: Add “make a list” and put it in “list”

Feedback 5: Add colors to list

4. Make the ball change direction. (This is called heading in App Inventor.) Change ball's direction to 90 degrees more than its current direction whenever it is touched.

Feedback 1: Drag out “when Ball1.Touched”

Feedback 2: Add “Set Ball1.Heading” and put in “do”

Feedback 3: From math, add “+” block

Feedback 4: Add “Ball1.Heading” and “90” to the “+” block

APPENDIX C. ASSESSMENT ONE, SUBGOAL FEEDBACK

Complete the following tasks in App Inventor.

Italicize the “Ask the Magic 8-Ball a question” command on your Magic 8 Ball app.

Feedback 1: You need to “Set Properties”.

Feedback 2: Here is an example of when you “Set Properties” earlier:

Set properties

Change the Button's **Text** property:

Delete "Text for Button1", leaving the Button's text property blank so that there is no writing over the 8-Ball image

1. You can create a ball that moves around your screen at a set heading (in degrees, 0 degrees is towards the right, 90 degrees is towards the top), set interval (in milliseconds), and set speed (in pixels). Make a ball that moves at a rate of 5 pixels every 250 milliseconds towards the right of the screen (hint: animation components must be on a canvas).

Feedback 1: You need to Create Components.

Feedback 2: Here is an example of when you “Created Components” earlier:

Create Components

From the **User Interface** palette, drag and drop the **Button** component to Screen1

Feedback 3: You need to Set Properties.

Feedback 4: Here is an example of when you “Set Properties” earlier:

Set properties

Change the Button's **Text** property:

Delete "Text for Button1", leaving the Button's text property blank so that there is no writing over the 8-Ball image

2. Create a list of colors and make the ball change to a random color whenever it collided with something. (Hint: Use the blocks editor).

Feedback 1 : You need to “Handle Events from My Blocks”

Feedback 2 : Here is an example of when you “**Handled Events from My Blocks**” earlier:

Handle Events from My Blocks

Under **Screen1**, click on **Button1**. Drag the “when Button1.Click” block into the work area.

Feedback 3: You need to “Set Outputs from My Blocks”

Feedback 4: Here is an example of when you “Set Outputs from My Blocks” earlier:

Set Outputs from My Blocks

Click on the **Player1** drawer, drag the “call Player1.Start” block into the work area and

insert it into the “when Button1.Click” block.

Feedback 5: **You need to “Set Conditions from Built-in”**

Feedback 6: Here is an example of when you “**Set Conditions from Built-in**” earlier:

Set Conditions from Built-in

From the Built-In drawer, click on the **Lists** drawer. Drag over the “pick random item” block and connect it to the open socket of the “set Label2.Text” block.

3. Make the ball change direction. (This is called heading in App Inventor.) Change ball's direction to 90 degrees more than its current direction whenever it is touched.

Feedback 1 : You need to “Handle Events from My Blocks”

Feedback 2 : Here is an example of when you “**Handled Events from My Blocks**” earlier:

Handle Events from My Blocks

Under **Screen1**, click on **Button1**. Drag the “when Button1.Click” block into the work area.

Feedback 3: You need to “**Set Outputs from My Blocks**”

Feedback 4: Here is an example of when you “**Set Outputs from My Blocks**” earlier:

Set Outputs from My Blocks

Click on the **Lists** drawer. Drag over the pick random item block and connect it to the open socket of the set Label2.Text block.

Feedback 5: **You need to “Set Conditions from Built-in”**

Feedback 6: Here is an example of when you “**Set Conditions from Built-in**” earlier:

Set Conditions from Built-in

From the Built-In drawer, click on the **Lists** drawer. Drag over the “pick random item” block and connect it to the open socket of the “set Label2.Text” block.

APPENDIX D. STEP FEEDBACK EXAMPLE FOR ASSESSMENT 2,

TASK 1

Task: Italicize the “Ask the Magic 8-Ball a question” on your Magic 8-Ball app.

First interaction with experimenter (disguised as ITS):

Participant: asks for help or has not made progress in 30 seconds:

ITS: It looks like you are taking steps to italicize the “Ask the Magic 8-Ball a question”. Do you need help?

Participant: Takes incorrect actions.

ITS: Hmm, the current configuration doesn't italicize the “Ask the Magic 8-Ball a question”. Please try again.

Participant: Takes correct actions but the task is not completed:

ITS: You have made progress towards italicizing the “Ask the Magic 8-Ball a question”. Keep going!

