Refusing to Try: Characterizing Early Stopout on Student Assignments

Anthony F. Botelho
Worcester Polytechnic Institute
abotelho@wpi.edu

Ashvini Varatharaj
Worcester Polytechnic Institute
avaratharaj@wpi.edu

Eric G. Van Inwegen
Independent Researcher
research@ericvaninwegen.net

Neil T. Heffernan
Worcester Polytechnic Institute
nth@wpi.edu

ABSTRACT

A prominent issue faced by the education research community is that of student attrition. While large research efforts have been devoted to studying course-level attrition, widely referred to as dropout, less research has been focused on finer-grained assignment-level attrition commonly observed in K-12 classrooms. This later instantiation of attrition, referred to in this paper as "stopout," is characterized by students failing to complete their assigned work, but the cause of such behavior are not often known. This becomes a large problem for educators and developers of learning platforms as students who give up on assignments early are missing opportunities to learn and practice the material which may affect future performance on related topics; similarly, it is difficult for researchers to develop, and subsequently difficult for computer-based systems to deploy interventions aimed at promoting productive persistence once a student has ceased interaction with the software. This difficulty highlights the importance to understand and identify early signs of stopout behavior in order to provide aid to students preemptively to promote productive persistence in their learning. While many cases of student stopout may be attributable to gaps in student knowledge and indicative of struggle, student attributes such as grit and persistence may be further affected by other factors. This work focuses on identifying different forms of stopout behavior in the context of middle school math by observing student behaviors at the sub-problem level. We find that students exhibit disproportionate stopout on the first problem of their assignments in comparison to stopout on subsequent problems, identifying a behavior that we call "refusal," and use the emerging patterns of student activity to better understand the potential causes underlying stopout behavior early in an assignment.

CCS CONCEPTS

• **Social and professional topics → K-12 education**; *Student assessment*; • **Applied computing → Education**; • **Computing methodologies →** *Cluster analysis*; *Model verification and validation.*

KEYWORDS

Student Attrition, Persistence, Dropout, Stopout, Refusal

ACM Reference Format:
Anthony F. Botelho, Ashvini Varatharaj, Eric G. Van Inwegen, and Neil T. Heffernan. 2019. Refusing to Try: Characterizing Early Stopout on Student Assignments. In *The 9th International Learning Analytics & Knowledge Conference (LAK19), March 4–8, 2019, Tempe, AZ, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3303772.3303806

1 INTRODUCTION

Persistence is an essential factor of student learning as it is important for students to have the opportunity to work through problems and apply deliberate practice, particularly when exhibiting early struggle when learning new material. The study of this construct of learning has led to research into such student attributes as grit [1], perseverance [2], as well as other representations of high student persistence such as academic tenacity [3], productive struggle [4], and productive failure [5]. All of these theories of learning recognize that persistence is necessary in order for students to effectively overcome difficulties faced when learning new material. It is similarly understood that the lack of persistence can deprive students of the opportunity to effectively learn new and difficult material which may then propagate to affect the students' ability to learn subsequent post-requisite content. It is important, therefore, to ensure that students are able to take advantage of practice opportunities when they will be productive for learning and identify struggling students early to provide them with the help they need to succeed.

While not all representations of persistence are productive, such as the case of wheel spinning behavior (e.g. see [6]), it is often beneficial for students to exhibit high persistence during early learning opportunities. In this way, early student attrition becomes a significant problem for instructors and learning platforms as it is difficult to develop and deploy learning interventions and provide aid to students who cease interaction with the course or learning software. Not all student attrition, however, is exhibited in the same way and can emerge at varying levels of granularity.

With the emergence of massive open online courses (MOOCs), attrition in the form of student dropout has received a large amount of attention and research. The reasoning for which a student exhibits

LAK19, March 4–8, 2019, Tempe, AZ, USA

ACM ISBN 978-1-4503-6256-6/19/03…$15.00
https://doi.org/10.1145/3303772.3303806

dropout, characterized as ceasing interaction with or explicitly leaving a course, has also been a well-studied problem within MOOCs [7][8][9][10][11] as such courses often observe high attrition rates. Although dropout of this nature is not commonly observed in K-12 classrooms, attrition is still a prominent problem within this context and has received significantly less attention and research focus in previous years. Particularly as more classrooms begin to utilize computer-based learning platforms to assign classwork and homework, supplement instruction, and provide aid to students, there are new opportunities to study student attrition at fine granular levels.

