

When practice does not make perfect:  
Differentiating between productive and unproductive persistence

Ma. Victoria Q. Almeda

Submitted in partial fulfillment of the  
requirements of the degree of  
Doctor of Philosophy  
under the Executive Committee  
of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018



## ABSTRACT

When practice does not make perfect: Differentiating between productive and unproductive persistence

Ma. Victoria Q. Almeda

Research has suggested that persistence in the face of challenges plays an important role in learning. However, recent work on wheel-spinning—a type of unproductive persistence where students spend too much time struggling without achieving mastery of skills—has shown that not all persistence is uniformly beneficial for learning. For this reason, Study 1 used educational data-mining techniques to determine key differences between the behaviors associated with productive persistence and wheel-spinning in ASSISTments, an online math learning platform. This study's results indicated that three features differentiated between these two modes of persistence: the number of hints requested in any problem, the number of bottom-out hints in the last eight problems, and the variation in the delay between solving problems of the same skill. These findings suggested that focusing on number of hints can provide insight into which students are struggling, and encouraging students to engage in longer delays between problem solving is likely helpful to reduce their wheel-spinning. Using the same definition of productive persistence in Study 1, Study 2 attempted to investigate the relationship between productive persistence and grit using Duckworth and Quinn's (2009) Short Grit Scale. Correlational results showed that the two constructs were not significantly correlated with each other, providing implications for synthesizing literature on student persistence across computer-based learning environments and traditional classrooms.

## Table of Contents

List of Tables .....	vi
List of Figures .....	vii
CHAPTER I – INTRODUCTION .....	1
Background .....	1
Statement of the Problem .....	2
Research Questions .....	3
CHAPTER II – REVIEW OF LITERATURE.....	5
Benefits of Persistence .....	5
Limitations of Persistence .....	7
Persistence Towards Productivity .....	10
Factors Impacting Productive Persistence .....	12
Self-regulation of Learning Strategies.....	13
Academic Mindsets.....	14
Academic Emotions .....	15
Studying Persistence Using Educational Data-Mining Methods .....	16
Modeling Persistence in Intelligent Tutoring Systems .....	18
Summary of the Literature Review .....	22
CHAPTER III – METHODOLOGY .....	24
Modeling in the ASSISTments System.....	24
The ASSISTments System .....	24
Goals of ASSISTments. ....	25

Functions of ASSISTments.....	25
Study 1 (Modeling in Wheel-Spinning in ASSISTments System: Methods and Results) .....	28
Participants.....	28
Wheel-Spinning Behavior.....	29
Operational definition .....	29
Feature engineering.....	32
Machine learning model .....	33
Results.....	34
Multi-feature model .....	34
Top three features in the J48 decision tree.....	36
Exploring model features .....	39
Summary of Study 1 Results.....	42
Understanding relations between productive persistence and grit.....	43
Study 2 (Correlating Productive and Grit: Methods).....	45
Participants.....	45
Measures .....	46
Short Grit Scale (Grit-S) .....	46
Productive persistence and unproductive persistence.....	47
Procedure .....	48
Main analyses.....	48
Student-level correlations .....	48
Hierarchical linear modeling.....	48

Secondary analyses .....	49
Time of year .....	49
Mastery speed .....	49
Difficulty of a Skill Builder .....	50
Grade level .....	50
Results .....	51
Main results .....	51
Student-level correlations .....	51
Hierarchical linear modeling .....	51
Secondary results .....	52
Time of year .....	52
Mastery speed .....	54
Difficulty of a Skill Builder .....	55
Grade level .....	57
Summary of Study 2 results .....	58
CHAPTER IV – DISCUSSION .....	60
Persistence in Computer-based Learning Environments .....	60
Hint use .....	60
Difficulty of a Skill Builder .....	62
Persistence Within and Beyond Computer-based Learning Environments .....	63
Spaced practice .....	64
Relationship between grit and online persistent behaviors .....	65

Implications for Intervention .....	67
Limitations and Future Work.....	69
Conclusion .....	71
REFERENCES .....	73
Appendix A: J48 Decision Tree of the Multi-feature Model.....	80
Appendix B: Features Selected with J48 Decision Tree Algorithm .....	83
Appendix C: Short Grit Scale (Grit-S).....	84
Appendix D: Skill Builders with Hard Difficulty Levels .....	86

## LIST OF TABLES

### Table

1	Criteria of Productive Persistence and Wheel-Spinning, Using Performance Metrics in the Skill Builders System .....	30
2	Relationships Between the First Combination of Features at the Top of the Tree and Wheel-Spinning Within the J48 Model .....	38
3	Relationships Between the Second Combination of Features at the Top of the Tree and Wheel-Spinning Within the J48 Model .....	39
4	Spearman Correlation of Persistent Behaviors to Grit-S Scores, Across Students.....	52
5	Spearman Correlation of Mastery Speed to Grit-S Scores, Across Student-problem Set Pairs .....	55

## LIST OF FIGURES

### Figure

1	Mastery of a Skill Builder in ASSISTments .....	26
2	Types of hints in ASSISTments .....	26
3	A precision-recall curve for the J48 model's predictions of wheel-spinning .....	35
4	Showing (a) Histogram of the minimum number of hints requested in any problem across student-problem set pairs (min-hintTotal) and (b) Proportion of student-problem set pairs identified as wheel-spinning for each min-hintTotal range of values .....	40
5	Showing (a) Histogram of the maximum number of bottom-out hints requested in the last eight problems across student-problem set pairs (max-past8BottomOut) and (b) Proportion of student-problem set pairs identified as wheel-spinning for each max-past8BottomOut range of values.....	41
6	Showing (a) Histogram of the standard deviation for the amount of time since the current set was last seen across student problem set pairs (std-timeSinceSkill) and (b) Proportion of student-problem set pairs identified as wheel-spinning for each max-past8BottomOut range of values.....	42
7	Proportions of productive and unproductive persistence for each month in the academic year (2016-2017) .....	53
8	Frequency of problem set pairs for each month in the academic year (2016-2017) .....	54

## ACKNOWLEDGMENTS

This work would have been impossible without the encouragement and support of my mentors, family, and friends. First and foremost, to Dr. Ryan Baker - I could have not asked for a better dissertation sponsor, career counselor, and life advisor; it has been the utmost honor and privilege being your student. Your guidance has been a transformative force, helping me shape my dissertation and explore my passion in learning analytics within K-12 education. Thank you for showing me the meaning of rigorous, ethical research and compassionate, intentional teaching. The lessons you imparted will continue to be a source of light and inspiration for me, as I embark on the next chapter as an educational researcher.

To the members of my dissertation committee - my heartfelt thanks for your generous time and invaluable feedback for improving my work. I would like to thank Dr. Herbert Ginsburg for reminding me of the bigger picture when doing research with children. Thank you for your unyielding support and dynamic conversations about children's mathematical development. I would like to thank Dr. Sandra Okita for encouraging me to reflect on the narrative of my work, especially its implications for educational interventions. To Dr. Stephen Peverly - thank you for imparting your expertise in cognition and emotions; your comments have urged me to deeply ruminate about my research. To Dr. Neil Heffernan - I truly appreciate your insights during weekly meetings and dissertation presentations. Thank you for your commitment in helping students further their research and career.

To Dr. Angela Duckworth, the leading expert on grit - thank you for the opportunity to present my dissertation at your lab. Having followed your work, I am profoundly grateful for our conversation, which will continue to serve as a wonderful reminder to never stop learning.

In addition, I would also like to thank members of the ASSISTments team, Korinn Ostrow and Seth Adjei, for going above and beyond as colleagues throughout my data collection and analysis. Thank you for your incredible patience and guidance with all my questions. My sincere thanks to my dear friend, Sergio Salmeron, for his programming expertise. Without your help, completing my dissertation would have taken twice as long. I am blessed to have your support and friendship.

To the alumni and current members of the Educational Data Mining Lab at Teachers College, especially Dr. Luc Paquette, Dr. Jaclyn Ocumpaugh, Dr. Sweet San Pedro, Dr. Elle Wang, Dr. Jeanine DeFalco, Aaron Hawn, Yang Jiang, Shimin Kai, Miggy Andres and Stefan Slater – you have all made Teachers College feel like home. Thank you for your unwavering encouragement and companionship through stressful and challenging times. I will cherish all of our adventures together and will miss having workmates like family.

To my Ph.D. support group, Colleen Uscianowski, Laura Malkiewich, Bess Sherry, Hanadi Shatara – I feel extremely lucky to have experienced graduate school alongside such talented and supportive friends. Thank you for listening and cheering me on during moments of worry and anxiety. To Myra Luna-Lucero, Ama Awotwi, Kelly Oppenzato, Eileen Wu, Bona Lee– thank you for the friendship and openness to learn from each other.

To my friends and relatives both in the U.S. and overseas, especially Liana Cruz, Nica Bannister, Emily Bautista, Marie La Vina, Mel Oliveros, and Kitkat Pecson– thank you for the overwhelming love and support. Your kind and encouraging words have sustained me throughout the years.

To my amazing partner, Ed – thank you for being my anchor and solace throughout this entire experience as a graduate student. Your love and strength has taught me to celebrate the

quiet triumphs and to laugh even in the most trying times. Our relationship continues to give me a fresh perspective on life; thank you for always inspiring me to be a better person.

Lastly, I owe a great debt of gratitude to my family, Mom, Dad, Jon, Ola, Mimi, and John, for your unconditional love and support. While many events have led to the fruition of my dissertation, this journey would have never begun without your first and ultimate decision to support and believe in me six years ago. Thank you for everything you've done for me, and I hope I have made you proud.

## DEDICATION

*For my first teachers, Mom and Dad.*

## CHAPTER I

### INTRODUCTION

#### **Background**

Persistence has been considered to be an important factor for achieving success in any endeavor. Recent studies have shown that persistence and similar constructs are associated with creativity (Prabhu, Sutton, & Sauser, 2008), academic achievement (Andersson & Bergman, 2011; Duckworth, Peterson, Matthews, & Kelly, 2007), and success in the workplace (Wrzesniewski, 2012).

Throughout the years, researchers and educators have used different terms to refer to the same ability of persevering, regardless of the person's inclination towards the task. Of particular interest is the popular term *grit*, defined as perseverance and passion of interest over long periods of time, which has been associated with long-term outcomes such as educational attainment and retention (Duckworth et al., 2007).

However, recent studies have suggested that not all persistence is uniformly positive. Within the context of computer-based learning environments, Beck and Gong (2013) asserted that some persistence may be “wheel-spinning,” defined as when a student spends too much time struggling to learn a topic without achieving mastery. A student who persists unsuccessfully may face eventual reduced motivation (Sedek & Kofta, 1990). Ultimately, wheel-spinning has also been linked to undesirable behaviors such as gaming the system, where students systematically

guess the correct answer or abuse the help functions within the online system (Beck & Rodrigo, 2014).

As such, we have two related phenomena—productive persistence and wheel-spinning, which is also referred to as unproductive persistence. The goal is to encourage students to persist productively and prevent them from persisting when the eventual outcome will be negative. Unfortunately, many students engage in unproductive persistence, often not taking the corrective action towards learning (Beck & Gong, 2013; Beck & Rodrigo, 2014; Gong & Beck, 2015). Specifically, as much as 35% of students were found to be wheel-spinning in two widely used math tutoring systems, ASSISTments and Cognitive Algebra Tutor (Beck & Gong, 2013). These findings suggest a need to distinguish between these two phenomena to refine the pedagogical theory on student persistence and to design intervention on their motivational experiences.

### **Statement of the Problem**

Given the academic interest in measuring the role of persistence in academic achievement, researchers are increasingly exploring the development of methods and instruments that measure persistence, both quantitatively and qualitatively. However, most prior studies on persistence have not attempted to distinguish between students' unproductive and productive struggle, until they have reached a point of failing or succeeding. This gap in the literature is problematic, as teachers may fail to understand that the same interventions that are appropriate for productively persistent students may not necessarily work for those who are unproductively persisting. For instance, students who engage in a negative cycle, repetitively using the same ineffective strategy to solve similar problems, are unlikely to benefit from demonstrating more grit. Further work is needed to understand the difference between productive and unproductive struggle to support persistence in a way that aligns with each student's needs.

Another issue that has not been fully explored within persistence research is the integration of findings on key terms in the field (e.g., perseverance, academic tenacity, grit). While a few studies in computer-based learning environments (CBLEs) have investigated students' incidence of wheel-spinning or unproductive persistence, the findings from these studies have not been linked to research on more global measures of persistence—such as grit. The interest in grit is likely growing because this concept more readily allows for intervention, as opposed to the other cognitive (e.g., intelligence) or sociopolitical factors (e.g., socioeconomic status) of the student (Duckworth & Gross, 2014). However, the research on grit has been largely localized, with little research unifying findings from other similar constructs that target effort in the face of challenges. Synthesizing empirical findings between conceptual distinctions can help sharpen interventions for improving learning. For instance, Duckworth and Gross (2014) asserted that the psychological mechanisms underlying grit may be, in part, related to deliberate and focused practice. The research on wheel-spinning, which involves micro-level information of students' problem solving within CBLEs, can help inform teachers about how and where deliberate practice can be more productive. As such, systematic work is needed to create a more consolidated framework about students' persistence and to enhance teacher practice that helps support productive struggle across different contexts, whether or not technology is integrated into the classroom.

### **Research Questions**

The first goal of this dissertation was to raise the importance of differentiating productive and unproductive persistence in CBLEs. In addition to this, the second goal of this dissertation was to extend persistence research by correlating the proportions of students' productive persistence with measures of grit. To accomplish these goals, the researcher conducted two

studies that answered the following research questions within an online learning platform, ASSISTments:

1. What are the key differences between productive and unproductive persistence in ASSISTments? (Study 1)
2. What is the relationship between productive persistence and grit in ASSISTments? (Study 2)

## CHAPTER II

### REVIEW OF LITERATURE

This chapter summarizes the literature on persistence, integrating key findings and conclusions from studies on grit and academic tenacity. While the benefits of grit have been extensively discussed in the literature, the potential costs of persisting in the face of challenges have not gained as much traction or interest. Within computer-based learning environments (CBLEs), one of the few studies that investigated this issue revealed that persistence may actually be detrimental to student learning. Therefore, one of the primary goals of this dissertation was to emphasize the need to investigate the key differences between productive and unproductive persistence, using educational data-mining techniques. Adopting an educational data-mining perspective affords the opportunity to investigate students' online interactions, revealing micro-level information about their learning struggles and processes. In understanding the trajectories of different modes of persistence, another major goal of this dissertation was to assert the importance of examining the relations between productive persistence and grit. Exploring these relations can help build a pedagogical framework that encourages and supports students' productive struggle.

#### **Benefits of Persistence**

Persisting in the face of setbacks or challenges plays an important role in learning across various academic domains. Peterson and Seligman (2004) regarded persistence as one human

character strength that determines each of our unique profiles, defining it as “voluntary continuation of goal-directed action in spite of obstacles, difficulties, or discouragement” (p. 229). Based on this definition, time alone is not a good measure of persistence, as it also involves expending efforts on tasks that are not necessarily enjoyable (Peterson & Seligman, 2004). Prior studies have shown a relationship between students’ ability to persist on a task and their achievement (Boe, May, & Boruch, 2002; Deater-Deckard, Petrill, Thompson, & DeThorne, 2005). For instance, Andersson and Bergman (2011) found that persistence during early adolescence predicted academic achievement 3 years later, even when controlling for previous GPA scores. Specifically, an increase of one standard deviation in student persistence was associated with a 0.19 standard deviation increase in academic achievement 3 years later, corresponding to a learning gain from an additional year in high school (Andersson & Bergman, 2011).

Many other terms have been used to describe the same process of persisting toward long-term goals in the face of setbacks. Consistent with previous research on persistence, these similar modes of engagement have also demonstrated positive impacts on students’ academic success. For example, the burgeoning research on grit, defined as “perseverance and passion for long-term goals,” has shown this construct to be a strong indicator of motivation and long-term accomplishment (Duckworth et al., 2007, p. 1087). For example, adults with higher grit scores attained higher levels of education and earned higher GPAs, as compared to adults with lower grit scores. Within an elite military academy, grit was also found to be the best performing predictor of summer retention among military cadets. Lastly, in a younger age group (i.e., 7-15 years old), grittier children ranked higher in a National Spelling Bee competition than less gritty

counterparts. In general, these findings provided empirical evidence that illustrate an association between high achievers and their willingness to persist consistently towards a specific objective.

Research has also shown that the concept of academic tenacity is a particularly important factor of students' academic performance. In one of the most widely cited papers related to the topic on persistence, Dweck, Walton, and Cohen (2011) defined academic tenacity as "mindsets and skills that allow students to look beyond short-term concerns to long-term or higher order-order goals, and withstand challenges and setbacks to persevere toward these goals" (p. 4). In this theoretical review, Dweck and colleagues asserted that motivational interventions, which teach that intelligence can be grown with effort, can boost academic performance. Specifically, previous studies have consistently shown that students who experience effort-oriented interventions obtain higher test scores than students who do not (Aronson, Fried, & Good, 2002; Blackwell, Trzesniewski, & Dweck, 2007; Good, Aronson, & Inzlicht, 2003). Drawing from interviews from prior literature, Dweck et al.'s (2011) review attributed these findings to changes in students' effort, with increased tenacity from learners who received the motivational intervention.

Because grit and academic tenacity are associated with gains in student outcomes, studies on these constructs have suggested a link between persistence and student achievement. In particular, these findings showed that there are strong benefits for persevering in terms of long-term achievements.

### **Limitations of Persistence**

Although persistence has been shown to be associated with student achievement, sustained effort in the face of challenge may not always be sufficient or useful for learning. In particular, persistence may be beneficial with attainable goals but detrimental when objectives

are unattainable. Early work by Janoff-Bulman and Brickman (1982) asserted that when the latter condition is experienced, there are graver negative consequences in persisting unproductively than in giving up too soon. Although withdrawing is typically seen as a negative response, it may actually be adaptive when it provides an opportunity to invest resources more effectively. For instance, Wrosch, Scheier, Miller, Schulz, and Carver (2003) indicated that giving up can lead participants to re-evaluate their aspirations and invest time on more manageable objectives—resulting in better psychological well-being. In contrast, persisting toward an unattainable goal may serve as a maladaptive response, due to resources being invested in a counterproductive fashion (Peterson & Seligman, 2004).

In line with this argument, Shechtman, DeBarger, Dornsife, Rosier, and Yarnall (2013) suggested that there may be potential costs in overemphasizing grit at home and in school. Currently, 200 KIPP charter schools across the United States promote grit as one of the character strengths essential for students' academic success. While character education can help drive achievement, too much emphasis on individual traits—such as grit—could potentially cause stress for students and impede their learning (Shechtman et al., 2013). For example, students may be blamed for lacking grit, even when giving up is potentially more strategic for their learning. Lucas, Gratch, Cheng, and Marsella (2014) demonstrated this possibility, finding that having more grit may be less helpful in achieving a goal. Lucas and colleagues designed an experiment that involved 37 difficult anagrams, including 16 items that were impossible to solve. The purpose of the anagram task was to solve as many items as possible within 20 minutes. To do well in this task and earn an entry to a \$100 lottery, participants would need to skip unsolvable items and move on to easier ones. Results showed that scores on the grit measure were negatively associated with the number of anagrams completed. This suggested that grittier

individuals may have spent more time solving more difficult items at the cost of lowering their performance. Thus, having more grit was actually detrimental in performing well on this task, suggesting that persistence may not be uniformly beneficial across challenging conditions.