Participant: Takes correct steps and completes the task:

ITS: You have italicized the “Ask the Magic 8-Ball a question”. Great job!

Subsequent interactions:

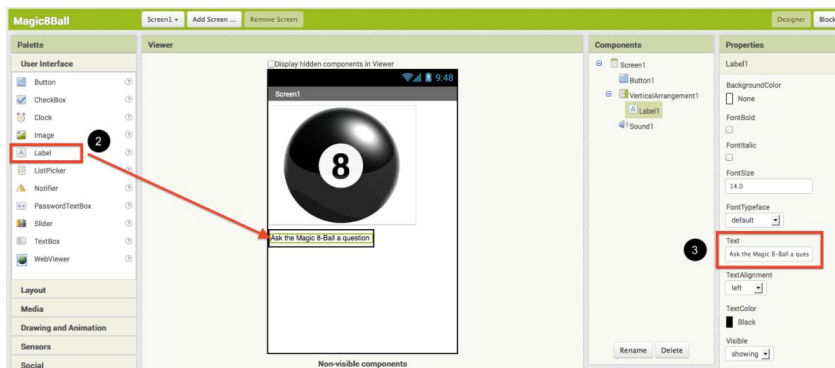
Participant: Takes incorrect actions or asks for feedback:

ITS provides step level feedback:: You need to select the text label.

Participant: Again takes incorrect actions or asks for feedback

ITS provides step level feedback: Here is an example of when you changed a text property earlier:

Under the **Properties** pane, change the *Text* property of **Label 1** to "Ask the Magic 8-Ball a question".



Participant: Again takes incorrect actions or asks for feedback

ITS provides step-level feedback: You need to select the text label.

OR

ITS provides step-level feedback: You need to Click “Font Italic”.

APPENDIX E. SUBGOAL FEEDBACK EXAMPLE FOR ASSESSMENT 2, TASK 1

Task Italicize the “Ask the Magic 8-Ball a question” on your Magic 8-Ball app.

First interaction with experimenter (disguised as ITS):

Participant: asks for help or has not made progress in 30 seconds:

ITS: It looks like you are taking steps to italicize the “Ask the Magic 8-Ball a question”. Do you need help?

Participant: Takes incorrect actions.

ITS: Hmm, the current configuration doesn't italicize the “Ask the Magic 8-Ball a question”. Please try again.

Participant: Takes correct actions but the task is not completed:

ITS: You have made progress towards italicizing “Ask the Magic 8-Ball a question”. Keep going!

Participant: Takes correct steps and completes the task:

ITS: You have italicized the “Touch the Bee!” command. Great job!

Subsequent interactions:

Participant: Takes incorrect actions or asks for feedback:

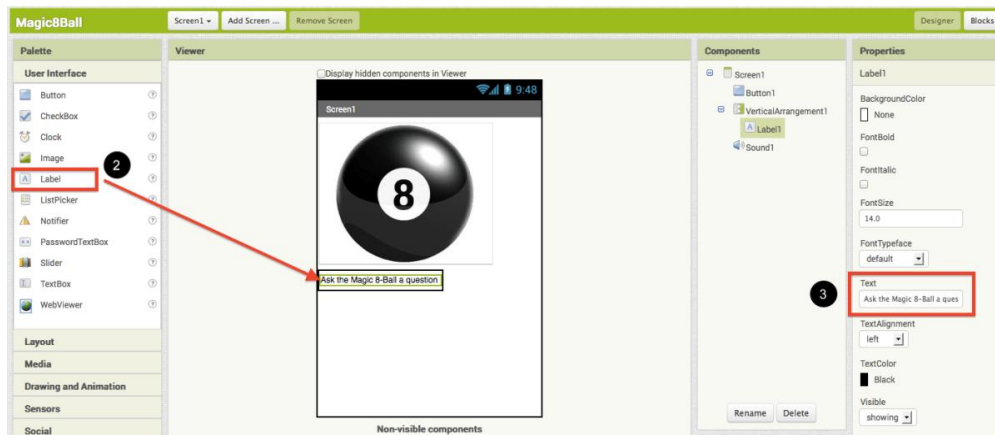
ITS provides subgoal level feedback: You need to “Set Properties”

Participant: Again takes incorrect actions or asks for feedback

ITS provides subgoal feedback: Here is an example of when you “Set Properties” earlier:

Set Properties

2. Under the **Properties** pane, change the *Text* property of **Label 1** to "Ask the Magic 8-Ball a question".



Participant: Again takes incorrect actions or asks for feedback
ITS: cycles through above feedback options.