In the context of K-12 classrooms, it is common to observe student attrition at the assignment-level, where students begin an assignment but fail or choose not to complete the assigned work. This behavior, which we call "stopout," is distinctly different from the course-level dropout that is observed in MOOCs as students likely return to work on subsequent assignments; the student remains in the course, but did not finish the assigned work. Similar to the study of dropout, the reasoning for stopout behavior is not often known, but observing the immediate prior action that a student takes before stopout occurs within a given assignment may help to provide insight into the cause of the behavior. A student who exhibits stopout early in an assignment may do so for different reasons than a student who exhibits the behavior after attempting several problems, or learning opportunities as they will be referred in this work.

1.1 Student Refusal

In order to provide sufficient context for the goals and motivation of the current work, we must first describe a student behavior that emerged during a previous unpublished analysis of student stopout on a per-problem level conducted in 2015; this analysis is repeated here and will be described with greater detail in Section 4.2.

In observing when stopout occurs within student assignments, what quickly became apparent was that there seemed to be a disproportionate number of students exhibiting stopout on the first learning opportunity. Assuming that there would be a reasonably consistent failure rate over each opportunity, we found that student stopout by opportunity followed an exponential, or more specifically, Weibull distribution as is commonly observed in survival analyses [12]. However, while most of the data followed this trend, the number of students exhibiting the behavior on the first opportunity was nearly double what would be expected by the fit exponential curve, as will also be demonstrated by Figure 4 in Section 4.2.

This behavior, which we call "refusal" was first used to identify problematic content within the learning system in which it was discovered, and is explored further in this work in an effort to better understand student interactions with the learning platform that may be indicative of early stopout behavior. The goal of this work is to explore the student actions associated with stopout and refusal behavior to better understand the potential causes of assignment-level student attrition within a computer-based learning platform. As students who exhibit refusal stop out of their assignments with little-to-no recorded interactions, it is these students who are arguably most important to identify in order to develop effective learning interventions to address any potential causes of this unproductive behavior.

In this research, we conduct a set of fine-grained analyses to determine the frequency of stopout as it correlates to the to the estimated knowledge level of each student in conjunction with the specific actions taken within the system immediately prior to their stopout. We also then extend these analyses to include the dataset collected by Lang et al. [13] wherein they study the role of confidence on student learning using self-report surveys in a randomized controlled trial.

We seek to show in this paper that:

(1) Student stopout after the first problem can be stochastically modelled as an exponential decay, but that this model fails to account for roughly half of the stopout that occurs on the first problem.
(2) Specific actions (immediately prior to stopout) by students correlate with different patterns of stopout over time.
(3) High stopout on the first problem correlates to low levels of self-reported confidence.

2 BACKGROUND

The study of stopout in computer-based systems has largely focused on MOOCs in recognition of the often large attrition rates experienced by such courses. While the actions available to students in such courses often makes for feature-rich datasets with which to study attrition, the dropout behavior exhibited within such systems tend to observe contextual factors including the attitude of the student [7], the estimated knowledge level of the student [14] combined with the effort exhibited by the student [9], as well as several other contextual factors such as technology, time management [15], and other social factors [10].

Within these, however, it becomes clear that stopout behavior is not random but is seemingly motivated by more internal factors than external. The student is ultimately making the choice to dropout or stopout; many times, this is predictively so [16], supporting the need to further understand why attrition occurs.

The problem of student stopout, however, is more prominent in K-12 classrooms than that of dropout experienced more in MOOC settings. In many cases, students choose to enroll in MOOCs, and can easily dropout due to a host of reasons briefly described above with little consequence. The problem of stopout in younger students is much more associated with a lack of persistence or motivation at an assignment-level rather than at the course-level.