Lastly, Beck and Gong (2013) also investigated the consequences of persistence in the context of Intelligent Tutoring Systems. In this study, the concept of unproductive persistence was referred to as “wheel-spinning,” defined as “a student who spends too much time struggling to learn a topic within achieving mastery” (p. 432). To put it another way, wheel-spinning describes the phenomenon of students who are given several opportunities to practice a skill but are unable to arrive at the correct solutions. Within the tutoring system, students who have not mastered the skills, despite 10 practice opportunities, are considered to be wheel-spinning. Using this operational definition, as much as 35% of students were found to wheel-spin across two tutoring systems (Beck & Gong, 2013). Given the considerable percentage of wheel-spinners, there is a need to minimize students’ unproductive persistence and guide them towards a more effective use of their learning opportunities.

While previous studies have shown learning benefits in persistence, other prior research has indicated otherwise. At times, sustained effort may be detrimental for students’ learning, especially when the amount of time spent struggling is not used effectively. As such, the path towards successful learning is not necessarily contingent on whether or not one persists on a task; rather, it may be more related to the ability to distinguish when persistence is productive or unproductive (Janoff-Bulman & Brickman, 1982). Unfortunately, as unproductive persistence and wheel-spinning have not been thoroughly studied, there is a gap in the literature investigating what the differences are between learners who persist in a productive or unproductive manner. As a result, it becomes pertinent for researchers to investigate the

differentiation between productive and unproductive persistence. Understanding these key differences is important in providing insight into students' struggles and how to best address challenges towards successful learning.

### **Persistence Towards Productivity**

There have also been attempts in recent research to examine how students' efforts can be productive for learning. For instance, Kapur (2014) compared posttest math performance between two types of teaching sequences: direct instruction and a contrasting method, where students engage in problem solving before being taught the concepts of the lesson. While students who first encountered the problem-solving phase initially struggled in finding the correct solutions, they eventually became successful in learning a new math topic. Kapur referred to this phenomenon as Productive Failure. In a sample of ninth-grade students, Kapur's results indicated that students who engaged in productive struggle performed significantly better on tests of conceptual understanding and transfer, compared to students receiving direct instruction. Similar results were found in a second study, where a productive struggle condition led to significantly higher levels of cognitive load and conceptual understanding than a direct instruction condition or a vicarious failure condition, where students evaluated previous (unsuccessful) solutions from their peers. These findings suggested that there are advantages in persisting despite greater cognitive load, especially when mental efforts are directed towards the production and exploration of math solutions.

In addition to this, Warshauer (2015) also investigated the role of productive struggle in learning among middle school students. Based on classroom observations, Warshauer developed a framework to categorize the different struggles that emerged as students engaged in tasks on proportional reasoning. The categories for the kinds of student struggles were as follows: *Get*

*started* involved confusion and uncertainty about the goal of the task; *Carry out process* occurred when students experienced an impasse, typically due to an inability to implement an algorithm or remember formulas; *Give mathematical explanation* entailed difficulty in explaining reasons behind solutions; and *Express misconception or error* occurred when students used misconceptions to carry out solutions to problems. Without the appropriate response from the teacher, each kind of struggle could result in an unproductive resolution, where students continued to struggle without achieving any progress. Based on this definition, Warshauer's results indicated that 18% of students' struggles were resolved in an unproductive fashion. Additionally, she found the same type of struggle can lead to different resolutions, depending on how a teacher responds to the student. For instance, in an activity that required the comparison of quantities between two different containers, students struggled to *Give mathematical explanations*. Productive resolutions for this type of struggle were achieved in different ways: one teacher helped a student by guiding him or her to use a more appropriate representation of the problem, while another teacher encouraged the student to reflect on his or her answer. These findings suggested that the path to a productive resolution appears to be more nuanced than associating one type of struggle to one particular teacher response. Varying responses can work for different students, as teachers have to take into account individual student factors (e.g., prior knowledge). In general, the findings from this study provided more information on the different kinds of struggles that can occur within middle school math classroom; however, more research is still needed to elucidate which factors contribute to students' productive struggle.

The work of Yeager, Muhich, Torres, and Asers (2012) provided further insight into this area of research by developing a framework, known as productive persistence, to help direct students towards academic success. Specifically, the goal of productive persistence is to develop

students' tenacity and their use of good learning strategies (i.e., study habits and skills). Preliminary findings on Carnegie's Pathways, which provide year-long math courses within college community settings, showed promising results in applying the productive persistence framework. Historically, only about 21% of students in the Pathways courses completed their math course requirements within a year of continuous enrollment. When the productive persistence framework was integrated, 56% of students fulfilled the math course requirements in a span of one semester—considerably increasing the previous success rate in a shorter timeframe (Strother, Van Campen, & Grunow, 2013). Although persistence and the effective use of learning strategies have each been previously studied, the combination of these factors presented an innovative and actionable framework to investigate how students can learn more productively. More importantly, the work on productive persistence has suggested there are primary drivers that push students towards academic success (e.g., academic mindsets). These primary drivers provide actionable information for educators to intervene and support student performance, extending the implications from prior research on productive struggle and productive failure. With early signs of success, the productive persistence framework can potentially inform interventions to help students persist and learn successfully.

### **Factors Impacting Productive Persistence**

To guide the design of effective interventions, it may be worthwhile to investigate which factors influence students' productive persistence. Previous research has revealed that non-cognitive factors, such as students' beliefs and learning strategies, play an important role in students' academic success. Given this association, Yeager, Bryk, Muhich, Hausman, and Morales (2013) asserted that the same non-cognitive factors linked to student achievement are also likely related to students' productive persistence. These non-cognitive individual factors

may include students' self-regulated learning and academic mindsets (Yeager et al., 2013) as well as their academic emotions (Pekrun & Linnenbrink-Garcia, 2012). Drawing from prior literature, the following section discusses each of these factors in detail.

### **Self-regulation of Learning Strategies**

Skills related to students' self-regulated learning may be an important contributor to their productive persistence. Aligned with the concept of self-reflective practice, previous research on social cognitive theory has focused on the importance of self-regulated learning, which is defined as “self-generated thoughts, feelings, and actions that are planned and cyclically adapted to the attainment of personal goals” (Zimmerman, 2000, p. 14). Using a three-phase model, Zimmerman and Campillo (2003) argued that self-regulation takes place in the following phases: the first phase, forethought, refers to the goal-setting processes and motivational beliefs that precede and prepare the stage for a learning effort; the second phase, performance, refers to processes related to self-control and self-observation that occur during a learning effort; the third phase, self-reflection, refers to processes related to self-judgment and self-reaction that occur after a learning effort. In particular, the third phase completes the cycle and influences the succeeding forethought phase—informing subsequent learning. As a whole, self-regulation entails a cyclical process, as learners modify current solutions based on feedback from previous performance.

For more than a decade, previous studies have shown significant associations between self-regulated learning and academic performance (Schoenfeld, 1992; Zimmerman & Martinez-Pons, 1988; Zimmerman & Schunk, 2001). More recent studies have indicated that self-regulated learning interventions related to goal-setting can potentially impact students' persistence. For instance, Duckworth et al. (2011) examined the effects of MCII, mental contrasting (i.e.,

elaboration of an ideal future and potential obstacles that impede on the goal) combined with implementation intentions (i.e., proposal of solutions for potential obstacles), on the completion of Preliminary SAT (PSAT) workbooks. Results from this randomized study showed that the students in the MCII intervention significantly completed more practice questions from the PSAT workbook than students in the control group, while controlling for PSAT course participation and gender. Similar findings were also found in the work of Gollwitzer, Oettingen, Kirby, Duckworth, and Mayer (2011), who compared the effects of two learning strategies—mental contrasting and only thinking of positive future outcomes—on the acquisition of foreign words in elementary school children. The mental contrasting group had higher performance on vocabulary quizzes than a positive-future only group. As a consequence of implementing a self-regulatory strategy such as mental contrasting, these positive outcomes could have likely stemmed from realizing that obstacles can be surpassed with sustained effort, encouraging the intervention group to persist in the completion of tasks (Gollwitzer et al., 2011). As a whole, these findings suggested that self-regulated learning strategies could serve as a potentially effective technique to deal with challenges and promote persistence in learning.

### **Academic Mindsets**

In addition to self-regulated learning, another factor that may impact the path towards productive persistence is the students' academic mindset. According to Dweck, Walton, and Cohen (2011), some students may possess a “fixed mindset,” viewing intelligence as an inherent, fixed quantity, whereas other students may possess a “growth mindset” and believe that intelligence is a malleable quantity that can be changed with hard work. These theories of intelligence have been shown to influence students' efforts in the face of academic challenges, as well as corresponding academic outcomes. For instance, in a longitudinal study on seventh-grade

students, Blackwell, Trzesniewski, and Dweck (2007) found that students who endorsed growth mindsets significantly reported higher grades over 2 years in high school than students who endorsed fixed mindsets, even when controlling for prior achievement scores. Their findings also revealed that effort beliefs (i.e., extent to which a student believes hard work will lead to positive outcomes) and positive strategies (i.e., extent to which a student would try harder in response to failure) mediated the relationship between academic mindsets and student achievement. In other words, students who viewed intelligence as a malleable quantity were also more likely to engage in effort-oriented strategies and to believe in the benefits of hard work, relative to students who viewed intelligence as a fixed quantity. These differential motivational patterns between varying academic mindsets resulted in differences in student performance, such that students with fixed mindsets and positive motivational beliefs about efforts and persistence obtained improvements in grades.

### **Academic Emotions**

Lastly, students' academic emotions may play a potentially important role in the incidence of productive persistence. Pekrun and Linnebrink-Garcia (2012) argued that academic emotions, such as boredom and enjoyment, are associated with student engagement in an activity. Students who typically persist on a task may also experience flow, defined as "a holistic sensation that people feel when they act with total involvement, when there is a balance between their skills and level of difficulty" (Csikszentmihalyi, 1975, p. 72). On the basis of flow theory, this balance appears to be unstable, such that a student's emotional state may likely shift towards boredom or anxiety. For instance, students may experience boredom when they find the task too easy, suggesting that the level of skill exceeds the level of difficulty. Conversely, students may also become anxious or frustrated when the level of difficulty is much higher relative to their

skill and capacity. These conditions illustrate that experiencing boredom and frustration could impact how students engage and persist on a task.

Other studies on affect have indicated that different levels of emotions can lead to varying responses to learning successes and failures. For instance, some students may experience moderate levels of frustration, “pleasurable frustrations,” in which challenges are difficult yet doable—motivating learners to persevere through hard problems (Gee, 2007). In contrast, prolonged or intense frustration during learning activities could potentially lead to detrimental learning effects. D’Mello and Graesser (2012) suggested that when students experience persistent frustration due to an unresolved impasse, they are more likely to become bored and disengaged from the task. Contrary to flow theory, however, San Pedro, Baker, Gowda, and Heffernan (2013) found that boredom was more prevalent when problems were overly difficult, not easy. In addition to this, frustration was found to be more common for easy than moderately challenging problems—a result that is difficult to explain using flow theory (San Pedro et al., 2013).

In general, these findings suggested a potential link between students’ emotions and their ability to sustain effort and engagement during learning tasks. Although the relationship between emotions and persistence has not been fully explored, it may be helpful to investigate the relations between these constructs. The ability to navigate through difficult emotions, such as frustration, can potentially help students persist and confront challenges in the face of adversity.

### **Studying Persistence Using Educational Data-Mining Methods**

While a handful of factors is potentially associated with productive persistence, these factors are likely difficult to study and address in traditional classrooms, where teachers need to manage individual student progress as well as the needs of the entire class. For this reason,

investigating productive persistence in CBLEs may be useful, given the opportunity to study learning processes across different types of learners in real time. In addition, CBLEs can provide a wealth of data about students' online behaviors, which can help contribute to a better understanding of their motivational and cognitive processes. As such, CBLEs provide an environment conducive for investigating which factors are most associated with persisting productively in the face of challenging problems.

With the growing interest in integrating computer-based learning into K-12 classrooms, the field of educational data mining (EDM) provides the means to analyze students' detailed log data and develop models of student attributes (e.g., student's current knowledge, motivation, and attitudes) that are potentially pertinent to learning (Baker & Yacef, 2009). For example, student modeling in previous EDM research has been used to infer and automatically predict student disengagement (Baker & Rossi, 2013) and affective states (Baker & Ocumpaugh, 2014). Modeling student characteristics plays an important role in personalizing computer-based learning, as it allows educational computer software to adapt automatically to students based on their individual differences (Baker & Yacef, 2009).

Considering the scarce literature investigating the differences between unproductive and unproductive persistence, the emergent EDM methods used within CBLEs provides a promising analytic approach for differentiating between these two types of persistence. Insights from fine-grained data (i.e., student's online interactions) can help advance pedagogical theory about the ways in which students persist when learning, informing instructional interventions within and beyond the context of CBLEs.

Understanding the divergent paths of persistence involves a variety of reasons for why a wide range of learners continue to persist without learning. Arguably, one potential reason is that

it may be difficult for teachers to determine which students are unproductively persisting and what interventions are appropriate to support them better. Another plausible reason entails students' lack of self-regulation, such that they fail to monitor which strategies lead to further confusion versus learning. Analytic approaches in EDM present potentially actionable information for encouraging productive persistence and reducing wheel-spinning or unproductive persistence. For example, the analysis of students' online interactions can help teachers identify struggling students and determine when to best intervene, encouraging instructional support at more targeted and opportune moments during learning (Baker & Yacef, 2009). Additionally, findings from students' online performance can potentially empower design researchers to develop learning environments that promote balance between persistence and the use of effective learning strategies.

### **Modeling Persistence in Intelligent Tutoring Systems**

There have been a few attempts to model student persistence within the context of CBLEs, such as Intelligent Tutoring Systems. As previously mentioned, Beck and Gong (2013) used students' log data in ASSISTments and Cognitive Tutor—both online tutoring systems—to model wheel-spinning. Specifically, their criteria for wheel-spinning were as follows: a mastery criterion, which entailed getting three correct problems in a row, and a time criterion, which referred to a threshold of 10 practice opportunities. While the threshold selected was somewhat arbitrary, Beck and Gong showed that the distribution of students who achieved mastery remained flat after encountering 10 problems. This suggested that students' proportions of mastery were unlikely to vary considerably with a higher threshold. Additionally, ASSISTments prevents students from making a large number of attempts in a short period of time, by locking out learners after 10 problems on a skill within a day. For these reasons, they argued that 10

practice opportunities served as a reasonable threshold for modeling wheel-spinning. According to this definition, results revealed a considerable percentage of wheel-spinning, with between 9.8% and 35% of students engaging in unproductive behavior across both systems.

Following this approach, Beck and Rodrigo (2014) used the same definition of wheel-spinning to model unproductive persistence in a sample of Philippine students who worked on the scatterplot generation and interpretation unit in Cognitive Tutor; that is, students were considered wheel-spinning if they encountered 10 problems without achieving mastery (i.e., getting three responses correct in a row). Their results indicated that wheel-spinning is also prevalent outside American classrooms, specifically in the Philippines. Their findings also suggested that wheel-spinning is associated with specific affective states, namely confusion and flow, but not with boredom. To explain this pattern of findings, Beck and Rodrigo argued that the phenomenon of wheel-spinning may be more related to affective factors that involve some cognitive reflection when learning content materials, rather than the lack of motivation to do the work. Given the smaller sample used in their study, further work is needed to explore fully the relations between unproductive persistence and affect, especially frustration, which has been previously linked to how students continue to engage in learning.

Using the same definition as wheel-spinning in Beck and Gong (2013), Gong and Beck (2015) built a logistic regression model to predict whether a student will wheel-spin or master a skill at each practice opportunity. They adopted this approach to track the likelihood of wheel-spinning over time using log data from two tutoring systems, with 96,919 problems attempted by 567 students in Cognitive Tutor and 133,061 problems attempted by 5,103 students in ASSISTments. They built additional features into their wheel-spinning detectors relevant to the following aspects: student in-tutor performance (e.g., number of prior problems solved correctly

on the first attempt); features argued to represent the seriousness of the learner's attitude (e.g., number of previous problems that the student solved correctly and responding more than one standard deviation faster than the mean response time for the current problem); and the number of prior problems practiced by the students for a given skill. Gong and Beck conducted a three-fold cross-validation at the student level, where the model was repeatedly built on two groups of students and tested on the remaining group. Results showed that as students encountered more practice opportunities without achieving mastery, they were more likely to wheel-spin. Additionally, this combination of aspects in a multi-feature model achieved high values of Area Under Curve (AUC) and Percent Correct in both the training and test groups. In general, these findings suggested that this model generalizes well to new and unseen students of Cognitive Tutor and ASSISTments.

Lastly, Matsuda, Chandrasekaran, and Stamper (2016) also developed a wheel-spinning detector from 122 students who used Cognitive Tutor Geometry. They built a neural network model to detect wheel-spinning as classified by expert coders, from data on students' response accuracy while learning a skill during the fifth through tenth practice opportunity. Matsuda and colleagues argued that the data were insufficient to identify wheel-spinning in student-skill pairs with fewer than five practice opportunities. Thus, only student-skill responses that had five or more practice opportunities were included in analysis, resulting in 2,883 student-skill sequences. As a first step, two human coders were asked to identify cases of wheel-spinning based on students' response data. Using this human coding system, a set of neural-network models were built to classify response sequences into wheel-spinning or non-wheel-spinning, out of all the students who had not mastered the skill. The detectors were developed using 10-fold cross-validation and assessed using model precision and recall. As Matsuda and colleagues were trying

to figure out the best number of practice opportunities to predict wheel-spinning, they built a separate model for each number of practice opportunities a student had encountered on a specific skill (i.e., from 5 to 10). Overall, each of these neural-network models achieved high recall but low precision, with the model for the eighth practice opportunity performing slightly better than the rest. These findings suggested that the neural-network detectors were able to detect wheel-spinning early in the learning process (i.e., as early as five practice opportunities), but with a high false-positive rate.

In sum, automatically detecting wheel-spinning can spare teachers the impossible task of assessing and monitoring each student's likelihood of struggling unproductively on a considerable number of problems. In general, modeling wheel-spinning offers the potential for scalability because researchers have previously built detectors of unproductive behaviors across many students in two different tutoring platforms. However, efforts to improve wheel-spinning modeling approaches have been made, no research has focused on the ability of these detectors to differentiate between productive and unproductive persistence. While previous studies based their detectors on whether or not a student reached 10 problems without achieving mastery (Beck & Gong, 2013; Beck & Rodrigo, 2014; Gong & Beck, 2015), this definition of wheel-spinning did not take into account whether the student ever reached mastery—in ASSISTments, correctly responding to three consecutive problems of the same skill following this threshold of 10 practice opportunities. This limitation raises questions about the validity of the models because they treated a student who mastered on the 11th problem the same as one who never mastered after 141 problems. Additionally, this wheel-spinning definition also did not take into consideration students who may have asked for external help or were lucky in getting three answers in a row correctly—thereby potentially overestimating the incidence of productive

persistence. Additional work should be done in refining the definitions of productive and unproductive persistence to help ensure the validity of these constructs.