APPENDIX F. ASSESSMENT TWO

A ten-item questionnaire for the measurement of IL (Items 1, 2, and 3), EL (Items 4, 5, and 6), and GL (Items 7, 8, 9, and 10) developed from Leppink et al. (2013).

All of the following questions refer to the activities that just finished. Please respond to each of the questions on the following scale (0 meaning *not at all the case* and 10 meaning *completely the case*). 0 1 2 3 4 5 6 7 8 9 10

- [1] The topic/topics covered in the activity was/were very complex.
- [2] The activity covered procedures that I perceived as very complex.
- [3] The activity covered concepts and definitions that I perceived as very complex.
- [4] The instructions and/or explanations during the activity were very unclear.
- [5] The instructions and/or explanations were, in terms of learning, very ineffective.
- [6] The instructions and/or explanations were full of unclear language.
- [7] The activity really enhanced my understanding of the topic(s) covered.
- [8] The activity really enhanced my knowledge and understanding of programming.
- [9] The activity really enhanced my understanding of the procedures covered.
- [10] The activity really enhanced my understanding of concepts and definitions.

APPENDIX G. ASSESSMENT THREE – NO FEEDBACK

1. Bold the “Ask the Magic 8-Ball a question” command on your Magic 8 Ball app.
2. You have learned how to create a button that has a picture of an 8 Ball, and then program the button so that when it is clicked the ball will show a fortune!
Now, make an app for a cat that meows.

Start a new project and name it "FinalKitty"

Include a label that says "Touch the Cat"

APPENDIX H. ASSESSMENT FOUR – EXPLANATION TASK

Below are the steps to the solutions of the some of the problems that you were just working on. The steps are correct and in the correct order. Please group the steps of these solutions that you think go together by entering them below in brackets. “Go together” is open to your interpretation, but think of it as if you were trying to put headers into the solution to group steps in some meaningful way. If you do not think any of the steps go together, you do not have to group any steps. If you group steps together, please provide a label or description of why you think those steps go together.

1. Italicize the “Ask the Magic 8-Ball a question” command on your Magic 8 Ball app.

Select Label

Click “FontItalic”

2. You can create a ball that moves around your screen at a set heading (in degrees, 0 degrees is towards the right, 90 degrees is towards the top), set interval (in milliseconds), and set speed (in pixels). Make a ball that moves at a rate of 5 pixels every 250 milliseconds towards the right of the screen (hint: animation components must be on a canvas).

Drag out Ball

Set Heading to 0

Set	Interval	to	250
Set	Speed	to	5

3. Create a list of colors and make the ball change to a random color whenever it collided with something. (Hint: Use the blocks editor).

Drag out “when Ball1.CollidedWith”

Add “set Ball1.PaintColor” and put it in “do”

Add “pick a random list item” and put it in “to”

Add “make a list” and put it in “list”

Add colors to list

4. Make the ball change direction. (This is called heading in App Inventor.) Change ball's direction to 90 degrees more than its current direction whenever it is touched.

Drag out “when Ball1.Touched”

Add “Set Ball1.Heading” and put in “do”

From math, add “+” block

Add “Ball1.Heading” and “90” to the “+” block

REFERENCES

- Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In G. Gauthier, C. Frasson, & K. VanLehn (Eds.), *Proceedings of the 5th International Conference on Intelligent Tutoring Systems, ITS 2000* (pp. 292–303). Berlin: Springer-Verlag.
- Atkinson, R. K., Catrambone, R., & Merrill, M. M. (2003). Aiding Transfer in Statistics: Examining the Use of Conceptually Oriented Equations and Elaborations During Subgoal Learning. *Journal of Educational Psychology*, 95(4), 762-773.ca
- Ashford, S. J., Blatt, R., & VandeWalle, D. (2003). Reflections on the looking glass: A review of research on feedback-seeking behavior in organizations. *Journal of Management*, 29, 773–799.
- Azevedo, R., & Bernard, R. M. (1995). A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, 13(2), 111–127.
- Bailey, B. P., & Konstan, J. A. (2006). On the need for attention-aware systems: Measuring effects on interruption on task performance, error rate, and affective state. *Computers in Human Behavior*, 22, 685–708.
- Bangert-Drowns, R. L., Kulik, C. C., Kulik, J. A., & Morgan, M. T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213–238.
- Baumeister, R. F., Hutton, D. G., & Cairns, K. J. (1990). Negative effects of praise on skilled performance. *Basic and applied social psychology*, 11(2), 131-148.
- Baron, R. A. (1993). Criticism (informal negative feedback) as a source of perceived unfairness in organizations: Effects, mechanisms, and countermeasures. In R. Cropanzano (Ed.), *Justice in the workplace: Approaching fairness in human resource management* (pp. 155–170). Hillsdale, NJ: Lawrence Erlbaum.
- Bisra, K., Liu, Q., Nesbit, J., Salimi, F., Winne, P. (2018). Inducing Self Explanation: A Meta-Analysis. *Educational Psychology Review* 30(3) 703-725

- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). How people learn: Brain, mind, experience, and school (Rev. ed.). Washington, DC: National Academies Press.
- Butler, R. (1987). Task-involving and ego-involving properties of evaluation: Effects of different feedback conditions on motivational perceptions, interest, and performance. *Journal of Educational Psychology*, 79(4), 474–482.
- Catrambone, R. (1995). Aiding subgoal learning: Effects on transfer. *Journal of Educational Psychology*, 87(1), 5-17. doi: 10.1037/0022-0663.87.1.5
- Catrambone, R. (1998). The subgoal learning model: Creating better examples so that students can solve novel problems. *Journal of Experimental Psychology: General*, 127, 355–376.
- Catrambone, R., Gane, B. D., Adams, A. E., Bujak, K. R., Kline, K. A., & Eiriksdottir, E. (2012). Task Analysis by Problem Solving (TAPS): A Method for Uncovering Expert Knowledge. Unpublished manuscript, School of Psychology, Georgia Institute of Technology, Atlanta, GA.
- Catrambone, R. & Holyoak, K.J. (1990). Learning subgoals and methods for solving probability problems. *Memory & Cognition*. 18(6), 593-603
- Chandler, P. & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, 8 (4), 293-332.
- Chen, D. & Catrambone, R. (2015). Paper vs. screen: Effects on reading comprehension, metacognition, and reader behavior. Proceedings of the 59th Annual Meeting of the Human Factors and Ergonomics Society (pp. 332-336). Santa Monica, CA: Human Factors and Ergonomics Society.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145-182. doi: 10.1207/s15516709cog1302_1
- Chi, M. T., Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, 18(3), 439-477.
- Clariana, R. B. (1990). A comparison of answer-until-correct feedback and knowledge-of-correct-response feedback under two conditions of contextualization. *Journal*

of Computer-Based Instruction, 17(4), 125–129.

- Collins, C., Brown, J., & Newman, S. (1989). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum.
- Corbett, A., & Anderson, J. R. (2001). Locus of feedback control in computer-based tutoring: Impact on learning rate, achievement and attitudes. In J. Jacko, A. Sears, M. Beaudouin-Lafon, & R. Jacob (Eds.), *Proceedings of ACM CHI 2001 Conference on Human Factors in Computing Systems* (pp. 245–252). New York, NY: ACM Press.
- Corno, L., & Snow, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 605–629). New York: Macmillan.
- Davis, W. D., Carson, C. M., Ammeter, A. P., & Treadway, D. C. (2005). The interactive effects of goal orientation and feedback specificity on task performance. *Human Performance*, 18, 409–426.
- Dihoff, R. E., Brosvic, G. M., Epstein, M. L., & Cook, M. J. (2003). The role of feedback during academic testing: The delay retention test revisited. *The Psychological Record*, 53, 533–548.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American psychologist*, 41(10), 1040.
- Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological review*, 95(2), 256.
- Fedor, D. B., Davis, W. D., Maslyn, J. M., & Mathieson, K. (2001). Performance improvement efforts in response to negative feedback: The roles of source power and recipient self-esteem. *Journal of Management*, 27(1), 79–97.
- Gaynor, P. (1981). The Effect of Feedback Delay on Retention of Computer-Based Mathematical Material. *Journal of Computer-Based Instruction*, 8(2), 28–34.
- Gilman, D. A. (1969). Comparison of several feedback methods for correcting errors by