The more general study of student persistence has led to a large amount of research exploring various aspects of the construct. Connotatively, persistence is often associated with positive learning behaviors, but in reality observes both beneficial and adverse effects depending on the context of which it is exhibited. It is intuitive that persistence can be beneficial when paired with productive learning behaviors, where learning occurs over time by making errors or receiving help. The productivity of persistence and perseverance is sometimes described by the construct of "grit" [1].

However, persistence may also be unproductive, as is the case of "wheel spinning" [6][17]. Wheel spinning describes the case when students attempt multiple problems but struggle to learn the material; this is analogous to a car that is stuck in mud or snow

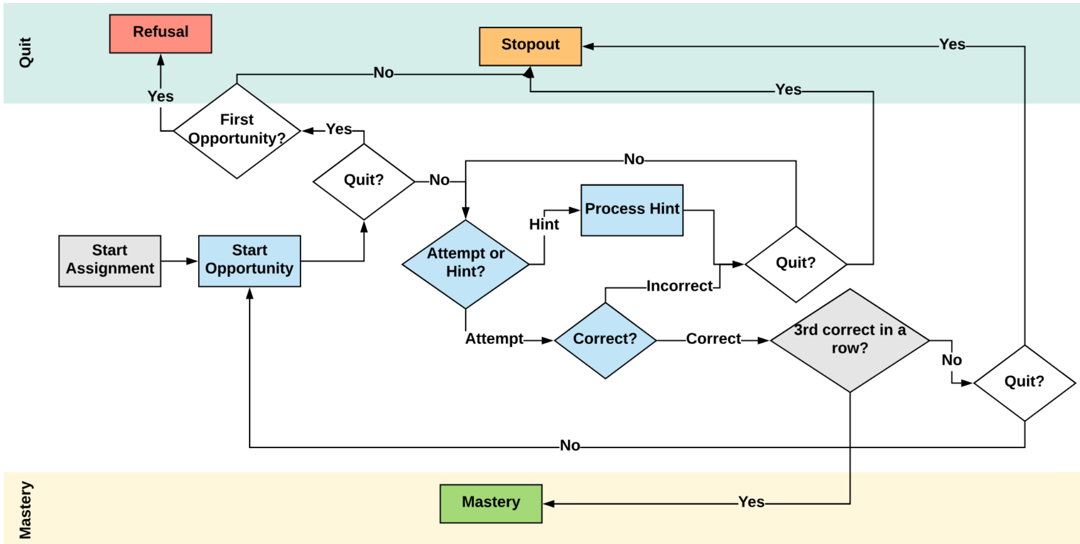

Figure 1: The flowchart of possible student actions resulting in either quitting (refusal or stopout) or mastery of the assignment.

that "spins its wheels" but makes little to no progress. In such cases, stopout is sometimes encouraged as a more productive action, so long as the student takes such an opportunity to seek help from an instructor or parent.

In this work, we examine student behaviors that suggest a lack of persistence, i.e. when students stopout early in the assignment. While stopout may be encouraged in very select scenarios, as in the case of wheel spinning, it is generally considered a negative learning behavior as students lose the opportunity to learn through additional practice opportunities.

3 DATASET

The dataset used in this work consists of student log data collected as real students work in ASSISTments [18][19], a web-based learning platform aimed at supporting teachers and providing students with immediate correctness feedback on homework and classwork. The system hosts content across K-12 grade levels and even some college content, but is focused largely on middle school math content. Within the system, teachers can use the content provided by the system or create their own to assign to their students. The data used in this work is comprised mastery-based assignments, referred to as "skill builders" within the system. These skill builders usually give students isomorphic questions (generated from one or closely related templates) that have been previously generated, but randomly presented to the students; templates and questions are tightly associated in a single skill or sub-skill. Since the problems that student see are randomly selected from a large pool, we examine data not per problem, but rather per opportunity - i.e. the first problem a student sees is opportunity 1, the second is opportunity 2, etc.