### **Summary of the Literature Review**

Previous studies on grit and academic tenacity have demonstrated that persistence plays a significant role in supporting students' successful learning. However, despite the learning benefits associated with tenacity, sustained effort may not always lead towards academic success. For instance, persisting on a task may not be helpful if students are struggling fruitlessly, with little or no progress in their learning. Recent studies on wheel-spinning, where too much time is spent struggling without achieving mastery, have indicated that a considerable percentage of learners engage in this mode of unproductive persistence within CBLEs. This suggests that the key to successful learning is not necessarily contingent on continued effort, but also on the ability to determine when effort is not productive and when to change strategy. Unfortunately, because the concept of unproductive persistence has been largely understudied, no research has yet compared productive persistence to wheel-spinning or how to support students who engage in either mode of persistence.

Thus, it becomes increasingly important to differentiate between productive and unproductive persistence to help support struggling learners. Understanding the key differences between these modes of engagement can begin to fill the gap in the literature and help teachers identify and support students who wheel-spin, while encouraging students to continue persisting if that persistence is productive. Findings in this area can potentially provide insight into unbundling the concept of persistence by encouraging students to work hard when it is beneficial and support them when it is not.

While recent modeling approaches on wheel-spinning have shown considerable efforts in detecting unproductive persistence within Intelligent Tutoring Systems, further work is still needed to understand better how this construct correlates with more global measures of tenacity. To the best of the author's knowledge, no systematic research has examined the relationships between productive persistence and grit. Exploring these relations could enhance the literature on how the expanding research on grit can be best applied to students identified as productively persisting.

By investigating how these constructs relate to one another, we can better understand the different trajectories of student persistence and identify potential protective factors to support successful learning. The key next step is to differentiate unproductive persistence from productive persistence—towards developing motivational research-based practice that enhances productive struggle and minimizes wheel-spinning.

## CHAPTER III

### METHODOLOGY

#### **Modeling in the ASSISTments System**

The main objective of this dissertation was to differentiate between students' productive and unproductive persistence, towards understanding the relationship between productive struggle and grit. This chapter describes the ASSISTments tutoring system, which is the primary source of data for this dissertation. Specifically, the data used in this dissertation consisted of students' online log data recording their online interactions with ASSISTments and their online responses to the Short Grit Scale (Grit-S), which was embedded in the tutor. To address the first research question, Study 1 discusses how models of wheel-spinning were created to differentiate students who productively and unproductively persisted in the tutoring system. To address the second research question, Study 2 discusses how students' proportions of wheel-spinning and productive persistence were correlated with their grit survey scores.

#### **The ASSISTments System**

This section describes the ASSISTments tutoring system, with particular attention to two functions that are particularly relevant to this dissertation: Skill Builders and the Automatic Reassessment and Relearning System (ARRS). Detailed descriptions of ASSISTments' goals and functions are provided below.

## **Goals of ASSISTments**

ASSISTments has been primarily used for assessing and assisting learning in middle school mathematics. In 2014, about 600 teachers across 43 states implemented ASSISTments, with approximately 50,000 students using the system each school year. Consistent with its goal to give out reports on student learning and performance to teachers, ASSISTments provides formative assessments of students' acquisition of specific math skills. Additionally, as an authoring tool, it allows teachers to create their own math content or choose from a library of ASSISTments problem sets that align with U.S. Mathematics Core Standards. In particular, problem sets contain a list of problems that tackle one of the following broad math topics in a grade level: Grades 6 and 7 (Ratios and Proportional Relationship, Number System, Expressions and Equations, Geometry, and Statistics and Probability) and Grade 8 (Number System, Expression and Equations, Functions, Geometry, and Statistics and Probability).

## **Functions of ASSISTments**

The system also utilizes a mastery-based learning approach in one of its popular features, Skill Builders. In Skill Builders, students are given the opportunity to practice on related problems until they master the associated math skill. By default, a student is considered to have mastered a skill in ASSISTments if he or she is able to get three questions correctly in a row within a Skill Builder, as illustrated in Figure 1. While students who know the skill may reach this proficiency threshold quickly, those who are struggling to learn the skill will have to work through more problems.

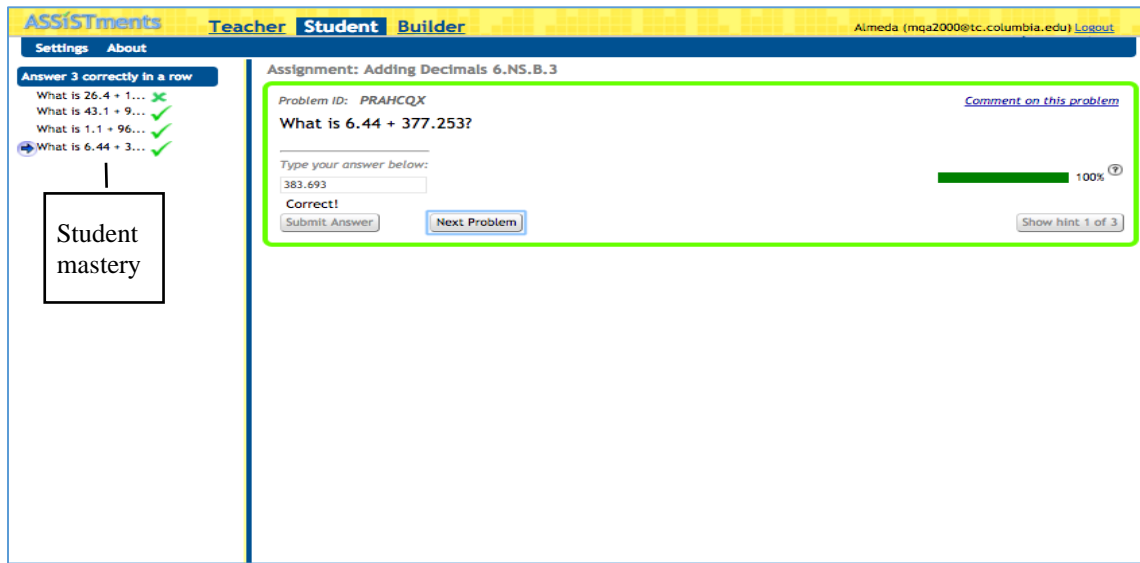


Figure 1. Mastery of a Skill Builder in ASSISTments

For each problem (i.e., original problem), students have the opportunity to access hints or scaffolding questions. Hints provide a sequence of clues that explain to students how to solve an original problem. As shown in Figure 2, the last hint in each hint sequence (called the bottom-out hint) provides students with the answer to the original problem.

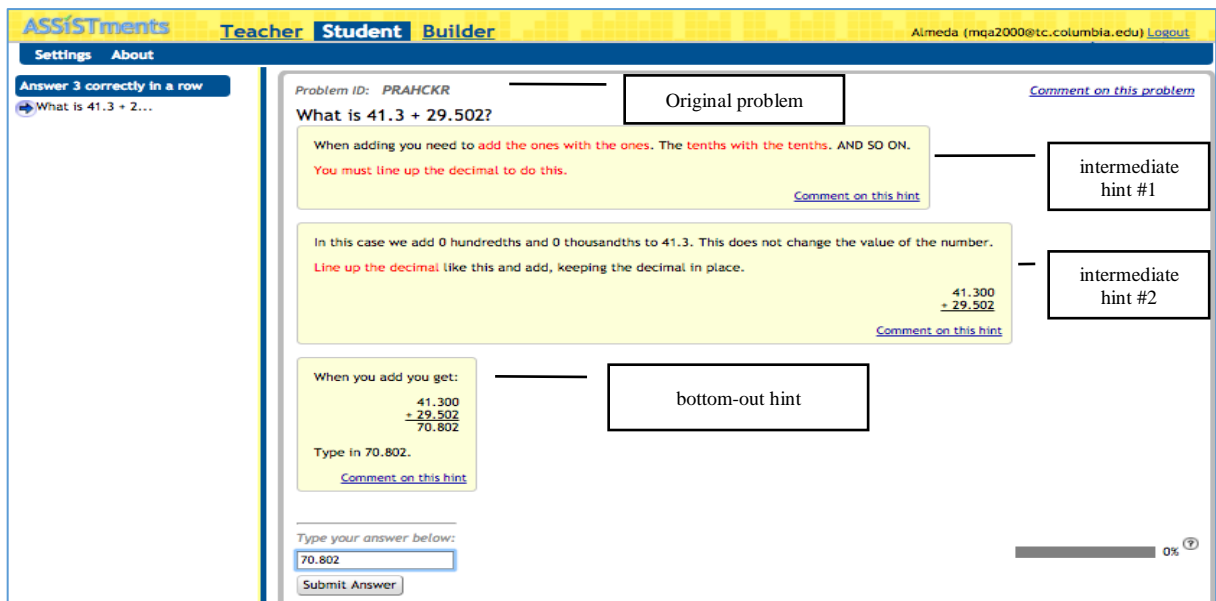


Figure 2. Types of hints in ASSISTments

ASSISTments also consists of scaffolding questions that break down the original problem into individual steps. These scaffolds are answered in a linear progression, where students must correctly answer the first scaffolding question to proceed to the next one. Once all scaffolding questions are completed, students may be prompted to answer the original question again.

Given the possibility that students may get lucky and correctly guess three questions in a row in a Skill Builder, or students may develop shallow knowledge that they fail to retain over time, the Automatic Reassessment and Relearning System (ARRS) was created to assess and aid in students' skill retention (Heffernan et al., 2012). Upon achieving mastery on a particular skill, students are given a single-item test on the same skill following a spaced schedule, with an increase in intervals between re-assessments (i.e., 7 days, 14 days, 30 days, and 60 days). If a student fails on any of these reassessments, he or she will be assigned relearning assignments to re-master the forgotten skill. Findings indicate that reassessment using ARRS leads to higher academic performance, with nearly half a grade higher for students in the ARRS conditions compared to non-ARRS counterparts (Heffernan et al., 2012).

As a whole, the features and functions of ASSISTments afford students the opportunity to practice as many problems as are needed to master a particular math skill. Such an opportunity would be difficult to recreate in a traditional classroom, where teachers have a limited time to go over a fixed number of problems. With the corpus of online data collected from students within ASSISTments, it becomes possible to examine closely and better understand how students persist through an ongoing series of problems. As a result, this structure of ASSISTments makes it an excellent platform to investigate student persistence. As indicated in previous wheel-spinning studies, one can easily imagine a student persevering through many problems but still not learning the corresponding skill. While little research has been conducted to reduce this

phenomenon, continued research on ASSISTments could potentially provide insight into addressing unproductive persistence within CBLEs.

### **Study 1 (Modeling Wheel-Spinning in ASSISTments: Methods and Results)**

To address the first research question, educational data-mining techniques were used to build models that differentiate between productive persistence and wheel-spinning in ASSISTments. The wheel-spinning model can indicate, after a student has completed 10 mathematics problems, whether or not he or she will eventually master and be able to demonstrate the skill in review later on. The section below details the modeling approach and what was learned from it.

#### **Participants**

Study 1 included data from the ASSISTments Skill Builders during the school year of 2014-2015. Data from a total of 26,497 students from a range of middle schools in the northeast United States were used in this set of analyses. The distribution of grade levels was as follows: 14% in K-5, 58% in Grades 6-8, and 25% in high school or higher education; the remaining 2% of students did not provide any grade-level information. Based on the students' user names, about 42% of students were inferred to be male, while 39% of students were inferred to be female. The gender of 19% of students could not be determined based on their user names. These 26,497 students attempted 940 Skill Builders problem sets over the course of the 2014 school year.

The objective of the analyses was to build machine learning models that differentiate between students who were persisting productively and those who were wheel-spinning during their work on the Skill Builders problem sets. Both of these groups are considered to be

persistent, but one group's persistence appears to produce positive results while the other group's persistence does not.

The author operationally defined students who are “persistent” to be students who worked on 10 or more problems within a single problem set. This cut-off was selected in part based on the design of the system, where students are stopped from working on the problem set after they have attempted 10 problems in a problem set within a day. For a student to continue onto the 11th problem in the problem set, he or she must return to the same problem set on a subsequent day.

From the set of persistent student-problem set pairs, instances of productive persistence and wheel-spinning were first identified. Given that the data were collected over the span of a whole year, there were instances in which students were able to attempt the same problem set more than once throughout the year. Students were thus able to achieve the mastery criteria of answering three problems correctly in a row—and then attempt the corresponding ARRS test—more than once throughout the year within a single problem set. For tractability, the measure of whether a student is “productive” or “wheel-spinning” was limited to the student's first ARRS test outcome and its corresponding set of three problems answered correctly in a row.

### **Wheel-Spinning Behavior**

**Operational definition.** As shown in Table 1, the operational definitions of “productive persistence” and “wheel-spinning” were based on two measures of learning: mastery (three correct problems in a row) and retention of knowledge (ARRS test). Specifically, students were classified as “productively persistent” if they answered three problems correctly in a row on or after the 10th problem in a problem set and passed the ARRS test. Conversely, students were classified as “unproductively persistent” if they answered three problems correctly in a row on or

Table 1

*Criteria of Productive Persistence and Wheel-Spinning,  
Using Performance Metrics in the Skill Builders System*

10 or More Problems	3 Correct in a Row (Mastery) on or After the 10th Problem	First ARRS Test	Definition
Yes	Yes	Passed	Productive persistence (this dissertation)
Yes	No	N/A	Wheel-spinning (this dissertation)
Yes	Yes	Failed	Wheel-spinning (Beck & Gong)
Yes	Any	Any	Neither Wheel-Spinning nor Productive Persistence
No	Any	Any	

after the 10th problem but did not pass the ARRS test. These students demonstrated correct performance immediately on a problem set but not in the longer term. Similarly, students who completed 10 problems but did not answer three problems correctly in a row on or after the 10th problem (and, as a result, never received an ARRS test) were also considered to be “unproductively persistent.”

Since the administration of the ARRS test could be customized by teachers, data from students who successfully answered three problems correctly in a row but were not given an ARRS test due to teacher customizations were considered to be missing data. It is worth noting that success on the ARRS test can be noisy—it is possible to get an incorrect answer by slipping or a correct answer by guessing (Baker, Corbett, & Aleven, 2008). As with all operational measures, this measure is therefore imperfect, but can still provide a basis for attempting to predict whether a student’s persistence will be productive.

In sum, the original dataset consisted of a total of 287,093 student-problem set pairs. Of this initial dataset, however, 211,612 student-problem set pairs were removed because students achieved three correct problems in a row at any point in time but did not attempt a corresponding ARRS test. Of the remaining 75,481 student-problem set pairs, only 8,948 student-problem set pairs were considered persistent, i.e., having attempted 10 or more problems in a problem set, regardless of mastery.

The final dataset used in these analyses thus consisted of 8,948 student-problem set pairs that were defined as persistent student-problem set pairs and used to build the final models. Within this final dataset, 2,093 student-problem set pairs were instances defined as productive persistence, while 6,855 student-problem set pairs were defined as wheel-spinning based on the criteria.

According to the study's definition, then, 9.08% of the total of 75,481 student-problem set pairs involved wheel-spinning, a lower proportion of wheel-spinning than reported in earlier studies that defined wheel-spinning as whether the student took a large number of problems to learn a skill (Beck & Gong, 2013). According to Beck and Gong's (2013) definition, where any student who completes 10 or more problems without getting three correct in a row and achieving mastery is wheel-spinning, 46.90% of the total 75,481 student-problem set pairs that had ARRS information would have been considered wheel-spinning. When considering the original dataset containing all student-problem set pairs, 12.33% of the original 287,093 student-problem set pairs would have been considered wheel-spinning, based on Beck and Gong's definition. Since many students do indeed eventually get three correct in a row and are then able to pass a retention test, one would argue that Beck and Gong's definition contains a great deal of productive persistence. Not all students who struggle are spinning their wheels.

However, even when using Beck and Gong's definition, there is still a substantial discrepancy in the proportion of wheel-spinning. A plausible explanation for the discrepancies in wheel-spinning proportions between this study and Beck and Gong's work lies in the choices of how the ASSISTments log data were aggregated. Whereas the present study focused on Skill Builders and the aggregation of student-problem set pairs, Beck and Gong's analysis included problem set types other than Skill Builders and the aggregation of student-skill pairs. Due to the design of Beck and Gong's analysis, many students might have seen a skill across a much wider range of time and contexts than in the present study.

**Feature engineering.** To differentiate wheel-spinning from productive persistence, a feature set previously created for another analysis (Baker, Goldstein, & Heffernan, 2011) was leveraged for Study 1. The initial core set of features consisted of 25 student actions and attributes within the ASSISTments Skill Builders platform that provided evidence on student persistence and learning. Following Beck and Gong's (2013) wheel-spinning threshold of 10 practice opportunities, the actions across the first 10 problems were used for feature engineering to build early predictors of wheel-spinning (i.e., whether they did not master the problem set or fail the ARRS test later on). A total of 125 problem-set-level features were then generated from these core features, based on their minimum, maximum, average, sum, and standard deviation across the problem set prior to the student having reached the 10-problem threshold for being persistent.

The features generated include attributes related to student use of hints in a problem set, number of student attempts, and features involving the time between these student actions. Examples of these attributes are listed below:

- number of last eight first responses that included a help request,
- number of last five actions that were a first response to a problem and incorrect,
- amount of time spent on the current problem so far,
- amount of time since the current problem set was last seen by the student,
- number of attempts made to solve each problem within a problem set, and
- number of bottom-out hints requested in the last eight problems.

Additionally, features related to the frequencies and occurrences of various student actions, and features related to the amount of time spent on/between certain student actions were created.

**Machine learning model.** After developing the feature set, a model was built to predict a binary variable: whether the student persisted productively or unproductively (i.e., wheel-spinning). The model was built using RapidMiner 5.3 data-mining software (Mierswa, Wurst, Klinkenberg, Scholz, & Euler, 2006) using the Weka J48 decision tree algorithm, which has been previously used in building detectors of engagement, affect, and meta-cognitive constructs.

Feature selection was conducted using an outer-loop forward selection process, attempting to determine the cross-validated goodness of specific sets of features. It is worth noting that conducting outer-loop forward selection tends to have an upward bias on model goodness relative to true training-test splits. The final features selected were from the set of attributes from the work of Baker, Goldstein, and Heffernan (2011).

The multi-feature model was validated using 10-fold student-level batch cross-validation, with AUC ROC as the primary measure of model goodness. The AUC ROC metric was computed using the A' implementation (rather than computing the integral of the area under the curve) to avoid having artificially high AUC ROC estimates because of having multiple data points with the same goodness—a feature of the integration-based estimates currently available

in most packages (Baker, 2015). A model with AUC ROC of 0.5 performs at chance, and a model with AUC ROC of 1.0 performs perfectly. It is worth noting that AUC ROC takes model confidence into consideration.