- computer-assisted instruction. *Journal of Educational Psychology*, 60(6), 503–508.
- Graesser, A. C., McNamara, D., & VanLehn, K. (2005). Scaffolding deep comprehension strategies through AutoTutor and iSTART. *Educational Psychologist*, 40, 225–234.
- Eiriksdottir, E., & Catrambone, R. (2011). Procedural instructions, principles, and examples: How to structure instructions for procedural tasks to enhance performance, learning, and transfer. *Human Factors*, 53(6), 749-770. doi: 10.1177/0018720811419154
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77, 81–112. doi:10.3102/003465430298487
- Hoska, D. M. (1993). Motivating learners through CBI feedback: Developing a positive learner perspective. In V. Dempsey & G. C. Sales (Eds.), *Interactive instruction and feedback* (pp. 105–132). Englewood Cliffs, NJ: Educational Technology Publications.
- Hundhausen, C. D., Farley, S. F., & Brown, J. L. (2009). Can direct manipulation lower the barriers to computer programming and promote transfer of training?: An experimental study. *ACM Transactions in CHI*, 16(3). doi: 10.1145/1592440.1592442
- Jiang, D., & Kalyuga, S. (2020). Confirmatory factor analysis of cognitive load ratings supports a two-factor model. *Tutorials in Quantitative Methods for Psychology*, 16, 216-225.
- Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition. *Journal of Applied Psychology*, 74, 657–690.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119(2), 254–284.
- Koedinger, K., Brunskill, E., Baker, R., McLaughlin, E., & Stamper, J. (2013). New Potentials for Data-Driven Intelligent Tutoring System Development and Optimization. *AI Magazine*. 34(3), 27-41.

- Kulhavy, R., White, M., Topp, B., Chan, J. & Adams (1985). Feedback complexity and corrective efficiency. *Contemporary Educational Psychology*, 10(3) 285-291.
- Knoblauch, C., & Brannon, L. (1981). Teacher Commentary on student writing; the state of the art. *Freshman English News*, 10(2), 1-4. Retrieved from <http://www.jstor.org/stable/43518564>
- Kornell, N., & Bjork, R. A. (2009). A stability bias in human memory: overestimating remembering and underestimating learning. *Journal of Experimental Psychology: General*, 138(4), 449.
- LeFevre, J., & Dixon, P. (1986). Do written instructions need examples? *Cognition and Instruction*, 3, 1-30. doi: 10.1207/s1532690xc0301_1
- Leppink, J., Paas, F., Van der Vleuten, C. M., Van Gog, T., & Van Merriënboer, J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods*, 45(4), 1058-1072. doi:10.3758/s13428-013-0334-1
- Locke, E. A., & Latham, G. P. (1990). A theory of goal setting & task performance. Englewood Cliffs, NJ: Prentice Hall.
- Margulieux, L. E., & Catrambone, R. (2016). Improving problem solving with subgoal labels in expository text and worked examples. *Learning and Instruction*, 42, 58-71. doi: 10.1016/j.learninstruc.2015.12.002
- Margulieux, L. & Catrambone, R. (2017). Using Learners' Self-Explanations of Subgoals to Guide Initial Problem Solving in App Inventor. *Proceedings of the 2017 ACM Conference on International Computing Education Research*. (pp. 21-29). Tacoma, Washington, USA
- Margulieux, L. E., Catrambone, R., Schaeffer, L. M. (February, 2018). Varying effects of subgoal labeled expository text in programming, chemistry, and statistics. *Instructional Science*. <http://rdcu.be/G1tB>
- Margulieux, L. E., Guzdial, M., & Catrambone, R. (2012). Subgoal-labeled instructional material improves performance and transfer in learning to develop mobile applications. In *Proceedings of the Ninth Annual International Conference on International Computing Education Research*, 71-78. doi: 10.1145/2361276.2361291