Within the ASSISTments system, after opening a given problem, students can either submit an answer (and will receive instant correctness feedback), or they may use a help feature, such as requesting a hint. Hints (the most common type of help in this dataset) are usually written as some version of a complete worked out solution, often broken into pieces; the last hint (colloquially referred to as the bottom-out hint) gives the answer to the problem. If a student enters an incorrect answer (or requests a hint), they may then enter any number of attempts and use as many or as few of the hints as needed; the student must enter the correct answer before they are able to proceed to the next question. In order to successfully complete a Skill Builder, a student must enter the correct answer on the first attempt, using no help features, three times in a row.

Thus, at any given moment, a student can be said to be in one of three mutually exclusive conditions: Quit (either refusal or stopout), Working, or Mastery, as illustrated by Figure 1. The primary dataset in this analysis was taken from a previous school year; we also used the dataset from [13], which also comes from a prior academic year. Thus, when looking at the datasets, students have either attained mastery or have quit.

As we examine the behavior of students who have quit, we also note the action taken immediately prior to quitting. In ASSISTments, there are four possible actions a student may take before quitting a Skill Builder: they may have Opened an Opportunity (but have done nothing else), entered a Correct Attempt, entered an Incorrect Attempt, or made a Help Request. In this analysis, we make no differentiation of whether the help requested gave an initial step in the solution or the final answer.

In this paper, we will use the term stopout to refer to any student who leaves (and never returns to) an unfinished assignment. Furthermore, for reasons discussed below, we refer to one specific type of stopout as refusal - that is, students who quit an assignment having only opened the first problem, without using any hint features or entering an attempt to answer it.

The data used in this work uses data from the 2016-2017 academic year and includes information recorded from 3,641 distinct students

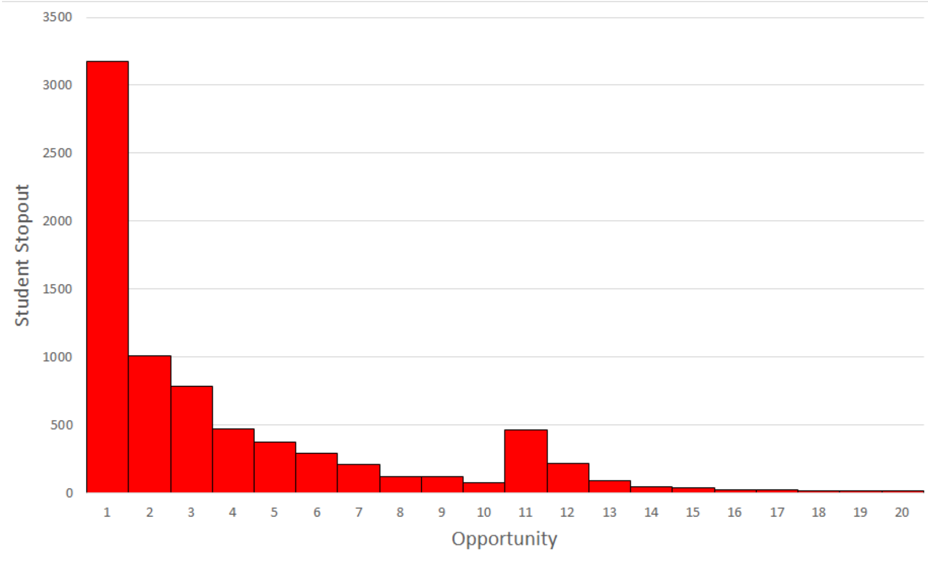

Figure 2: The frequency of student stopout by learning opportunity. Stopout on the first opportunity appears to be disproportionately larger than subsequent opportunities.

who exhibited stopout on skill builder assignments. Each row of the dataset corresponds to a single assignment attempted by a student. As this work is studying only those who exhibited stopout, students who complete each assignment are not included in the data or analyses. In an effort to remove cases where the completion of an assignment may have been optional, only assignments that had been started by at least 10 students and have an overall completion rate higher than 75% were considered for the analyses.