A precision-recall curve was created to identify the tradeoffs between precision and recall at different confidence thresholds of the model of unproductive persistence. Using a precision-recall curve facilitates understanding of how well the model functions across its predictive range, rather than at just one point, and can also be used to choose an optimal threshold for interventions with different costs or benefits (Davis & Goadrich, 2006). Precision represents the proportion of instances identified as wheel-spinning that are true instances of wheel-spinning, while recall represents the proportion of instances of true wheel-spinning that were identified as wheel-spinning. To put it another way, precision indicates how good the model is at avoiding false positives, while recall indicates how good the model is at avoiding false negatives. Together, precision and recall provide an indication of the model's balance between these two types of errors (Davis & Goadrich, 2006).

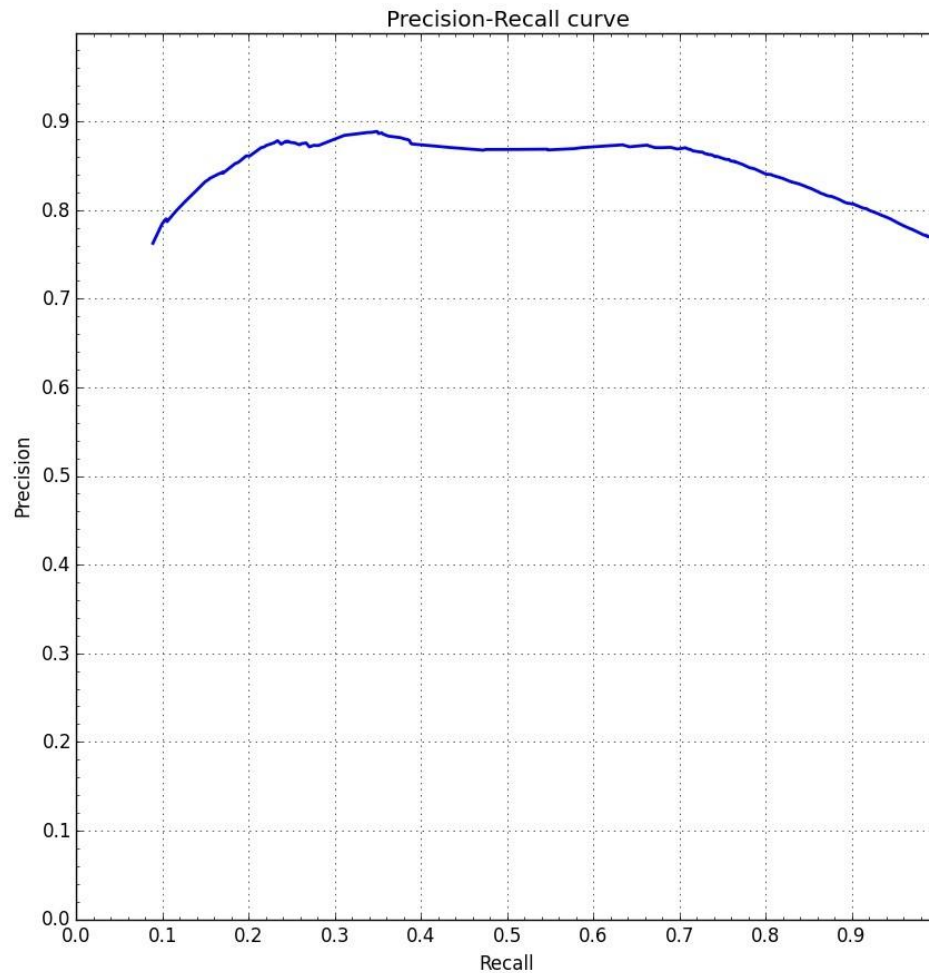
After creating the J48 decision tree with the initial set of features, further analysis on its structure was conducted. Features selected in the top three nodes in the J48 decision tree were of particular interest, as these play an important role in the tree's process of evaluating specific data points. Appendix A provides a complete summary of the very large decision tree structure.

## **Results**

### **Multi-feature Model**

The J48 model achieved an AUC value of 0.684. The standard error for the AUC metric for J48 was computed to be 0.004 (using the approach in Hanley & McNeil, 1982). This J48-based model used a combination of 15 features (see Appendix B for a description of each of

these features). The final decision tree generated using this algorithm had 95 leaf nodes and 189 decision nodes. The precision-recall curve generated for the J48 model of wheel-spinning is shown in Figure 3.



*Figure 3.* A precision-recall curve for the J48 model's predictions of wheel-spinning

From the precision-recall curve shown in Figure 3, there appears to be a clear tradeoff between precision and recall across thresholds, although the relationship is non-monotonic. The highest precision (nearly 0.90) is seen for relatively low values of recall (between 0.20 and 0.40). These high precision values remain stable as recall increases to 0.7, and only drop slightly

afterwards. Overall, it can be seen that the precision of the J48 model (the proportion of true positive predictions out of all positive predictions) remains high (above 0.75) even at high recall values—up to 100% recall.

When analyzing the features in the tree individually, only one feature from the J48 multi-feature model performed at a goodness substantially above chance—the standard deviation of the amount of time since the current problem set was last encountered by the student (std-timeSince Skill: AUC = 0.554). As the 15-feature model achieved considerably better predictive performance than the performance of each single feature, this suggests that it is in the interaction of these features that wheel-spinning and productive persistence can be differentiated.

**Top three features in the J48 decision tree.** One can better understand the pattern of relationships that distinguish wheel-spinning from productive persistence by examining the top nodes of the J48 decision tree. This tree only has leaves on one side of the first- and second-level nodes. As such, the author focused on the features in the three top nodes of the tree, which form a single branch. The tree spreads out below these levels. These features are as follows:

- minimum number of hints requested in any problem in the problem set (min-hintTotal),
- maximum number of bottom-out hints requested in the last eight problems (max-frPast8BottomOut), and
- standard deviation of the amount of time since the current problem set was last seen by the student (std-timeSinceSkill).

The second and third of these features deserve a bit of further examination. Both of these features only include actions in the current problem set—neither feature cuts across problem sets. Max-frPast8Bottomout refers to the maximum number of bottom-out hints requested in the past eight problems within the sequence of 10 problems in this problem set. Since the value of

max-frPast8BottomOut refers only to the number of bottom out hints in the most recent eight problems within a sequence of 10, a value of 1 could indicate that a student requests a bottom-out hint at any point within the sequence of 10 problems. It could also indicate more than one bottom-out hint requested in all 10 problems if the student requests one bottom-out hint at the beginning and one more at the end of the 10-problem sequence. In this case, the value of 1 will be obtained by the sequence {1, 0, 0, 0, 0, 0, 0, 0, 0, 1}, where 1 refers to a problem in which a student requested a bottom-out, while 0 refers to when he or she does not.

The std-timeSinceSkill variable gives the amount of variation in how much time the student took between each problem encountered in this problem set. A low value for std-timeSinceSkill indicates that the amount of time spent between problems in this problem set was relatively similar. For example, a student who attempted two to three problems in the same problem set every day over four consecutive days before achieving mastery and attempting the corresponding ARRS test would have a relatively low std-timeSinceSkill value, compared to other students in the dataset. In comparison, a student who attempted the maximum 10 problems allowed in a day and then only returned after 4 weeks to re-attempt the problem set before achieving mastery and trying the corresponding ARRS test would have a much higher std-timeSinceSkill value.

Based on analyses of the top three nodes described above, a considerable proportion of the data is explained by two feature combinations of these three features. Both feature combinations were associated with high probabilities of wheel-spinning, indicating that there are two distinct types of students who are likely to persist unproductively in the tutoring system.

The first feature combination indicates that students were likely to be wheel-spinning when they did not request any hints in at least one problem within the problem set but requested

more bottom-out hints in the last eight problems within the sequence of 10. 89.07% of the instances (2,544 out of 2,856 student-problem set pairs) (see Table 2).

Table 2

*Relationships Between the First Combination of Features at the Top of the Tree and Wheel-Spinning Within the J48 Model*

Selected Features	Feature Descriptions	Likely to Be Wheel-Spinning When
min-hintTotal	Minimum number of hints requested in any problem in the problem set	= 0
max-frPast8BottomOut	Maximum number of bottom-out hints requested in the last eight problems within the current problem set	> 1

While the first feature combination indicates that more bottom-out hint requests were associated with more wheel-spinning, wheel-spinning could still occur with fewer bottom-out hint requests. The second feature combination indicates that, even when the maximum number of bottom-out hint requests was 1 or 0, students could still wheel-spin if they did not request any hints in at least one problem and the standard deviation value of the amount of time since the current problem set was last seen was less than or equal to 2.53 days. The probability of wheel-spinning for the second feature combination was lower than for the first feature combination. A total of 77.85% of the instances (5,027 out of 6,457 student-problem set pairs with this feature combination) were labeled as wheel-spinning (see Table 3).

Table 3

*Relationships Between the Second Combination of Features at the Top of the Tree and Wheel-Spinning Within the J48 Model*

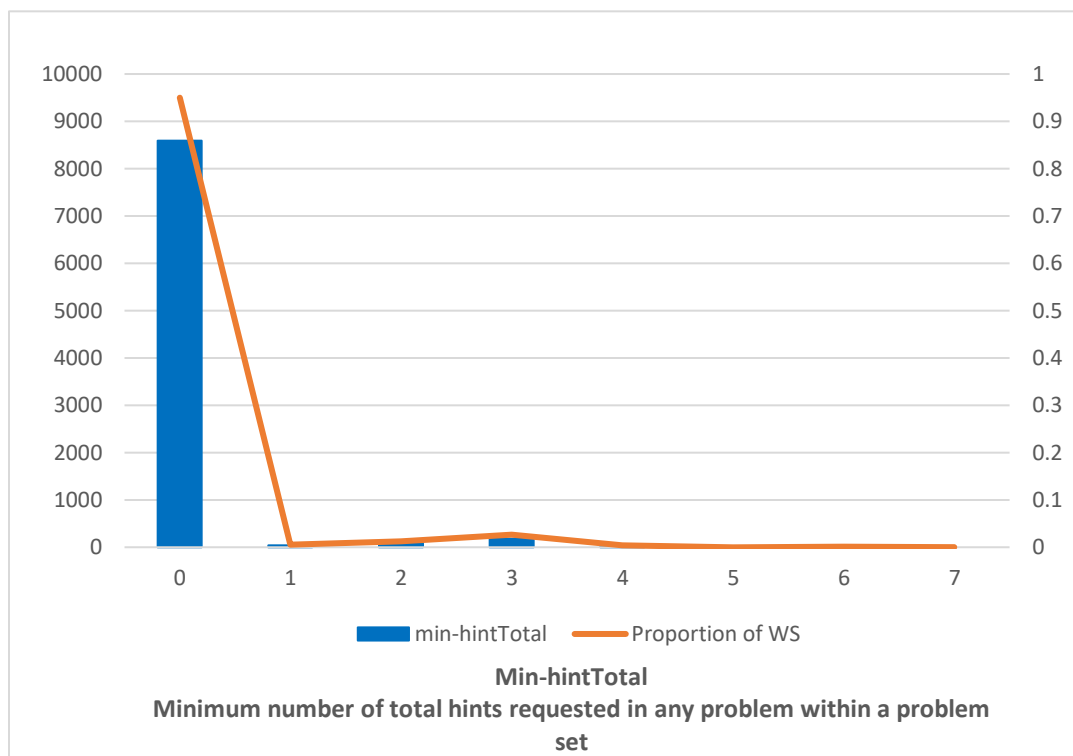
Selected Features	Feature Descriptions	Likely to Be Wheel-Spinning When
min-hintTotal	Minimum number of hints requested in any problem in the problem set	= 0
max-frPast8BottomOut	Maximum number of bottom-out hints requested in the last 8 problems within the current problem set	$\leq 1$
std-timeSinceSkill	Standard deviation of the amount of time since the current problem set was last seen by the student	$\leq 2.53$ days

These three selected features (min-hintTotal, max-frPast8bottomOut, and std-timesinceSkill) and their relationships with student wheel-spinning in Skill Builders are discussed in greater detail in the next section.

**Exploring model features.** To gain a better understanding of the range of values for the features at the top nodes of the prediction model, the author built histograms of min-hintTotal, max-frPast8BottomOut, and std-timeSinceSkill. Corresponding proportions of student wheel-spinning across the value ranges for these features were also created (Figures 4, 5, and 6, respectively).

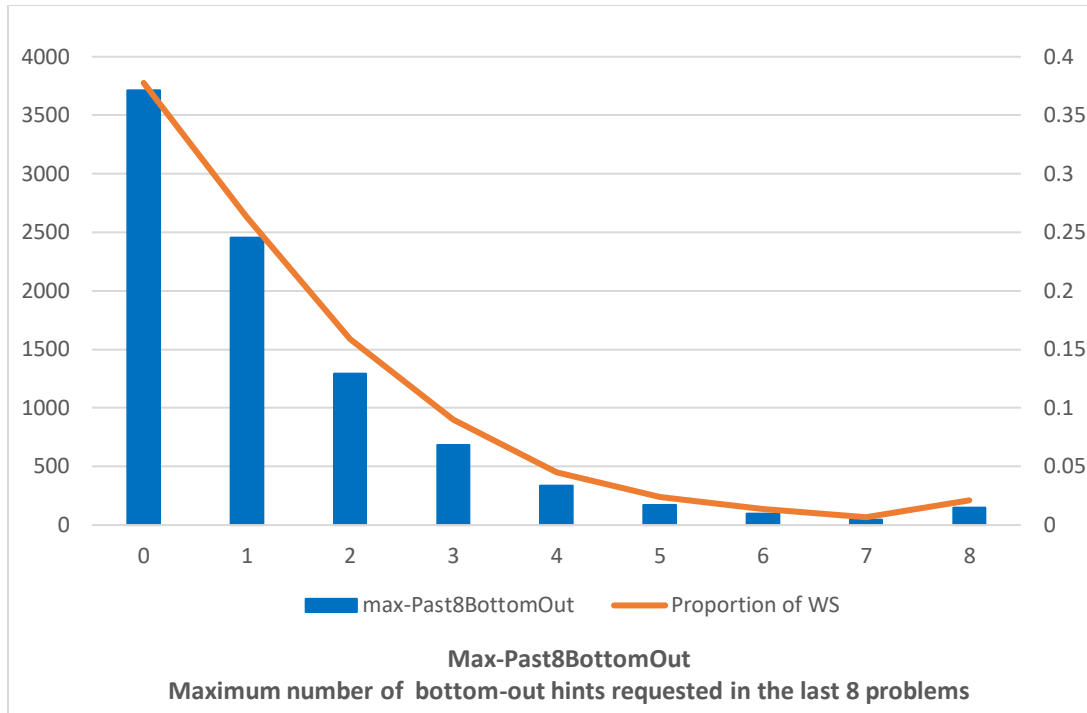
In the case of min-hintTotal, Figure 4 shows a unimodal distribution. In general, most students did not request any hints in at least one problem. The corresponding proportion of student-problem set pairs that were wheel-spinning based on the operational definition for each range of values in min-hintTotal is represented by the line graph. Here, one can see that the

proportion of student wheel-spinning is much lower if the minimum number of hints requested in any problem increases beyond 1.



*Figure 4.* Showing (a) Histogram of the minimum number of hints requested in any problem across student-problem set pairs (min-hintTotal) and (b) Proportion of student-problem set pairs identified as wheel-spinning for each min-hintTotal range of values

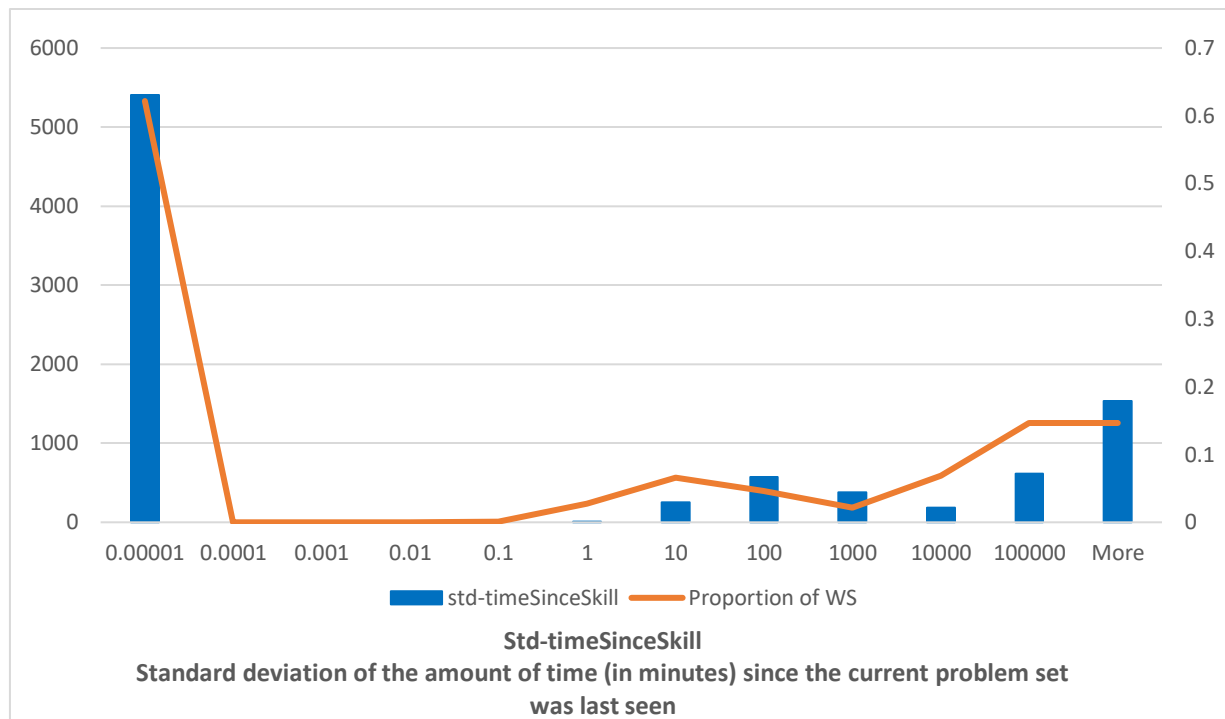
As shown in the histogram for max-past8BottomOut in Figure 5, the distribution for this feature is relatively skewed to the right. Most students did not make any bottom-out hint requests in the last eight problems within the sequence of 10 problems. The second most frequent number of bottom-out hint requests was 1, with 2 as the third most frequent. The proportion of students who were wheel-spinning across the values of bottom-out hints requested is represented with a line graph. This line graph shows a steady decrease in the proportion of student wheel-spinning as the maximum number of bottom-out hints requested in the past eight problems increases.