- Mason, B. J., & Bruning, R. (2001). Providing feedback in computer-based instruction: What the research tells us. *Center for Instructional Innovation*, University of Nebraska–Lincoln: 14. Retrieved on January 16, 2017 from https://www.researchgate.net/publication/247291218_Providing_Feedback_in_Computer-based_Instruction_What_the_Research_Tells_Us
- Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational psychologist*, 38(1), 43-52.
- McCloskey, W., & Leary, M. R. (1985). Differential effects of norm-referenced and self-referenced feedback on performance expectancies, attribution, and motivation. *Contemporary Educational Psychology*, 10, 275–284.
- McLaughlin, A. C., Rogers, W. A., & Fisk, A. D. (2006). How effective feedback for training depends on learner resources and task demands. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 50, No. 25, pp. 2624-2628). SAGE Publications.
- Mook, D. G. (1983). In defense of external invalidity. *American psychologist*, 38(4), 379.
- Moreno, R. (2004). Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery-based multimedia. *Instructional Science*, 32, 99–113. doi:10.1023/B:TRUC.0000021811.66966.1d
- Moreno, R., & Valdez, A. (2005). Cognitive load and learning effects of having students organize pictures and words in multimedia environments: The role of student interactivity and feedback. *Educational Technology Research and Development*, 53, 35–45. doi:10.1007/BF02504796
- Narciss, S., & Huth, K. (2004). How to design informative tutoring feedback for multimedia learning. In H. M. Niegemann, D. Leutner, & R. Brunken (Ed.), *Instructional design for multimedia learning* (pp. 181–195). Munster, NY: Waxmann.
- Noyes, J., Garland, K., & Robbins, L. (2004). Paper-based versus computer-based assessment: is workload another test mode effect?. *British Journal of Educational Technology*, 35(1), 111-113.

- Paas, F., & van Gog, T. (2009). Principles for designing effective and efficient training of complex cognitive skills. In F. T. Durso (Ed.) *Reviews of Human Factors and Ergonomics (Vol. 5)*, Santa Monica, CA: HFES, pp. 166-194.
- Phye, G. D., & Bender, T. (1989). Feedback complexity and practice: Response pattern analysis in retention and transfer. *Contemporary Educational Psychology*, 14(2), 97–110.
- Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, 19(2), 221-243.
- Renkl, A., & Atkinson, R. K. (2002). Learning from examples: Fostering self-explanations in computer-based learning environments. *Interactive Learning Environments*, 10(2), 105-199. doi:10.1076/ilee.10.2.105.7441
- Salmoni, A. W., Schmidt, R. A., & Walter, C. B. (1984). Knowledge of results and motor learning: a review and critical reappraisal. *Psychological Bulletin*, 95(3), 355-386.
- Schaeffer, L. M., Margulieux, L. E., & Catrambone, R. (2016). Interaction of instructional materials order and subgoal labels on learning in programming. In A. Papafragou, D. Grodner, D. Mirman, & J.C. Trueswell (Eds.), *Proceedings of the 38th Annual Conference of the Cognitive Science Society* (pp. 271-276). Austin, TX: Cognitive Science Society.
- Schroth, M. L. (1992). The effects of delay of feedback on a delayed concept formation transfer task. *Contemporary Educational Psychology*, 17(1), 78–82.
- Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78, 153–189. doi:10.3102/0034654307313795
- Shute, V. J., Woltz, D. J., & Regian, J. W. (1989, May). An investigation of learner differences in an ITS environment: There's no such thing as a free lunch. Paper presented at the 4th International Conference on Artificial Intelligence and Education–AI-ED '89, Amsterdam, Holland.
- Song, S. H., & Keller, J. M. (2001). Effectiveness of motivationally adaptive computerassisted instruction on the dynamic aspects of motivation. *Educational Technology Research and Development*, 49(2), 5–22.

- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, 12(2), 257-285.
- Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), 123-138.
- Van der Kleij, F. M., Eggen, T. J., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, 58(1), 263-272.
- Van der Kleij, F. M., Feskens, R. W., and Eggen, T. M. (2015). Effects of feedback in a computer-based learning environment on students' learning outcomes. *Review Of Educational Research*, 85(4), 475-511.
- VandeWalle, D. (1997), Development and validation of a work domain goal orientation instrument, *Educational and Psychological Measurement*, 8, 995-1015.
- VanLehn, K. A. (2011). The relative effectiveness of human tutoring, intelligent tutoring systems, and other tutoring systems. *Educational Psychologist*, 46(4), 197-221. doi:10.1080/00461520.2011.611369.
- Wästlund, E. (2007). *Experimental studies of human-computer interaction: Working memory and mental workload in complex cognition*. (Doctoral dissertation). Retrieved from https://gupea.ub.gu.se/bitstream/2077/4693/1/gupea_2077_4693_1.pdf
- Wästlund, E., Reinikka, H., Norlander, T., & Archer, T. (2005). Effects of VDT and paper presentation on consumption and production of information: Psychological and physiological factors. *Computers in Human Behavior*, 21(2), 377-394.
- Wiliam, D. (2007). Changing classroom practice. *Educational leadership*, 65(4), 36.