A second dataset, described further in Section 4.4, was also used to observed the relationship between stopout behavior and student confidence. This data consists of students interacting with the ASSISTments learning platform for a randomized controlled trial studying student confidence [13]. From the dataset used in that work, we extracted all students from the treatment condition (e.g. the students who received a confidence survey prior to beginning their assignment) who exhibited stopout during the assignment; this excludes any student who stopped out on the initial survey as well as students who finished the survey but did not begin the first non-survey problem of the assignment. The resulting dataset used in this work consists of 438 distinct students who exhibited stopout.

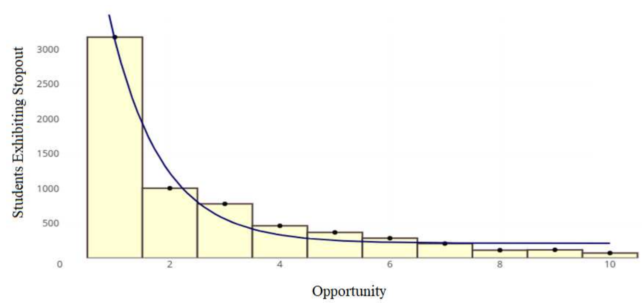

Figure 3: The exponential curve fit to stopout on the first ten learning opportunities. The line is a poor fit seemingly due to stopout on the first item.

4 METHODOLOGY

4.1 Characterizing Early Stopout and Refusal

It is important to clarify, before describing our analyses, how we have defined stopout within the data. In any sense, just as it has been described in earlier sections, stopout is exhibited when a student begins an assignment and fails or refuses to finish that assignment. It follows, then, that students who never begin an assignment did not exhibit stopout and are therefore not included in our data or analyses[1]. It is found that when students do stopout, however, it occurs after four distinct kinds of actions taken in the system. Students stopout either during a problem, or exhibit stopout after completing a problem but before progressing to the subsequent problem; in this later case, the student managed to enter the correct answer, but stopped out before seeing the next problem. In such a case, we mark the student as stopping out on the following opportunity. For example, if the student enters the correct answer to the first problem, or opportunity, but does not begin the second problem, that student is said to have stopped out on the second opportunity as the first problem was sufficiently completed. When students stopout during

[1]Although we would have preferred to include these students in our analyses, given the variety of grading policies of individual teachers we would be unable to determine how many students were required to complete an assignment, but never even opened it. We can state for certain how many students opened the assignment and failed to complete it; we cannot state for certain how many students should have opened the assignment, but did not.

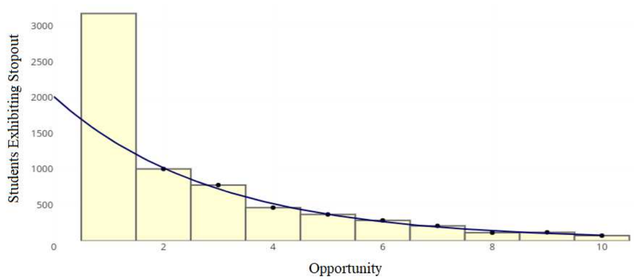

Figure 4: The exponential curve fit to stopout on opportunities 2 through 10, extended to show predicted stopout on the first problem.

a problem, before entering the correct response, those students are said to have stopped out on that learning opportunity (e.g. the student opens the first problem makes an incorrect attempt, or even no attempt, and then stops out is defined as the student stopping out on the first opportunity).

In order to better understand the behavior associated with stopout on skill builder assignments, it is important to first understand how stopout is exhibited independent of students and assignments. As was introduced in Section 1.1, we can explore this by simply observing the trends of stopout over all student assignments in the data. We first observe the distribution of where stopout occurs in an assignment by plotting the frequency of stopout by opportunity, as illustrated in Figure 2. Again, as introduced in Section 1.1, it is clear that there is a large number of students who stopout on the first, and subsequently the the eleventh opportunities; this observed spike on the eleventh opportunity can be attributed to students reaching the "daily limit" within the system which stops students who have not completed the assignment by the tenth opportunity, suggesting that they seek help and return to complete the subsequent day (e.g. to help prevent wheel spinning behavior). While the increased stopout observed on the eleventh opportunity to students who do not return after reaching the daily limit, no such reasoning can easily be given to explain the increased stopout observed on the first opportunity.