*Figure 5.* Showing (a) Histogram of the maximum number of bottom-out hints requested in the last eight problems across student-problem set pairs (max-past8BottomOut) and (b) Proportion of student-problem set pairs identified as wheel-spinning for each max-past8BottomOut range of values

From Figure 6, the most frequent range of standard deviation values for the amount of time between problems of the same skill is 0; this means that the student started the next problem immediately after completing the previous problem in the system, for every case within the problem set. However, many student-problem set pairs have standard deviation values of 100,000 minutes or more (approximately 69 days or more). The line graph shows that the proportion of students defined as wheel-spinning is highest at very low standard deviation values of the length of time spent between encountering problems in a problem set. As previously mentioned, low standard deviation values typically represent situations where a student experiences shorter periods of time between solving problems, while high standard deviation values typically represent situations where a student experiences longer periods of time between problem

attempts of the same problem set. As such, this finding implies that students who have shorter delays between solving problems are more likely to be wheel-spinning.



*Figure 6.* Showing (a) Histogram of the standard deviation for the amount of time since the current set was last seen across student problem set pairs (std-timeSinceSkill) and (b) Proportion of student-problem set pairs identified as wheel-spinning for each max-past8BottomOut range of values

## Summary of Study 1 Results

In general, these findings indicated (unsurprisingly) that no single-feature model performed as well as the multi-feature model. The J48 multi-feature model was reasonably effective in predicting whether or not a student will engage in wheel-spinning, achieving an AUC ROC of 0.684. By comparison, the AUC performance of gaming detectors was found to be around 0.80 (Baker, Corbett, Koedinger & Roll, 2006; Pardos, Baker, San Pedro, Gowda, & Gowda, 2014), which is only slightly better than the current performance. Sensor-free affect

detectors of student affective states in ASSISTments, found to be effective in several-year longitudinal predictions (San Pedro et al., 2013), have had AUC values ranging between from 0.63-0.74 (Pardos et al., 2014), slightly lower than the current value of the multi-feature model. AUC values in the 0.74-0.81 range are used in medical decision making with real-world impact, such as the choice of which anti-retroviral therapy to use for HIV patients (Revell et al., 2013). As such, while there is considerable room for improvement in the model presented here, this is at the level of goodness where it can be used for basic research and intervention, given appropriate caution.

The findings presented here suggest that there is more predictive power in the combination of features, rather than in each of the features alone. Upon further analysis of the J48 decision tree, three selected features differentiated particularly well between students productively and unproductively persisting after 10 or more problems in the ASSISTments. The minimum number of hints requested in any problem (min-hintTotal) and the standard deviation of the amount of time since the problem set was last seen (std-timeSinceSkill) are both negatively associated with wheel-spinning. Specifically, more wheel-spinning is likely to occur when there are no hints requested in any problem and there is lower variation in the delay between solving problems of the same problem set. In contrast, a nuanced relationship was found between the use of bottom-out hints and wheel-spinning. More wheel-spinning is likely to occur when students request too few or too many bottom-out hints in the last eight problems of a problem set.

**Understanding the relationship between productive persistence and grit.** The findings of Study 1 extend prior wheel-spinning research by including a student's eventual mastery and success on a delayed ARRS test as part of the wheel-spinning criteria. As prior

wheel-spinning research focused on whether or not students had reached 10 problems without achieving mastery (which may have included some students who were productively persistent), the new definition used in Study 1 arguably provided a clearer differentiation between students who productively and unproductively persisted following 10 problems. When considering eventual mastery and ARRS performance, Study 1 revealed that approximately three times as many student-problem set pairs involved wheel-spinning as productive persistence in ASSISTments (9.08% vs 2.77%). This finding suggests that some intervention designed to minimize students' unproductive struggle is warranted to help better support struggling learners in ASSISTments.

While some work has correlated students' wheel-spinning to their affective states (Beck & Rodrigo, 2014), no research has examined the relationship of this construct to popular measures of tenacity, such as grit. Exploring how productive persistence relates to grit can potentially provide practical implications for supporting a variety of struggling learners across different learning environments. Determining which students are productively persisting could help indicate when research findings of how to enhance grit are most appropriate and helpful. Previous research has suggested that developing students' growth mindset can be an effective strategy to promote grit (Laursen, 2015). As such, students who are already being productive when they are persistent (but are not being persistent frequently enough) may likely stick with challenging problems if teachers emphasize and praise their efforts over ability. As further work is needed to understand how productive persistence relates to grit, the next section, Study 2, discusses the methods to investigate the correlation between the two constructs.

## **Study 2 (Correlating Productive Persistence and Grit: Methods)**

To address the second research question, the author examined the relationship between productive persistence and grit. Using the same definition of productive persistence as used in Study 1, the incidence of productive struggle was correlated with grit scores on the Short Grit Scale (Grit-S). Findings in this study indicated the extent to which productive persistence was associated with grit, providing potential implications for synthesizing persistence research and designing interventions focused on supporting productive struggle across various learners.

### **Participants**

The author leveraged log data from students who used ASSISTments during the middle school mathematics classes of schools in the United States throughout the school year of 2016-2017. Through the help of the Worcester Polytechnic Institute (WPI) research team, the author collected and analyzed survey data from middle school students who took the grit survey in ASSISTments from May to June 2017. To ensure that an adequate number of students would take the online grit survey, the survey was embedded in 2017 into the Skill Builders most commonly used from May to June of 2016. These Skill Builders are as follows: Grade 6 (“Finding the Surface Area of a Rectangular Prism” and “Finding the Mean, Median, Mode, or Range”); Grade 7 (“Probability of a Single Event” and “Combining like Terms”); and Grade 8 (“Finding the Volume of a Cylinder” and “Square Roots”).

A total of 486 students who took the grit survey and had ASSISTments log data during the academic school year 2016-2017 were included in Study 2. These 486 students attempted 230 Skill Builders, resulting in 10,772 unique student-problem set pairs.

Similar to Study 1, there were instances when teachers customized the administration of the ARRS test. Due to these customizations, the following were considered to be missing data:

instances when students successfully achieved mastery but were not given the ARRS test, and instances when students did not achieve mastery but were still given the ARRS test. Following these exclusion criteria, 6,568 student-problem set pairs were removed. A total of 4,204 student-problem set pairs, with 213 students who attempted 192 Skill Builders, were included in the Study 2 analyses.

## Measures

**Short Grit Scale (Grit-S).** Students' grit levels were measured using the Short Grit Scale (Grit-S) (Duckworth et al., 2007; Duckworth & Quinn, 2009), an eight-item scale that was embedded in the ASSISTments tutoring system. The Grit-S Scale was designed to capture two latent factors: *Consistency of Interest* (i.e., a student's passion over time) and *Perseverance of Effort* (i.e., a student's ability to sustain effort in the face of challenges). Students were given a 5-point response scale ranging from 1 (i.e., *not at all like me*) to 5 (i.e., *very much like me*). A complete version of Grit-S is shown in Appendix C. Example items included the following: "*New ideas and projects sometimes distract me from previous ones*" and "*I am a hard worker.*" Scores were averaged to obtain an index of students' grit level, with a maximum score of 5 (very gritty) and minimum score of 1 (not at all gritty).

Previous results have shown empirical evidence for the validity and reliability of Grit-S in a sample of young National Spelling Bee participants. For instance, Duckworth and Quinn (2009) revealed that the overall scale of Grit-S achieved high internal consistency ( $\alpha = 0.80$ ), with acceptable reliability for each of the factors: *Consistency of Interest*,  $\alpha = 0.69$ ; *Perseverance of Effort*,  $\alpha = 0.76$ . In terms of predictive validity, Grit-S scores were found to be a significant predictor of GPA scores 1 year later, while controlling for age among Grade 7-11 students.

In the present study, the overall scale of Grit-S achieved an internal consistency coefficient of 0.59. As for the subscales, *Consistency of Interest* achieved a coefficient of 0.74, while *Perseverance of Effort* achieved a coefficient of 0.72. As both factors achieved moderate internal consistency, the author correlated proportions of students' persistent behaviors to the scores of the overall scale as well as to the subscale scores on the *Interest* and *Effort* factors.

**Productive persistence and unproductive persistence.** The same operational definitions of productive and unproductive persistence used in Study 1 were applied in Study 2. From a total of 4,204 problem set pairs, the author identified 9.02% of these ( $n = 371$  student-problem set pairs) as persistent, defined previously as those who worked on 10 problems within a single problem set. Among these persistent students, 3.19% of student-problem set pairs ( $n = 134$ ) were productive: where the student answered three problems in a row on or after 10 problems in a problem set and passed the ARRS test of the same problem set. In other words, student-problem set pairs were considered as productively persistent if, after completing 10 or more problems, students reached mastery and succeeded on the delayed test. In contrast, student-problem set pairs were considered unproductively persistent if, after completing 10 or more problems, they were still unable to reach mastery or failed on the ARRS test. Following this definition, 5.83% of student-problem set pairs ( $n = 245$ ) were identified as unproductively persistent.

Once each persistent student-problem set pair was categorized as productive or unproductive, the ratio of how often the students' persistence was productive was computed. This produced an average proportion of productive persistence per student across problem sets, which was used for the main correlational analysis. The author also explored the relationship between grit and unproductive persistence based on prior definitions in Study 1, using the

average proportion of wheel-spinning per student across problem sets. For these metrics, students who were never persistent were dropped. The remaining 213 persistent students were included in the correlational analysis.

In addition, the author also computed the ratio of how often the student was productively persistent versus giving up prior to 10 problems without reaching mastery. For this metric, students who were never productively persistent or never gave up prior to 10 problems without reaching mastery were dropped. Only 39 students were included in this analysis.

## **Procedure**

### **Main analyses.**

***Student-level correlations.*** To examine the relationship between productive persistence and grit, the author used Spearman correlation to correlate students' proportions of persistent behaviors to their average scores on the overall Short Grit Scale (Grit-S), *Interest* and *Effort* subscales. A standard two-tailed t-test was used to determine if relationships were statistically significant. The same methods were used to explore how grit related to students' wheel-spinning and the ratio of how often they productively persisted versus gave up prior to 10 practice opportunities without achieving mastery. The Benjamini-Hochberg (1995) correction was used to control for multiple comparisons.

These main correlational analyses can help provide insight into how grit relates to different modes of persistence, specifically whether grit is more strongly associated with students' productive or unproductive struggle.

***Hierarchical linear modeling.*** In addition to the student-level correlations, the author also explored the existence of classroom-level variance in grit levels. Considering that distinct classrooms may have differing Skill Builders completion rates that could influence responses on

persistent measures, hierarchical linear modeling (HLM) addressed this issue by examining the role of classroom variables on grit levels. The existence of these classroom-level effects was evaluated by estimating the null model, where the variance in the grit survey responses was split into student- and classroom-level components. The intraclass correlation (ICC) indicated the extent to which differences in grit levels existed between classrooms. From the total of 213 persistent students, 26 classrooms were included in the HLM analysis.

**Secondary analyses.** As several factors may potentially contribute to the relationship between students' grit levels and their online persistent behaviors in ASSISTments, the author conducted a set of secondary analyses. Specifically, the author explored the role of the following factors on the different measures of persistence: time of year, students' mastery speed, difficulty of Skill Builders, and students' grade level.

**Time of year.** To better understand the progression of wheel-spinning and productive persistence over the course of a year, the author computed the average proportion of each mode of persistence for each month of the academic year (i.e., August 2016 to June 2017) across 4,204 student-problem set pairs.

**Mastery speed.** To account for how well each student performed on a specific math topic, the author correlated each student's mastery speed on a Skill Builder (i.e., the average number of problems to master a specific problem set) to his or her corresponding Grit-S scores. Due to the non-normal distributions of mastery speed, *Interest* and *Effort*, and the overall scale of Grit-S, Spearman correlation was used. Instances when students did not master a problem set were treated as missing data and excluded from consideration. As such, a total of 2,967 student-problem set pairs were included in this analysis.

***Difficulty of a Skill Builder.*** Considering that some math topics are likely easier to grasp than others, the difficulty level of different Skill Builders may influence students' frequency of wheel-spinning or productive persistence. To obtain categories representative of math topics with different difficulty levels, the author applied clustering based on the average mastery speed of each problem set in the current sample ( $n = 156$  Skill Builders). As this analysis involved the computation of mastery speed, instances when students never mastered a problem set were treated as missing data and excluded from consideration.

Clustering is defined as the process of grouping a set of objects, such that those belonging to the same cluster share similarities in their attributes. K-means, the most straightforward and popular clustering algorithm, was used because of its fast and easy implementation and analysis. For ease of interpretation, the number of clusters was prescribed in advance ( $K = 3$ ) to represent three levels of difficulty: sets of problems that are easy, moderately difficult, and hard.

Once clusters were defined, the author used a one-way ANOVA to investigate the effect of Skill Builders difficulty level (i.e., easy, moderately difficult, and hard) on students' proportions of wheel-spinning and productive persistence. Log data from 156 distinct Skill Builders were included in the ANOVA analysis.

***Grade level.*** Student grade level may also play a role in influencing grit levels and persistent behaviors. To investigate this issue, the author conducted a one-way ANOVA to analyze the effect of student grade level on grit levels. From the total of 4,204 student-problem set pairs, there were 433 unique students. Of these 433 students, the grade-level information was as follows: 1 student in Grade 5, 145 students in Grade 6, 145 students in Grade 7, 120 students in Grade 8, 1 student in Grade 9, 1 student in Grade 10, and 20 had missing grade-level information. As the majority of the students in this sample were in middle school, 410 students in

Grades 6 to 8 were included in the one-way ANOVA analysis. Students who were not in middle school or had missing grade-level information were dropped from analysis.

## Results

### Main results

***Student-level correlations.*** In addressing the main research question, the author summarized the results of the Spearman correlations between each persistent behavior and the Grit-S scores in Table 4. Surprisingly, there was no significant relationship between students' proportions of productive persistence and their grit levels (*Interest*,  $r_s = 0.064$ ,  $t(211) = 0.93$ ,  $p = 0.351$ ; *Effort*,  $r_s = 0.023$ ,  $t(211) = 0.33$ ,  $p = 0.740$ ; Grit-S,  $r_s = 0.059$ ,  $t(211) = 0.86$ ,  $p = 0.395$ ). Similar to this, students' wheel-spinning proportions were not significantly related to their grit levels (*Interest*,  $r_s = -0.055$ ,  $t(211) = -0.80$ ,  $p = 0.426$ ; *Effort*,  $r_s = -0.088$ ,  $t(211) = -1.28$ ,  $p = 0.203$ ; Grit-S,  $r_s = -0.106$ ,  $t(211) = -1.55$ ,  $p = 0.124$ ). In addition, students' ratios of productive persistence versus giving up prior to 10 problems were also not significantly associated with their grit levels (*Interest*,  $r_s = -0.209$ ,  $t(37) = -1.30$ ,  $p = 0.202$ ; *Effort*,  $r_s = 0.189$ ,  $t(37) = 1.17$ ,  $p = 0.248$ ; Grit-S,  $r_s = -0.002$ ,  $t(37) = -0.01$ ,  $p = 0.989$ ). Overall, the author did not find evidence to suggest that grit manifested behaviorally within ASSISTments. In other words, these results indicated that grit, a significant predictor of long-term outcomes, was not significantly related to persistent behaviors linked with short-term outcomes in an online learning platform.

***Hierarchical linear modeling.*** To examine whether there was systematic variance in grit levels between classrooms, the author conducted HLM on the scores for *Interest*, *Effort*, and Grit-S. Based on the intraclass correlations on *Interest* (ICC = 0.02), *Effort* (ICC = 0.03), and Grit-S (ICC = 0.03), the proportion of variance in grit explained by classrooms was extremely low. According to Heck and Thomas (2015), most researchers used an ICC value of 0.05 as the

cut-off for conducting multi-level analysis. As such, these findings suggested that there was no systematic classroom variance for grit levels, suggesting that HLM was likely unnecessary.

Table 4

*Spearman Correlation of Persistent Behaviors to Grit-S Scores, Across Students*

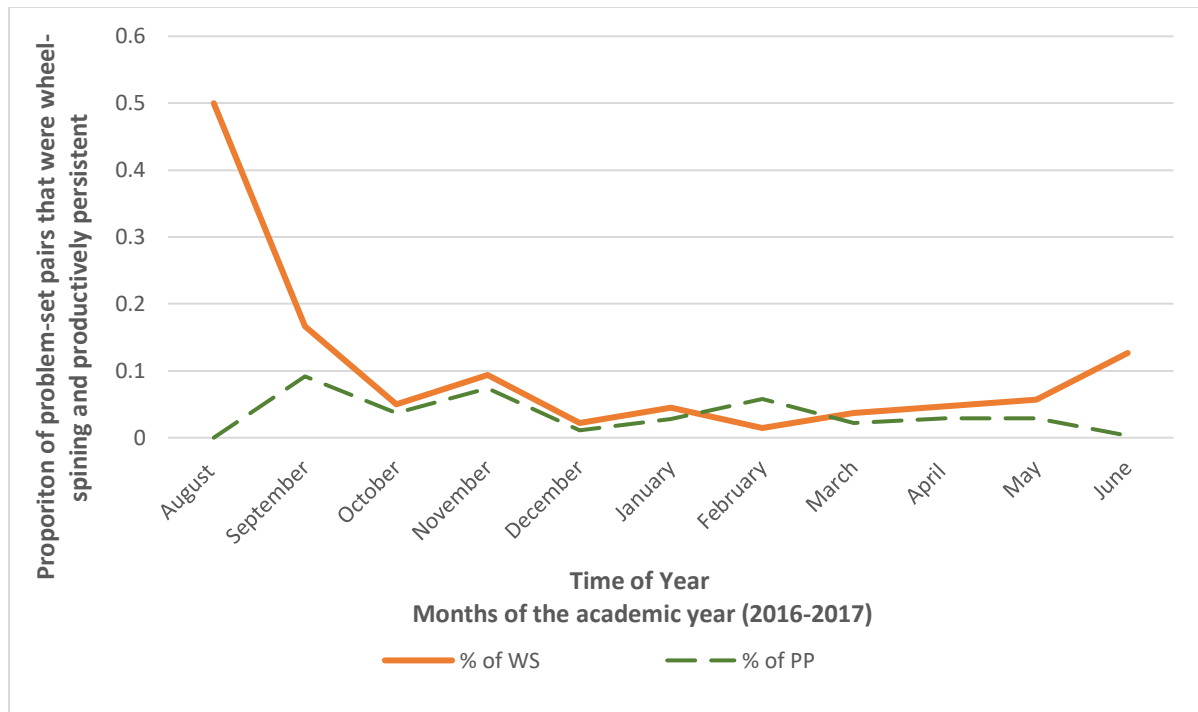
Persistent Behavior	Spearman Rho ( $r_s$ )	p-value	Adjusted alpha
Productive Persistence			
<i>Consistency of Interest</i>	0.064	0.351	0.028
<i>Perseverance of Effort</i>	0.023	0.740	0.044
Grit-S	0.059	0.395	0.033
Wheel-spinning			
<i>Consistency of Interest</i>	-0.055	0.426	0.039
<i>Perseverance of Effort</i>	-0.088	0.203	0.017
Grit-S	-0.106	0.124	0.006
Ratio of Productive Persistence versus Giving up			
<i>Consistency of Interest</i>	-0.209	0.202	0.011
<i>Perseverance of Effort</i>	0.189	0.248	0.022
Grit-S	-0.002	0.989	0.050

Note: \*marginally significant after using Benjamini-Hochberg correction

### **Secondary results.**

***Time of year.*** To examine temporal changes in students' persistent behaviors, Figure 7 presents a summary of students' proportions of wheel-spinning and productive persistence for each month, from August 2016-June 2017. Specifically, productive persistence remained consistently low throughout the academic year, with the lowest incidence found in August and

the highest incidence found in September. With regard to unproductive struggle, wheel-spinning was found to be highest in the beginning of the fall term, followed by a steep decline in October. Throughout the fall term (October-December) and spring term (January-May), the proportions of wheel-spinning were relatively low (0.10 or below), except for a slight increase in June.