While visually it appears that there is disproportionate stopout on the first item as compared to subsequent opportunities, we first attempt to show this by exploring the modeling of stopout by opportunity. As the distribution appears to fit an exponential decay function, we fit two such curves to compare the goodness of model fit. We first fit an exponential curve to opportunities 1 through 10, as seen in Figure 3. We compare this model to another exponential curve that uses just opportunities 2 through 10, as seen in Figure 4. The comparison of these two models shows that there is disproportionate stopout that occurs on the first item. The R-squared values confirm this, with the first model exhibiting an R-squared value of .816 calculated over opportunities 2 through 10, and the second model exhibiting an R-squared value of .991 calculated over the same range. The model using just opportunities 2 through 10 fit an exponential curve nearly perfectly to the real data, illustrating where the expected stopout on the first opportunity is if it were to follow the same trend; in this regard, over twice as many students stopout on the first item as expected (an estimated 1,371 as compared to the observed 3,076 students). The observed difference between the expected and the observed number of students exhibiting stopout on the first learning opportunity is hypothesized to describe the estimated number of students exhibiting *refusal* as introduced in Section 1.1.

It is for this reason that it becomes even more pertinent to understand what causes so many students to exhibit refusal, as they stopout before even trying to learn the material. From this alone, it is unclear if students are exhibiting refusal due to a lack of knowledge or confidence, or if other behaviors are the cause, such as those associated with frustration or boredom. The analyses described in the next section, while non-causal, will help to provide insight into the behaviors associated with student stopout.

4.2 Categorizing Stopout Behavior

While the previous analysis observed stopout across all students, we further explore the behaviors associated with stopout for each student assignment. As described, there are several student level factors that may affect how the behavior is interpreted. For example, an estimated higher knowledge student who stops out on the first item without taking any action is likely to do so for different reasons than an estimated lower knowledge student with the same recorded activity; in the first sense, it may be boredom that causes the student to stop out after determining he/she is already comfortable with the material, while the later student may stopout due to low confidence in their ability to solve. It is likely that students cannot be dichotomized so cleanly, where a higher knowledge student stops out due to low confidence, but the analysis presented here will act as an initial step toward identifying these potential causes.

We use one student-level and 4 action-level covariates to group students by their last recorded activity before exhibiting stopout for each assignment. As the same student may stopout on different assignments for varying reasons, each student-assignment is treated as a separate sample, with grouping performed at the assignment level.

At the student-level, we estimate student knowledge based on the percent of correctly answered items attempted before beginning the observed assignment. This estimate will help to identify students who commonly answer problems correctly from those who often struggle to learn new material. As this covariate exhibits a positive skew, the value is squared to produce a more normal distribution representing estimated student knowledge. This transformed prior percent correct for each student will be used in subsequent analyses and referred to simply as prior correctness for simplicity.

The action-level covariates used in this work describe the last action recorded by the system for each student in each assignment. As all students in the dataset exhibited stopout, this represents the last activity taken by the student before stopping out of the assignment. Each action is represented as a binary value, and is limited to just the last action taken by the student. These actions are as follows:

- Opened Problem - denoting that the student opened the problem but made no subsequent action.

- Correct Attempt - the student entered a correct response to complete the problem, but did not progress to the subsequent problem.
- Help Request - the student requested an on-demand hint or scaffolded question, but made no further attempt to answer the problem.
- Incorrect Attempt - the student entered a response but the answer was incorrect.

We group students by their prior correctness and last recorded action using k-means clustering to gain an understanding of the different behaviors that emerge associated with student stopout. Determining the correct value of k in this type of analysis is important to the interpretability of the results. We determine this value using a short grid-search using different values of k between 2 and 15 and observing the variance of within-sum of squares between the emerging clusters similar to a skree plot used in principal component analysis. From this step, a value of 6 is determined to best partition the data; values 5 and 7 were additionally explored, but did not lead to large differences in interpretation, further supporting the usage of 6 groups to summarize the data.