*Figure 7. Proportions of productive and unproductive persistence for each month in the academic year (2016-2017)*

Figure 8 shows the frequency of student-problem set pairs throughout August 2016-June 2017. The lowest frequency occurred in August 2016 while the highest frequency occurred in May 2017.

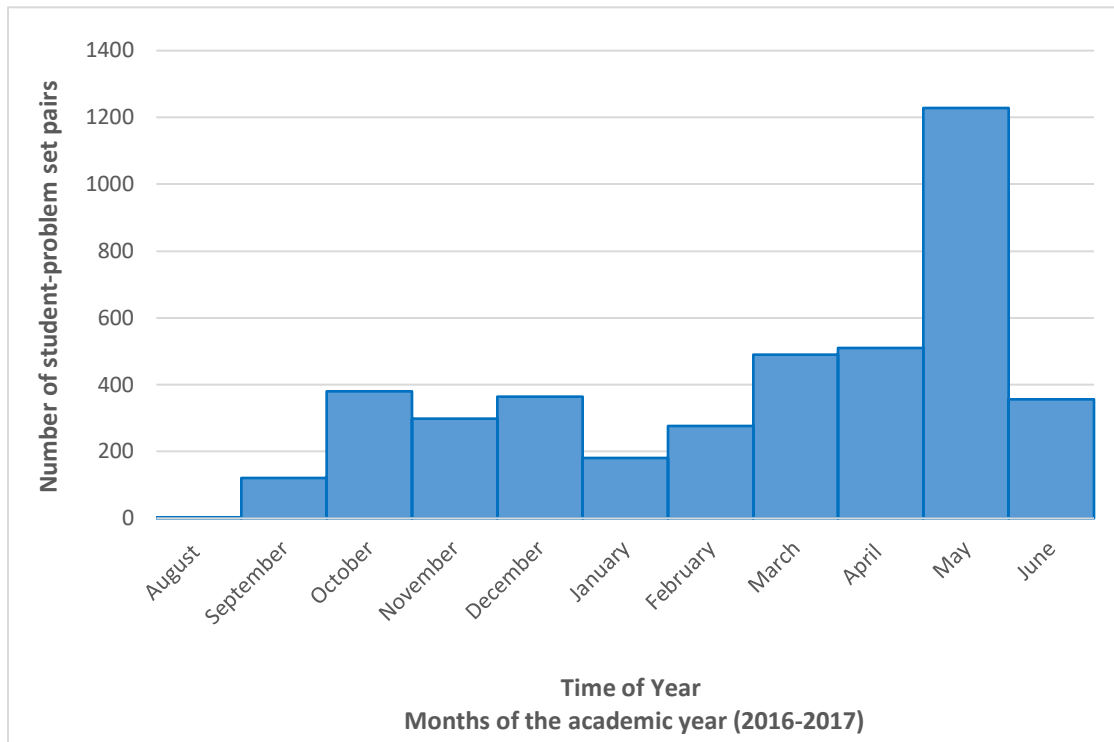


Figure 8. Frequency of problem set pairs for each month in the academic year (2016-2017)

**Mastery speed.** Table 5 presents the correlational results between the average number of problems needed for a student to master a specific problem set and his or her average Grit-S scores. Mastery speed was not significantly correlated with *Interest* ( $r_s = -0.037$ ,  $t(2965) = -2.02$ ,  $p = 0.045$ , adjusted alpha = 0.017) or *Effort* ( $r_s = -0.023$ ,  $t(2965) = -1.25$ ,  $p = 0.203$ ). There was a marginally significant and negative association between mastery speed and overall scores on Grit-S ( $r_s = -0.036$ ,  $t(2965) = -1.96$ ,  $p = 0.047$ , adjusted alpha = 0.033, marginal adjusted alpha = 0.066). However, it is important to note the strength of this correlation was very weak, indicating that the number of problems needed for a student to achieve mastery was not closely associated with his or her overall grit scores.

Table 5

*Spearman Correlation of Mastery Speed to Grit-S Scores, Across Student-problem Set Pairs*

Spearman Rho ( $r_s$ )	p-value	Adjusted alpha
<i>Consistency of Interest</i>		
$r_s = -0.037$	0.045	0.017
<i>Perseverance of Effort</i>		
$r_s = -0.023$	0.203	0.050
<i>Grit-S</i>		
$r_s = -0.036$	0.047	0.033*

Note: \*marginally significant after using Benjamini-Hochberg correction

***Difficulty of a Skill Builder.*** K-means clustering was conducted to obtain categories of difficulty levels across Skill Builders, using mastery speed as a proxy for difficulty (i.e., problem sets that take more problems for students to master can be viewed as being more difficult). With the prescribed number of clusters ( $K = 3$ ), the resultant clusters varied in size (cluster 1,  $n = 92$ ; cluster 2,  $n = 49$ ; cluster 3,  $n = 15$ ). Table 6 provides a summary of the descriptives for each cluster. Cluster 1 represents easier problems sets (mastered more quickly), Cluster 2 represents problem sets with moderate difficulty, and Cluster 3 represents harder problem sets (mastered more slowly).

Table 6

*Mean and Standard Deviation of the average mastery speed across Skill Builders*

Cluster	Mean	Standard Deviation
Cluster 1 (n = 92)	3.87	0.51
Cluster 2 (n = 49)	5.74	0.73
Cluster 3 (n = 15)	9.12	1.33

Using the groups from the cluster analysis, the author conducted a one-way ANOVA to test whether difficulty level (i.e., easy, moderately difficult, and hard Skill Builders) was significantly associated with students' average proportions of productive persistence and wheel-spinning. Levene's tests indicated that assumptions of homogeneity were not met for productive persistence and wheel-spinning. As such, the Welch F statistic was used for both variables. There was a significant effect of difficulty level on proportions of productive persistence, Welch  $F(2, 30.31) = 21.84, p < 0.0001, \eta^2 = 0.487$ . Similarly, there was a significant difference in proportions of wheel-spinning between difficulty levels of Skill Builders, Welch  $F(2, 29.54) = 11.57, p < 0.0001, \eta^2 = 0.293$ .

In general, hard Skill Builders had the highest proportions of productive persistence of all problem sets. Planned contrasts using Games-Howell procedures indicated that Skill Builders with hard levels ( $M = 0.241, SD = 0.176$ ) had higher proportions of productive persistence than Skill Builders with moderately difficult ( $M = 0.057, SD = 0.064$ ) and easy levels ( $M = 0.013, SD = 0.029$ , all  $ps < 0.01$ ). Additionally, proportions of productive persistence were significantly higher in moderately difficult Skill Builders than in easy Skill Builders,  $p < 0.0001$ .

In terms of unproductive struggle, hard Skill Builders also had the highest proportions of wheel-spinning of all problem sets. Planned contrasts using Games-Howell procedures revealed that Skill Builders with hard levels ( $M = 0.148$ ,  $SD = 0.176$ ) had significantly higher proportions of wheel-spinning than Skills Builders with easy levels ( $M = 0.004$ ,  $SD = 0.019$ ),  $p < 0.0001$ . However, there was no significance difference in wheel-spinning proportions between hard Skill Builders and moderately difficult Skill Builders ( $M = 0.039$ ,  $SD = 0.064$ ),  $p = 0.078$ . Additionally, moderately difficult Skill Builders had higher proportions of wheel-spinning than easy Skill Builders,  $p < 0.01$ .

While the greater proportions of unproductive struggle found in hard Skill Builders was expected, the similar finding found for productive persistence was surprising. One might have expected higher proportions of productive persistence with easy Skill Builders. A possible explanation for this finding is that problem sets that take longer to master are more likely to have students who achieve mastery and skill retention on or after 10 problems. In sum, these findings suggested that hard Skill builders are likely important to focus on, given the high incidences of both wheel-spinning and productive persistence found in these problem sets.

**Grade level.** A one-way ANOVA was conducted to test the statistical difference in average grit scores between middle school grade levels. Levene's tests revealed that the assumption of homogeneity was met for *Effort* but not for *Interest* and Grit-S. As a result, the Welch F statistic was used for the two latter variables. There was a significant effect of student grade level on the *Interest* subscale scores (Welch  $F(2, 267.88) = 4.61$ ,  $p < 0.05$ ,  $\eta^2 = 0.19$ ) and Grit-S scores (Welch  $F(2, 269.37) = 6.86$ ,  $p < 0.01$ ,  $\eta^2 = 0.027$ ). In contrast, there were no significant differences in the average scores on the *Effort* subscale between grade levels,  $F(2, 407) = 1.51$ ,  $p = 0.210$ ,  $\eta^2 = 0.008$ .

Planned contrasts using Games-Howell procedures revealed that seventh graders ( $M = 3.06$ ,  $SD = 0.90$ ) had significantly higher average grit scores than eighth graders ( $M = 2.73$ ,  $SD = 0.89$ ) on the *Interest* subscale,  $p < 0.01$ . However, there were no significant differences in *Interest* scores between seventh graders and sixth graders ( $M = 2.95$ ,  $SD = 1.08$ ),  $p = 0.610$ . In addition, sixth graders did not have significantly different *Interest* scores relative to eighth graders,  $p = 0.160$ .

Similarly, sixth graders ( $M = 3.31$ ,  $SD = 0.70$ ) and seventh graders ( $M = 3.36$ ,  $SD = 0.58$ ) had significantly higher average grit scores than eighth graders ( $M = 3.11$ ,  $SD = 0.53$ ) on the overall Grit-S scale, both  $ps > 0.05$ . In contrast, there were no significant differences in Grit-S scores between sixth and seventh graders,  $p = 0.800$ .

Overall, students' reports of effort were not significantly different between grade levels. However, eighth graders reported significantly less passion and less grit than sixth and seventh graders. This pattern of results was not consistent with Duckworth et al.'s (2007) assertion that grit tends to grow with age. While one would expect older middle school students to report more grit than their younger counterparts, the author found the opposite result in the current sample.

## **Summary of Study 2 Results**

In general, students' proportions of productive persistence and wheel-spinning were not significantly associated with their Grit-S scores. In other words, grit did not appear to be associated with persistent behaviors in an online platform. While grit was marginally correlated with mastery speed (the average number of problems to master a Skill Builder), the strength of this relationship was very weak. In contrast, grit was significantly associated with student grade level. Findings indicated that eighth graders reported significantly lower overall grit levels and effort, as compared to sixth and seventh graders.

When looking at factors related to persistent behaviors, the findings from Study 2 suggested that the difficulty levels of Skill Builders and the time of year may play a role in influencing students' productive and unproductive persistence. In particular, productive persistence was found to be relatively low throughout the entire academic term, 2016-2017. In contrast, high proportions of wheel-spinning were found in the beginning of the fall term (i.e., August to September) and at the end of the Spring term (i.e., June). Lastly, hard Skill Builders had the highest incidence of both productive and unproductive persistence. In other words, problems that took longer to master were more likely to have students both productively and unproductively struggling on or after 10 problems.

## CHAPTER IV

### DISCUSSION

#### **Persistence in Computer-based Learning Environments**

As persistence plays a significant role in learning, it is important for teachers and researchers to understand whether sustained effort is productive or unproductive. This topic is increasingly pertinent to investigate in the context of CBLEs, such as Intelligent Tutoring Systems, where students have been found to engage in wheel-spinning (Beck & Gong, 2013; Beck & Rodrigo, 2014; Gong & Beck, 2015; Matsuda et al., 2016). Results from both Study 1 and Study 2 suggested that specific factors, such as hint use and difficulty of a Skill Builder, contributed to students' persistent behaviors in ASSISTments. As both factors are closely tied to functions in ASSISTments and other related platforms, the findings on these factors can provide insight into how to better identify and support wheel-spinning students in CBLEs.

#### **Hint Use**

In Study 1, the findings from the J48 multi-feature model indicated that three key features—the minimum number of hints requested in any problem (min-hintTotal), the maximum number of bottom-out hints requested in the last eight problems (max-frPast8 BottomOut), and the standard deviation of the amount of time since the current problem set was last seen by the student (std-timeSinceSkill)—distinguish students who productively persisted from those who did wheel-spinning. Specifically, features related to hint usage (i.e., min-hintTotal

and max-frPast8BottomOut) were particularly important for distinguishing persistent behaviors in online learning systems.

First, results involving min-hintTotal indicated that not requesting any hints in at least one problem was related to greater wheel-spinning. In other words, students who did not always request hints in the first 10 problems were more likely to wheel-spin later on, suggesting that early hint usage is likely beneficial for students. This result aligned with what previous literature has shown about the relationship between students' help avoidance and their learning. In particular, Aleven et al. (2006) found that avoiding the use of help was negatively correlated with posttest scores, while controlling for pretest scores. Additionally, Almeda, Baker, and Corbett (2017) found that help avoidance had a generally stable negative correlation with aspects of robust learning, specifically transfer, retention, and preparation for learning. As such, students who wheel-spin and avoid hints could potentially be failing to realize they need support to learn the math content.

Second, the findings on max-frPast8BottomOut showed that the relationship between the number of bottom-out hints requested and wheel-spinning was more nuanced than min-hintTotal. In particular, the heavy use of bottom-out hints in the last eight problems was related to greater wheel-spinning. It is possible that some of these students were gaming the system (Baker, Corbett, Koedinger, & Wagner, 2004) and were less likely to read the intermediate hints (i.e., clues about how to solve the problem) because requesting bottom-out hints provided them with the answer (Aleven & Koedinger, 2000). Similarly, scarce use of bottom-out hint requests in the last eight problems was also associated with more wheel-spinning. It is possible that struggling students who infrequently use bottom-out hints are not checking the correct answers from these hints to learn the math content effectively. Given these findings, it is not hard to imagine that

students who avoid or abuse the bottom-out hint are more likely to struggle unproductively in ASSISTments.

Overall, `min-hintTotal` and `max-frPast8BottomOut` are features specific to student choices during problem solving in a tutoring system. It may be worthy of teachers' time to pay attention to students' use of intermediate and bottom-out hints in ASSISTments and other similar online learning platforms. Actionable information on the use of different types of hints is potentially helpful for teachers to identify students who are at risk of wheel-spinning and in need of additional support towards content mastery and skill retention.

In particular, these findings about hint use provide implications for developing randomized controlled trials that help students be more aware of their help-seeking behaviors in online learning systems. For instance, it may be useful to investigate whether online prompts, which remind students to request help within the first 10 problems, lead to reduced wheel-spinning in ASSISTments. Additionally, when students have reached the threshold for wheel-spinning at 10 practice opportunities, the tutoring system can quickly ask them to reflect on why they have not succeeded, and ask them what types of hints they would like to receive moving forward. Asking them to articulate why they are struggling can likely encourage self-reflection and allow teachers to create more effective hints that support student problem solving.

### **Difficulty of a Skill Builder**

In Study 2, the results indicated that the difficulty level of a Skill Builder had a significant effect on wheel-spinning and productive persistence. In particular, hard Skill Builders (i.e., problem sets that took a longer time to master) were associated with the highest proportions of unproductive and productive persistence (see Appendix D for a list of the 15 Skill Builders with hard difficulty levels). These findings suggested that struggling on challenging Skill Builders or

math topics is critical for students' learning, as this could eventually lead to either wheel-spinning or unproductive persistence. In particular, 11 out of 15 Skill Builders classified as hard involved problem sets with word problems. These results were consistent with the preponderance of math research that stressed the difficulty of word problems in arithmetic (Cummins, Kintsch, Reusser, & Weimer, 1988; Kintsch & Greeno, 1985) and algebra (Nathan, Kintsch, & Young, 1992; Paige & Simon, 1966).

Previous research has suggested that the difficulty of solving word problems stems from learners having to engage in two problem-solving phases: the comprehension phase, which requires linguistic processing knowledge to process the text of the problem; and the solution phase, which requires the use of strategies to arrive at a solution to the problem (Koedinger & Nathan, 2004). Analysis of algebra word-problem solutions revealed that most students use the unwind strategy, where learners work backwards from a given result value and use arithmetic tasks to produce a start value (Koedinger & MacLaren, 2002). As students primarily use implicit mathematical strategies to solve word problems, researchers have suggested that student mathematical development can be supported by helping students see the connection between the problem solution and the equation that represents the situation of the problem (i.e., the situation equation) (Nathan & Koedinger, 2000). This suggestion can be applied in ASSISTments by simultaneously presenting the problem solution and situation equation in the hints. Bridging implicit mathematical strategies into formal equations may potentially help students represent and solve challenging word problems in ASSISTments and other similar learning platforms.

### **Persistence Within and Beyond Computed-based Learning Environments**

The findings of this dissertation also indicated that specific factors are likely to be generalizable across different learning contexts. In particular, the findings of this dissertation

related to spaced practice are related to past findings in both online learning and traditional classrooms. This dissertation also has implications related to the relationship between grit and online persistence. Both of these factors have the potential to yield insight into how students persist within and beyond computer-based environments.

### **Spaced Practice**

Based on Study 1 findings, less variation in the number of days since the current problem set was last seen was associated with greater wheel-spinning. Taking a closer look at `std-timeSinceSkill`, it becomes apparent that a low value for this feature indicates consistently shorter periods between each problem attempt. For example, students who attempt a few problems each day for 4 consecutive days have relatively lower `std-timeSinceSkill` values and are more likely to wheel-spin than other students in the dataset. In contrast, wheel-spinning is less likely to occur when there are longer periods of time between problem solving of the same problem set. For instance, students who attempt several problems in a day, but only return after more than 1 month to reattempt more problems, are more likely to have higher values of `std-timeSinceSkill` and engage in less wheel-spinning. In general, these findings indicate that a greater degree of spaced practice (i.e., longer periods between problem solving) is associated with less wheel-spinning than a lower degree of spaced practice, with shorter periods between problem solving. Note that, in these examples, `std-timeSinceSkill` is likely contingent on teachers' choices of when to assign specific math skills to students. Differences in allocating student study time may have implications for shorter delays between problem solving and whether the result is poorer student mastery and retention of skills.

Prior research has shown the benefits of spacing on retention in math problem solving (Rohrer, Dedrick, & Burgess, 2014; Rohrer & Taylor, 2006). Other previous studies have

examined the effects of different spacing schedules on retrieval tasks. In a classic study by Landauer and Bjork (1978), they found that participants who encountered an expanded spacing schedule (i.e., increasing delays between rehearsals) had better retrieval performance compared to participants in a uniform spacing schedule (i.e., equal delays between rehearsals), suggesting evidence for a superior type of spaced practice. More recent work has investigated the effects of spaced practice in Duolingo, a free online language learning app. In particular, Settles and Meeder (2016) found that using a model of spaced practice led to improved recall for learners using the Duolingo platform. By comparing algorithms on spaced practice, Settles and Meeder were able to optimize each user's practice schedule based on content that is likely to decay from memory. As such, one can expect that practicing in a more spaced-out fashion will likely increase the effectiveness of persistence in ASSISTments and more traditional learning environments, given the present findings and the extensive literature on the benefits of spaced practice on retention.

### **Relationship Between Grit and Online Persistent Behaviors**

As shown in Study 2, students' levels of grit were not found to be associated with their online persistent behaviors (i.e., productive persistence and wheel-spinning) within ASSISTments. While a marginally significant correlation was found between mastery speed and grit, the small effect size of this relationship suggested that grit was not closely related to behaviors linked with mastery in an online learning platform. Overall, these findings suggested that personality questionnaires, predominantly studied in traditional classrooms, are not as predictive of student behaviors in blended learning systems. This pattern of findings may, in part, stem from the susceptibility of the self-report measures to reference bias, the tendency for different standards of reference points to influence survey responses (West et al., 2016). Contrary

to what was hypothesized, West and colleagues (2016) found that students in charter schools with higher standardized test scores surprisingly reported lower levels of grit than students in open-enrollment district schools. Researchers have asserted that the context of the highly academic and disciplinary charter schools led students to rate themselves more critically, thereby explaining the higher performance but lower grit scores relative to counterparts in the less demanding district schools.

Given this evidence on reference bias, the lack of a substantial relationship between grit and online persistent behaviors may likely be due to differing classroom practices between CBLEs and traditional learning environments. For the most part, traditional classrooms are characterized by teachers spending a significant amount of time lecturing to a large group of students. Often, in these traditional settings, the same material is used for instruction regardless of students' varying abilities. With the use of tutoring systems, students are given the opportunity for more individualized learning, emphasizing autonomy and self-learning. For instance, students who use ASSISTments can request help at any point during problem solving and practice a very large number of problems to master a topic if needed—a stark contrast to traditional classrooms where the entire class works on the same number of problems. Whereas teachers in traditional classrooms typically only provide comments for a few students, each student using ASSISTments and other related platforms are immediately given feedback (e.g., whether a response was correct or incorrect) for each problem-solving step. The affordances of practice and feedback in blended learning systems are likely to shape differences in standards by which online versus traditional learners assess persistence. Along with the results of this dissertation, previous studies have shown that self-reports of persistence are not predictive of learning outcomes in educational games (Ventura & Shute, 2013) and Intelligent Tutoring Systems (McCarthy et al., 2018),

suggesting that stealth assessments of students' persistent behaviors may be more useful than surveys in online environments. Thus, researchers and teachers should take caution in using self-reported measures to derive conclusions about learners in CBLEs, without fully understanding their sensitivity to differences in classroom practices.

### **Implications for Intervention**

Overall, the findings of this dissertation provide implications for teacher professional development in real-world settings. Considering the growing interest to assess and develop grit in the classroom, overemphasizing grittiness may have detrimental effects on students' psychological well-being. As previously discussed, teachers may attribute students' poor academic performance to their lack of grit, placing blame on students rather than focusing on critical supports potentially lacking in the environment (Shechtman et al., 2013). Additionally, encouraging grit in overly demanding learning contexts may have negative impacts on student persistence towards long-term goals. Students may demonstrate "fake grit" by responding on the scale based on what is expected of them, or persist only for the sake of immediate reward or punishment. As such, teachers should be aware of the costs of grit being misapplied in the classroom, and focus on creating conditions conducive for positive student motivation.

In addition to discussing these risks, it is increasingly important for teacher educators to incorporate the concept of unproductive persistence in professional development programs. Providing vignettes of wheel-spinning students can encourage teachers to reflect on how unproductive persistence manifests in the classroom. Furthermore, incorporating role play activities, where teachers demonstrate strategies to scaffold wheel-spinning students, provides an opportunity for fellow teachers and teacher educators to provide feedback and fine-tune interventions between the two modes of persistence. These professional development activities

can potentially train teachers to correctly differentiate wheel-spinning from productive persistence, and to effectively intervene during critical moments of student struggle.

In relation to this, the present findings also have implications for developing dashboards that help support teachers' needs, including the process of monitoring and decision-making in the classroom. Recent research has shown that teachers have intuitively thought about unproductive persistence in context of ITS in blended learning classrooms. When probing about teachers' wants and perceived challenges, teachers requested knowing when students were stuck as a superpower to help them during class sessions (Holstein, McLaren, and Aleven, 2017). Teachers believed that most of the students who raised their hands rarely needed support, while those who were actually struggling never requested help (Holstein, McLaren, and Aleven, 2017). With these findings in mind, the development of a real-time dashboard that provides teachers with information about students' wheel-spinning can help teachers prioritize scaffolding across students, by seeing which learner is in most of need of help at any given moment. Specifically, the dashboard can track and flag which students have reached the daily limit of 10 practice opportunities, so teachers can quickly interview them and understand why they are struggling. Drawing from Study 1 findings, the dashboard can also provide teachers with information about students' hint use that is likely related to their wheel-spinning behavior. In this way, teachers are made aware of the different profiles associated with wheel-spinning, and may be encouraged to think about how to best support each type of wheel-spinning student. For example, when a teacher is alerted about a wheel-spinning student who abuses hints, the teacher may approach the student and ask him or her to carefully read the intermediate hints before proceeding to the bottom-out hints. With further work on developing real-time dashboards and collecting qualitative data on wheel-spinning, one can imagine these tools recommending questions and strategies for teachers

to use, based on what students have done in the past 10 problems. As such, this line of research provides a new direction for supporting teachers – providing them with actionable information about student wheel-spinning that can potentially improve teacher interventions in the classroom.

### **Limitations and Future Work**

The dissertation is not without limitations. First, it is important to note that the aforementioned findings are correlational. Conducting small-scale randomized controlled trials (RCTs) in ASSISTments and other similar platforms will help establish which of these findings are causal, towards enhancing students' productive struggle and learning. While this work provides insight on the relationship between the number of different types of hints and wheel-spinning, future work should examine hint effectiveness and its relation to success and persistence within online learning platforms. Heiner, Beck, and Mostow (2004) compared the efficacy of several hint types across K-4 grade levels using a reading tutor designed to support children's oral reading. Similarly, researchers should also examine which types of hints lead to more productive persistence in ASSISTments, while controlling for varying skill levels of students. To improve hint design, it is also recommended that further research explore the effects of hint content (e.g., succinct versus verbose, or procedural versus conceptual) and the sequence of hints within the tutoring system. Specifically, the dashboard may eventually inform teachers when it is most helpful to present conceptual versus verbose hints during problem solving, in order to provide more differentiated support for different students at various types of problems.

Second, despite including indicators of eventual success on the problem set and retention, the threshold for defining wheel-spinning remains somewhat arbitrary. Although the cut-off of 10 problems maps to current practice within the ASSISTments system, other thresholds may better capture student persistence. As such, future work is needed to examine the threshold of wheel-

spinning further by creating a range of possible cut-offs for persistence (e.g., eight through 14 problems) and determining whether the predictors of wheel-spinning differ significantly across these thresholds. Given that the operationalization of persistence was largely based on the context of the ASSISTments platform, where the threshold of 10 problems is a meaningful transition point (since the system asks the learner to take a break if 10 problems are completed without mastery), studying other cut-offs for persistence may be particularly warranted when distinguishing wheel-spinning from productive persistence in other online platforms. Additionally, given that ASSISTments focuses on procedural math fluency, the bulk of the wheel-spinning research is closely tied to procedural math learning. Future work should also explore the definitions of wheel-spinning in the context of online systems that support the development of conceptual math understanding.

Furthermore, there may be other relevant indicators of student success, such as performance on standardized tests (cf. Pardos et al., 2014) or college readiness that can be incorporated into the wheel-spinning definition. Similar to the grit scale being a predictor of long-term outcomes, researchers should also look into developing long-term, stealth assessments of wheel-spinning that track unproductive persistence for years, rather than hours.

Lastly, the author asserts the need for future work to conduct a mixed-methods approach in assessing unproductive struggle. In addition to analyzing the log data within the tutoring system, it may be helpful for researchers to conduct moment-by-moment qualitative observations of when and how students wheel-spin to gain a better understanding of the deep motivations underlying unproductive persistence.

## **Conclusion**

This dissertation attempted to address two main research gaps in existing persistence literature. First, while most studies have investigated the benefits of grit, less work has examined how to identify and prevent unproductive persistence. Findings from Study 1 make a potentially important contribution in unbundling the concept of persistence, helping identify wheel-spinning and distinguishing this behavior from productive persistence that should be encouraged in online learning platforms. Second, no research has investigated the differences and similarities of these constructs with more global measures of persistence, such as grit. In attempting to obtain a broader understanding of related constructs of persistence, Study 2 results indicated a lack of significant relationship between grit scores and proportions of persistent behaviors in ASSISTments, which may be due to differing classroom practices between traditional classrooms and blended learning classrooms.

These findings are particularly relevant in light of the recent character education movement. With the excitement to incorporate non-cognitive skills into the classroom, teachers and principals may make hasty decisions about how to apply this research, integrating character assessments into high-stakes accountability systems. In fact, several California districts have proposed to implement this idea—an initiative that has been criticized as counter to the purpose of existing non-cognitive measures developed for self-discovery and research (Duckworth, 2016). While there is great promise in using Grit-S to evaluate persistence and cultivate character development, this tool is only meaningful if properly understood and utilized. As such, there is a critical need for further research to examine the costs of being gritty and acknowledge the limitations of using popular self-reported measures, such as Grit-S, in different learning environments. This dissertation represents a first step in this direction. By exploring how

persistence functions within and beyond CBLEs, the present work urges researchers and teachers to engage in deeper conversations about persistence in online learning platforms and traditional contexts—ultimately, to prevent oversimplifying the concept of grit and better guide a variety of learners towards productive motivation and improved learning.

## REFERENCES

- Aleven, V., & Koedinger, K. R. (2000, June). Limitations of student control: Do students know when they need help? In *International conference on intelligent tutoring systems* (pp. 292-303). Berlin, Heidelberg: Springer.
- Aleven, V., McLaren, B., Roll, I., & Koedinger, K. (2006). Toward meta-cognitive tutoring: A model of help seeking with a cognitive tutor. *International Journal of Artificial Intelligence in Education*, 16(2), 101-128.
- Almeda, M. V. Q., Baker, R. S., & Corbett, A. (2017, March). Help avoidance: When students should seek help, and the consequences of failing to do so. *Meeting of the Cognitive Science Society*, 2428, 2433.
- Andersson, H., & Bergman, L. R. (2011). The role of task persistence in young adolescence for successful educational and occupational attainment in middle adulthood. *Developmental Psychology*, 47(4), 950.
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology*, 38(2), 113-125.
- Baker, R. S. (2015). *Big data and education* (2nd ed.). New York, NY: Teachers College Press.
- Baker, R. S., Corbett, A. T., & Aleven, V. (2008). *Improving contextual models of guessing and slipping with a truncated training set*. Pittsburgh, PA: Carnegie Mellon University, Human-Computer Interaction Institute.
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Roll, I. (2006, June). Generalizing detection of gaming the system across a tutoring curriculum. In *International conference on intelligent tutoring systems* (pp. 402-411). Berlin, Heidelberg: Springer.
- Baker, R. S., Corbett, A. T., Koedinger, K. R., & Wagner, A. Z. (2004, April). Off-task behavior in the cognitive tutor classroom: When students game the system. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 383-390). New York, NY: Association for Computing Machinery.
- Baker, R. S., Goldstein, A. B., & Heffernan, N. T. (2011). Detecting learning moment-by-moment. *International Journal of Artificial Intelligence in Education*, 21(1-2), 5-25.
- Baker, R. S., & Rossi, L. M. (2013). Assessing the disengaged behaviors of learners. In R. Sottolare, A. Graesser, X. Hu., A. Olney, B. Nye, & A. Sinatra (Eds.), *Design recommendations for intelligent tutoring systems* (p. 153). Orlando, FL: U.S Army Research Laboratory.

- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM-Journal of Educational Data Mining*, 1(1), 3-17.
- Baker, R. S. J. D., & Ocumpaugh, J. (2014). Interaction-based affect detection in educational software. In R. A. Calvo, S. K. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford handbook of affective computing* (pp. 233-245). Oxford, UK: Oxford University Press.
- Beck, J. E., & Gong, Y. (2013, July). Wheel-spinning: Students who fail to master a skill. In *International conference on artificial intelligence in education* (pp. 431-440). Berlin, Heidelberg: Springer.
- Beck, J., & Rodrigo, M. M. T. (2014, June). Understanding wheel spinning in the context of affective factors. In *International conference on intelligent tutoring systems* (pp. 162-167). Berlin, Heidelberg: Springer.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)*, 289-300.
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78(1), 246-263.
- Boe, E. E., May, H., & Boruch, R. F. (2002). Student task persistence in the Third International Mathematics and Science Study: A major source of achievement differences at the national, classroom, and student levels. Retrieved from [www.eric.ed.gov](http://www.eric.ed.gov)
- Csikszentmihalyi, M. (1975). *Beyond boredom and anxiety*. San Francisco, CA: Jossey-Bass.
- Cummins, D. D., Kintsch, W., Reusser, K., & Weimer, R. (1988). The role of understanding in solving word problems. *Cognitive Psychology*, 20(4), 405-438.
- Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In *Proceedings of the 23rd International Conference on Machine Learning* (pp. 233-240). New York, NY: Association for Computing Machinery.
- Deater-Deckard, K., Petrill, S. A., Thompson, L. A., & DeThorne, L. S. (2005). A cross-sectional behavioral genetic analysis of task persistence in the transition to middle childhood. *Developmental Science*, 8(3), F21-F26.
- D'Mello, S., & Graesser, A. (2012). Dynamics of affective states during complex learning. *Learning and Instruction*, 22(2), 145-157.
- Duckworth, A. (2016). Don't grade schools on grit. *The New York Times*.

- Duckworth, A., & Gross, J. J. (2014). Self-control and grit: Related but separable determinants of success. *Current Directions in Psychological Science*, 23(5), 319-325.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087.
- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the Short Grit Scale (GRIT-S). *Journal of Personality Assessment*, 91(2), 166-174.
- Dweck, C., Walton, G. M., & Cohen, G. L. (2011). *Academic tenacity: Mindsets and skills that promote long-term learning*. Seattle, WA: Bill & Melinda Gates Foundation.
- Gee, J. P. (2007). *Good video games p good learning: Collected essays on video games, learning and literacy*. New York, NY: Peter Lang.
- Gollwitzer, A., Oettingen, G., Kirby, T. A., Duckworth, A. L., & Mayer, D. (2011). Mental contrasting facilitates academic performance in school children. *Motivation and Emotion*, 35(4), 403-412.
- Gong, Y., & Beck, J. E. (2015, March). Towards detecting wheel-spinning: Future failure in mastery learning. In *Proceedings of the Second (2015) ACM Conference on Learning@Scale* (pp. 67-74). New York, NY: Association for Computing Machinery.
- Good, C., Aronson, J., & Inzlicht, M. (2003). Improving adolescents' standardized test performance: An intervention to reduce the effects of stereotype threat. *Journal of Applied Developmental Psychology*, 24(6), 645-662.
- Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area and results from diagnostic interviews. *Psychiatry Res.* 15, under a receiver operating characteristic curve. *Radiology*, 143, 29-36.
- Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus*. New York, NY: Routledge.
- Heffernan, N., Heffernan, C., Dietz, K., Soffer, D., Pellegrino, J. W., Goldman, S. R., & Dailey, M. (2012). *Improving mathematical learning outcomes through automatic reassessment and relearning*. Annual meeting of American Educational Research Association, Vancouver, British Columbia.
- Heiner, C., Beck, J. E., & Mostow, J. (2004, June 17-19). Improving the Help Selection Policy in a Reading Tutor that Listens. *Proceedings of the InSTIL/ICALL Symposium on NLP and Speech Technologies in Advanced Language Learning Systems*, 195-198.

- Holstein, K., McLaren, B. M., & Aleven, V. (2017). Intelligent tutors as teachers' aides: exploring teacher needs for real-time analytics in blended classrooms. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 257-266). New York, NY: Association for Computing Machinery.
- Janoff-Bulman, R., & Brickman, P. (1982). Expectations and what people learn from failure. In N. T. Feather (Ed.), *Expectations and actions: Expectancy-value models in psychology* (pp. 207-237). Hillsdale, NJ: Erlbaum.
- Kapur, M. (2014). Productive failure in learning math. *Cognitive Science*, 38(5), 1008-1022.
- Kintsch, W., & Greeno, J. G. (1985). Understanding and solving word arithmetic problems. *Psychological Review*, 92(1), 109.
- Koedinger, K. R., & MacLaren, B. A. (2002). *Developing a pedagogical domain theory of early algebra problem solving*. Pittsburgh, PA: Carnegie Mellon University.
- Koedinger, K. R., & Nathan, M. J. (2004). The real story behind story problems: Effects of representations on quantitative reasoning. *The Journal of the Learning Sciences*, 13(2), 129-164.
- Landauer, T. K., & Bjork, R. A. (1978). Optimum rehearsal patterns and name learning. In M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory* (pp. 625-632). London: Academic Press.
- Laursen, E. K. (2015). The power of grit, perseverance, and tenacity. *Reclaiming Children and Youth*, 23(4), 19.
- Lucas, G. M., Gratch, J., Cheng, L., & Marsella, S. (2015). When the going gets tough: Grit predicts costly perseverance. *Journal of Research in Personality*, 59, 15-22.
- Matsuda, N., Chandrasekaran, S., & Stamper, J. C. (2016). How quickly can wheel spinning be detected? In *Proceedings of the 9th International Conference on Educational Data Mining*, 607-608.
- McCarthy, K., Likens, A., Kopp, K., Perret, C., Watanabe, M., and McNamara, D. (2018). The "LO"-down on grit: Non-cognitive trait assessments fail to predict learning gains in iSTART and W-Pal. Paper presented at LAK 2018, Workshop on Online Learning & Non-Cognitive Assessment at Scale, Sydney, NSW, AU.
- Mierswa, I., Wurst, M., Klinkenberg, R., Scholz, M., & Euler, T. (2006). Yale: Rapid prototyping for complex data mining tasks. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 935-940.

- Nathan, M. J., Kintsch, W., & Young, E. (1992). A theory of algebra-word-problem comprehension and its implications for the design of learning environments. *Cognition and Instruction*, 9(4), 329-389.
- Nathan, M. J., & Koedinger, K. R. (2000). Moving beyond teachers' intuitive beliefs about algebra learning. *The Mathematics Teacher*, 93(3), 218.
- Paige, J. M., & Simon, H. A. (1966). Cognition processes in solving algebra word problems. In B. Kleinmuntz (Ed.), *Problem solving: Research, method and theory* (pp. 119-151). New York, NY: Wiley.
- Pardos, Z. A., Baker, R. S., San Pedro, M. O. C. Z., Gowda, S. M., & Gowda, S. M. (2014). Affective states and state tests: Investigating how affect and engagement during the school year predict end of year learning outcomes. *Journal of Learning Analytics*, 1(1), 107-128.
- Pekrun, R., & Linnenbrink-Garcia, L. (2012). Academic emotions and student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 259-282). New York, NY: Springer.
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. Oxford, NY: American Psychological Association and Oxford University Press.
- Prabhu, V., Sutton, C., & Sauser, W. (2008). Creativity and certain personality traits: Understanding the mediating effect of intrinsic motivation. *Creativity Research Journal*, 20(1), 53-66.
- Revell, A. D., Wang, D., Wood, R., Morrow, C., Tempelman, H., Hamers, R. L., Alvarez-Uria, G., Streinu-Cercel, A., Ene, L., Wensing, A. M. J., DeWolf, F., Nelson, M., Montaner, J. S., Lane, H. C., & Larder, B. A. (2013). Computational models can predict response to HIV therapy without a genotype and may reduce treatment failure in different resource-limited settings. *Journal of Antimicrobial Chemotherapy*, 68(6), 1406-1414.
- Rohrer, D., Dedrick, R. F., & Burgess, K. (2014). The benefit of interleaved mathematics practice is not limited to superficially similar kinds of problems. *Psychonomic Bulletin and Review*, 21(5), 1323-1330.
- Rohrer, D., & Taylor, K. (2006). The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, 20(9), 1209-1224.
- San Pedro, M. O. Z., Baker, R. S., Gowda, S. M., & Heffernan, N. T. (2013, July). Towards an understanding of affect and knowledge from student interaction with an intelligent tutoring system. In *International Conference on Artificial Intelligence in Education* (pp. 41-50). Berlin, Heidelberg: Springer.

- Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). Reston, VA: National Council of Teachers of Mathematics.
- Sedek, G., & Kofta, M. (1990). When cognitive exertion does not yield cognitive gain: Toward an informational explanation of learned helplessness. *Journal of Personality and Social Psychology*, 58(4), 729.
- Settles, B., & Meeder, B. (2016). A trainable spaced repetition model for language learning. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (Vol. 1, pp. 1848-1858).
- Shechtman, N., DeBarger, A. H., Dornsife, C., Rosier, S., & Yarnall, L. (2013). *Promoting grit, tenacity, and perseverance: Critical factors for success in the 21st century*. Washington, DC: U.S. Department of Education, Department of Educational Technology.
- Strother, S., Van Campen, J., & Grunow, A. (2013). Community college pathways: 2011-2012 descriptive report. Stanford, CA: Carnegie Foundation for the Advancement of Teaching.
- Ventura, M., & Shute, V. (2013). The validity of a game-based assessment of persistence. *Computers in Human Behavior*, 29(6), 2568-2572.
- Warshauer, H. K. (2015). Productive struggle in middle school mathematics classrooms. *Journal of Mathematics Teacher Education*, 18(4), 375-400.
- West, M. R., Kraft, M. A., Finn, A. S., Martin, R. E., Duckworth, A. L., Gabrieli, C. F., & Gabrieli, J. D. (2016). Promise and paradox: Measuring students' non-cognitive skills and the impact of schooling. *Educational Evaluation and Policy Analysis*, 38(1), 148-170.
- Wrosch, C., Scheier, M. F., Miller, G. E., Schulz, R., & Carver, C. S. (2003). Adaptive self-regulation of unattainable goals: Goal disengagement, goal reengagement, and subjective well-being. *Personality and Social Psychology Bulletin*, 29(12), 1494-1508.
- Wrzesniewski, A. (2012). Callings. In K. S. Cameron & G. M. Spreitzer (Eds.), *The Oxford handbook of positive organizational scholarship* (pp. 45-55). New York, NY: Oxford University Press.
- Yeager, D., Bryk, A., Muhich, J., Hausman, H., & Morales, L. (2013). *Practical measurement*. Palo Alto, CA: Carnegie Foundation for the Advancement of Teaching.
- Yeager, D., Muhich, J., Torres, L., & Asers, R. (2012). Productive persistence: A practical theory of community college student success. Unpublished manuscript, Carnegie Foundation.

- Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1), 82-91.
- Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 233-262). New York, NY: Cambridge University Press.
- Zimmerman, B. J., & Martinez-Pons, M. (1988). Construct validation of a strategy model of student self-regulated learning. *Journal of Educational Psychology*, 80(3), 284.
- Zimmerman, B. J., & Schunk, D. H. (Eds.). (2001). *Self-regulated learning and academic achievement: Theoretical perspectives*. New York, NY: Routledge.

## Appendix A: J48 Decision Tree of the 15-feature Model

```

amin-hintTotal <= 0
| amax-past8BottomOut <= 1
| | std-timeSinceSkill <= 218115.7837
| | | sum-responseIsChosen <= 26
| | | | amax-totalFrAttempted <= 18: 1 (898.0/172.0)
| | | | amax-totalFrAttempted > 18
| | | | | amax-totalFrAttempted <= 491
| | | | | | sum-frIsHelpRequestScaffolding <= 2
| | | | | | | amax-totalFrSkillOpportunities <= 6
| | | | | | | | amax-totalFrSkillOpportunities <= 1
| | | | | | | | | amax-responseIsChosen <= 0
| | | | | | | | | | sum-totalFrAttempted <= 126: 0 (7.0/2.0)
| | | | | | | | | | sum-totalFrAttempted > 126: 1 (5.0)
| | | | | | | | | | amax-responseIsChosen > 0
| | | | | | | | | | | sum-responseIsChosen <= 4: 0 (20.0/6.0)
| | | | | | | | | | | sum-responseIsChosen > 4: 1 (2.0)
| | | | | | | amax-totalFrSkillOpportunities > 1
| | | | | | | | sum-frPast8WrongCount <= 34
| | | | | | | | | std-timeSinceSkill <= 6.988398
| | | | | | | | | | amax-totalFrPercentPastWrong <= 0.9
| | | | | | | | | | | sum-totalFrAttempted <= 422
| | | | | | | | | | | | sum-totalFrAttempted <= 209: 1 (2.0)
| | | | | | | | | | | | sum-totalFrAttempted > 209: 0 (5.0)
| | | | | | | | | | | | sum-totalFrAttempted > 422: 1 (20.0)
| | | | | | | | | | | | amax-totalFrPercentPastWrong > 0.9: 0 (4.0)
| | | | | | | | | | | std-timeSinceSkill > 6.988398: 1 (161.0/12.0)
| | | | | | | | | | sum-frPast8WrongCount > 34
| | | | | | | | | | | std-totalFrSkillOpportunitiesByScaffolding <= 0
| | | | | | | | | | | | amax-totalFrPercentPastWrong <= 0.363636: 1 (5.0/1.0)
| | | | | | | | | | | | amax-totalFrPercentPastWrong > 0.363636
| | | | | | | | | | | | | std-frWorkingInSchool <= 0.278325: 1 (3.0/1.0)
| | | | | | | | | | | | | std-frWorkingInSchool > 0.278325: 0 (4.0)
| | | | | | | | | | | | std-totalFrSkillOpportunitiesByScaffolding > 0: 0 (3.0)
| | | | | | | amax-totalFrSkillOpportunities > 6
| | | | | | | | amax-past8BottomOut <= 0
| | | | | | | | | std-totalFrSkillOpportunitiesByScaffolding <= 2.18526
| | | | | | | | | | amax-responseIsChosen <= 0
| | | | | | | | | | | std-totalFrSkillOpportunitiesByScaffolding <= 0.964126: 1 (992.0/344.0)
| | | | | | | | | | | std-totalFrSkillOpportunitiesByScaffolding > 0.964126
| | | | | | | | | | | | std-timeSinceSkill <= 23.786277: 0 (86.0/31.0)
| | | | | | | | | | | | std-timeSinceSkill > 23.786277
| | | | | | | | | | | | | std-timeSinceSkill <= 13108.93518: 1 (15.0/1.0)
| | | | | | | | | | | | | std-timeSinceSkill > 13108.93518
| | | | | | | | | | | | | | sum-frPast8WrongCount <= 30: 1 (3.0/1.0)
| | | | | | | | | | | | | | sum-frPast8WrongCount > 30: 0 (5.0/1.0)
| | | | | | | | | | amax-responseIsChosen > 0
| | | | | | | | | | | amax-totalFrSkillOpportunities <= 9
| | | | | | | | | | | | amax-totalFrPercentPastWrong <= 0.9: 1 (26.0/3.0)
| | | | | | | | | | | | amax-totalFrPercentPastWrong > 0.9
| | | | | | | | | | | | | std-totalFrSkillOpportunitiesByScaffolding <= 1.022268
| | | | | | | | | | | | | | mean-totalFrAttempted <= 105.538462
| | | | | | | | | | | | | | | mean-totalFrAttempted <= 67.8: 0 (2.0)
| | | | | | | | | | | | | | | mean-totalFrAttempted > 67.8: 1 (2.0)
| | | | | | | | | | | | | | | mean-totalFrAttempted > 105.538462: 0 (5.0)
| | | | | | | | | | | | | std-totalFrSkillOpportunitiesByScaffolding > 1.022268
| | | | | | | | | | | | | | sum-totalFrAttempted <= 756: 0 (4.0/1.0)
| | | | | | | | | | | | | | sum-totalFrAttempted > 756: 1 (8.0)
| | | | | | | | | | | amax-totalFrSkillOpportunities > 9: 1 (603.0/196.0)
| | | | | | | | | | | std-totalFrSkillOpportunitiesByScaffolding > 2.18526
| | | | | | | | | | | | amax-totalFrPercentPastWrong <= 0.363636
| | | | | | | | | | | | | sum-responseIsChosen <= 3: 1 (9.0/2.0)
| | | | | | | | | | | | | sum-responseIsChosen > 3: 0 (2.0)
| | | | | | | | | | | | amax-totalFrPercentPastWrong > 0.363636: 0 (22.0/1.0)
| | | | | | | amax-past8BottomOut > 0
| | | | | | | | amax-totalFrAttempted <= 121: 1 (1060.0/281.0)
| | | | | | | | | amax-totalFrAttempted > 121
| | | | | | | | | | amax-totalFrAttempted <= 302
| | | | | | | | | | | amax-responseIsChosen <= 0: 1 (207.0/86.0)
| | | | | | | | | | | amax-responseIsChosen > 0
| | | | | | | | | | | | sum-frPast8WrongCount <= 92
| | | | | | | | | | | | sum-responseIsChosen <= 10
| | | | | | | | | | | | | std-frWorkingInSchool <= 0.438019

```

```

sum-frIsHelpRequestScaffolding <= 1
std-timeSinceSkill <= 37505.97416
sum-frPast8WrongCount <= 35
sum-frPast8WrongCount <= 28
std-totalFrSkillOpportunitiesByScaffolding <= 0.943037: 0 (3.0)
std-totalFrSkillOpportunitiesByScaffolding > 0.943037: 1 (3.0/1.0)
sum-frPast8WrongCount > 28: 1 (6.0)
sum-frPast8WrongCount > 35
sum-frPast8WrongCount <= 56: 0 (9.0)
sum-frPast8WrongCount > 56
std-frWorkingInSchool <= 0.158114: 1 (4.0/1.0)
std-frWorkingInSchool > 0.158114: 0 (3.0)
std-timeSinceSkill > 37505.97416: 1 (4.0)
sum-frIsHelpRequestScaffolding > 1
mean-totalFrAttempted <= 187.12766: 1 (2.0)
mean-totalFrAttempted > 187.12766: 0 (3.0/1.0)
std-frWorkingInSchool > 0.438019: 1 (6.0)
sum-responseIsChosen > 10: 1 (20.0/2.0)
sum-frPast8WrongCount > 92
std-timeSinceSkill <= 28605.3143
amax-totalFrPercentPastWrong <= 0.96: 0 (6.0)
amax-totalFrPercentPastWrong > 0.96
sum-responseIsChosen <= 13: 0 (6.0)
sum-responseIsChosen > 13: 1 (2.0)
std-timeSinceSkill > 28605.3143: 1 (2.0)
amax-totalFrAttempted > 302
std-frWorkingInSchool <= 0.390205: 1 (30.0/2.0)
std-frWorkingInSchool > 0.390205
sum-frPast8WrongCount <= 37: 0 (2.0)
sum-frPast8WrongCount > 37: 1 (3.0/1.0)
sum-frIsHelpRequestScaffolding > 2
amax-totalFrAttempted <= 307
amax-responseIsChosen <= 0
std-frWorkingInSchool <= 0.162221: 1 (70.0/1.0)
std-frWorkingInSchool > 0.162221
sum-frIsHelpRequestScaffolding <= 7: 1 (15.0/1.0)
sum-frIsHelpRequestScaffolding > 7
std-timeSinceSkill <= 92.603468: 1 (4.0)
std-timeSinceSkill > 92.603468
sum-frIsHelpRequestScaffolding <= 21: 0 (6.0)
sum-frIsHelpRequestScaffolding > 21: 1 (2.0)
amax-responseIsChosen > 0: 1 (86.0/18.0)
amax-totalFrAttempted > 307
sum-frIsHelpRequestScaffolding <= 5: 1 (3.0)
sum-frIsHelpRequestScaffolding > 5
std-totalFrSkillOpportunitiesByScaffolding <= 1.542118: 0 (10.0/1.0)
std-totalFrSkillOpportunitiesByScaffolding > 1.542118: 1 (3.0/1.0)
amax-totalFrAttempted > 491: 1 (32.0)
sum-responseIsChosen > 26: 1 (675.0/90.0)
std-timeSinceSkill > 218115.7837
std-timeSinceSkill <= 12799929.2
amax-totalFrSkillOpportunities <= 5
std-timeSinceSkill <= 1820838.272
std-totalFrSkillOpportunitiesByScaffolding <= 0
sum-frPast8WrongCount <= 48: 1 (76.0/9.0)
sum-frPast8WrongCount > 48: 0 (4.0/1.0)
std-totalFrSkillOpportunitiesByScaffolding > 0
std-timeSinceSkill <= 684764.2974: 0 (3.0)
std-timeSinceSkill > 684764.2974: 1 (3.0)
std-timeSinceSkill > 1820838.272
sum-frPast8WrongCount <= 19
std-frWorkingInSchool <= 0.357935
std-timeSinceSkill <= 3294065.607: 0 (2.0)
std-timeSinceSkill > 3294065.607: 1 (3.0)
std-frWorkingInSchool > 0.357935: 0 (7.0)
sum-frPast8WrongCount > 19: 1 (4.0)
amax-totalFrSkillOpportunities > 5
amax-past8BottomOut <= 0
std-timeSinceSkill <= 2773779.763
amax-responseIsChosen <= 0: 0 (228.0/107.0)
amax-responseIsChosen > 0
amax-totalFrPercentPastWrong <= 0.363636
sum-responseIsChosen <= 23
std-frWorkingInSchool <= 0.347839

```



## Appendix B: Features Selected With J48 Decision Tree Algorithm

Features	Description
amax-past8BottomOut	(Maximum) Number of bottom-out hints requested within the last 8 problems in a given problem set
amax-responseIsChosen	(Maximum) Whether or not a problem requires the correct answer to be chosen from a list of answers (e.g., multiple choice)
amax-totalFrAttempted	(Maximum) The total number of problems attempted in the tutor so far
amax-totalFrPercentPastWrong	(Maximum) The percentage of all past problems that were incorrect on a given problem set
amax-totalFrSkillOpportunities	(Maximum) The total number of unique problems the user has encountered relevant to the current problem set
mean-totalFrAttempted	(Mean) The total number of problems attempted in the tutor so far
std- frIsHelpRequestScaffolding	(Standard deviation) Whether or not the first response to a scaffolding problem is a help request
std-frWorkingInSchool	(Standard deviation) Whether or not the first response was made during or after school hours (between 7:00 a.m. and 3:00 p.m.)
std-timeSinceSkill	(Standard Deviation) Length of time since a problem involving this skill type was last seen
std- totalFrSkillOpportunitiesByScaffolding	(Standard Deviation) The total number of scaffolding problems divided by the unique problems the user has encountered relevant to the current problem set
sum- frIsHelpRequestScaffolding	(Sum) Whether or not the first response to a scaffolding problem is a help request
sum-frPast8WrongCount	(Sum) Cumulative count of the number of first responses to a problem that were wrong answers in the past 8 problems
sum-responseIsChosen	(Sum) Whether or not a problem requires the correct answer to be chosen from a list of answers (e.g., multiple choice)
sum-totalFrAttempted	(Standard Deviation) The total number of problems attempted in the tutor so far

### Appendix C: Short Grit Scale (Grit-S)

Directions for taking the Grit Scale: Please respond to the following 8 items. Be honest – there are no right or wrong answers!

1. New ideas and projects sometimes distract me from previous ones.
  - ☐ Very much like me
  - ☐ Mostly like me
  - ☐ Somewhat like me
  - ☐ Not much like me
  - ☐ Not like me at all
2. Setbacks don't discourage me.
  - ☐ Very much like me
  - ☐ Mostly like me
  - ☐ Somewhat like me
  - ☐ Not much like me
  - ☐ Not like me at all
3. I have been obsessed with a certain idea or project for a short time but later lost interest.
  - ☐ Very much like me
  - ☐ Mostly like me
  - ☐ Somewhat like me
  - ☐ Not much like me
  - ☐ Not like me at all
4. I am a hard worker.
  - ☐ Very much like me
  - ☐ Mostly like me
  - ☐ Somewhat like me
  - ☐ Not much like me
  - ☐ Not like me at all
5. I often set a goal but later choose to pursue a different one.
  - ☐ Very much like me
  - ☐ Mostly like me
  - ☐ Somewhat like me
  - ☐ Not much like me
  - ☐ Not like me at all

6. I have difficulty maintaining my focus on projects that take more than a few months to complete.
- ☐ Very much like me
  - ☐ Mostly like me
  - ☐ Somewhat like me
  - ☐ Not much like me
  - ☐ Not like me at all
7. I finish whatever I begin.
- ☐ Very much like me
  - ☐ Mostly like me
  - ☐ Somewhat like me
  - ☐ Not much like me
  - ☐ Not like me at all
8. I am diligent.
- ☐ Very much like me
  - ☐ Mostly like me
  - ☐ Somewhat like me
  - ☐ Not much like me
  - ☐ Not like me at all

#### Appendix D: Skill Builders with Hard Difficulty Levels

1. Area and perimeter 4.MD.A.3
2. Combining Like Terms 8.EE.C.7b EX
3. Converting a Fraction to a Percent 6.RP.A.3c EX
4. Distributive Property 7.EE.A.1 EX
5. Finding Scale Factor 7.G.A.1
6. Finding Slope from Ordered Pairs 8.F.B.4 Ex
7. Finding the Percent from the Part and Whole 6.RP.A.3c
8. Finding the Ratio 6.RP.A.1 EX
9. Finding the Whole from the Percent and Part in a Word Problem 6.RP.A.3c
10. Prime Factorization 6.NS.B.4 EX
11. Pythagorean Theorem - Finding Leg or Hypotenuse 8.G.B.7 EX
12. Recognizing Statistical Questions 6.SP.A.1
13. Surface Area Rectangular Prism 7.G.B.6 EX
14. Writing an Equation from a Real-World Problem 6.EE.C.9 EX
15. Writing Inequalities from Situations 6.EE.B.8 EX