4.3 Stopout Behavior by Opportunity

Once student assignments have been grouped into the 6 clusters described in the previous section, we can further identify how the behaviors associated with stopout change with the opportunity. As we observe differential dropout on the first learning opportunity as compared with subsequent opportunities, we are hoping to observe differences in behaviors across learning opportunities to help explain this phenomenon. By observing how the distribution of the clusters changes with each learning opportunity, we can gain an understanding of which behaviors, if any, occur most on the first opportunity as compared to subsequent opportunities.

We limit our analysis to just the first three learning opportunities. As the number of students present decreases with each opportunity due to stopout, the number of students on later opportunities makes it difficult to make fair comparisons to earlier problems that are better represented by higher numbers of students. Additionally, as students know the threshold of completion being three consecutive correct responses, observing the first three opportunities highlights those students who exhibit the lowest persistence, stopping out on or before the earliest problem of which the assignment can be completed.

The distribution of the clusters is observed, filtering to include those who stopout on the first, second, and third opportunities and visualizing how this distribution changes. As fewer students are available for each opportunity, a proportional distribution is used by dividing the number of students included in each cluster by the total number of students who exhibit stopout at each respective opportunity.

4.4 Observing Student Confidence

Just as is the case with stopout behavior as a whole, refusal likely occurs as a result of many factors. In this work, however, we focus on exploring the relationship between two such possible factors with refusal behavior: lack of knowledge and confidence. As detailed in the description of our cluster analysis, we use prior correctness as an indicator of how well the student is expected to know the material; students who perform well on prior material often exhibit comparatively high performance on subsequent content as the student has demonstrated knowledge of foundational material. In this way, estimated knowledge, or lack thereof, can be explored amongst students exhibiting stopout and refusal behaviors.

In order to observe the relationship between these behaviors and confidence, however, we utilize an auxiliary dataset consisting of students who participated in a randomized controlled trial with the ASSISTments platform in an earlier academic year [13]. In this study, students assigned to the experimental condition were asked to answer a survey item before starting the assignment (and then subsequently asked again during the assignment, although only the initial survey was used in this work). Students were shown an example of the problems that would be seen in the assignment and asked them to self-report their level of confidence on a 5-point scale ranging from 0% (not confident at all) to 100% (very confident). Using the subsequent student data collected from the student assignments, we apply the clusters developed in Section 4.2 to observe any significant differences in reported confidence between each of the clusters. In regard to refusal behavior specifically, we also compare differences in reported confidence for students who exhibit stopout on the first opportunity.

5 RESULTS AND DISCUSSION

The resulting 6 clusters of student prior knowledge and last recorded action is illustrated in Figure 5. Being the only continuous variable, the prior correctness appears to be a distinguishing factor among the student activity. This measure, being close to normally distributed after the described transformation, is represented as a z-scored value across the 6 groups in the figure; cluster 6, for example, represents the highest knowledge students who stopped out after an incorrect answer. Again, this figure is the clustering as performed over the entire dataset independent of the learning opportunity on which students exhibited stopout. The resulting clusters further distinguish themselves by the last action taken by each student, with no cluster found to contain more than one type of action taken by students.

The number of student assignments that fall within each cluster is denoted under each column along with the cluster number. From this, it becomes clear that the majority of students, regardless of high or low knowledge, stop out at the start of a problem without taking action as illustrated by clusters 2 and 5. The clusters with the fewest students, clusters 1 and 3, appear to have the lowest knowledge students who stop out after a help request and after a correct response respectively. The remaining groups, clusters 4 and 6, both contain students who exhibit stopout after an incorrect response, but represent opposing knowledge estimates.

While the clusters themselves seem to offer some interpretation as to the types of behaviors exhibited by students in the context of estimated knowledge, the final analysis offers an opportunity to observe these groupings by opportunity as well. Figure 6 depicts the results of this comparison, observing the distribution of student assignments that belong to each cluster by opportunity. Cluster 3 is found to have the fewest overall students proportionally in the first three opportunities; as this is not the smallest cluster when

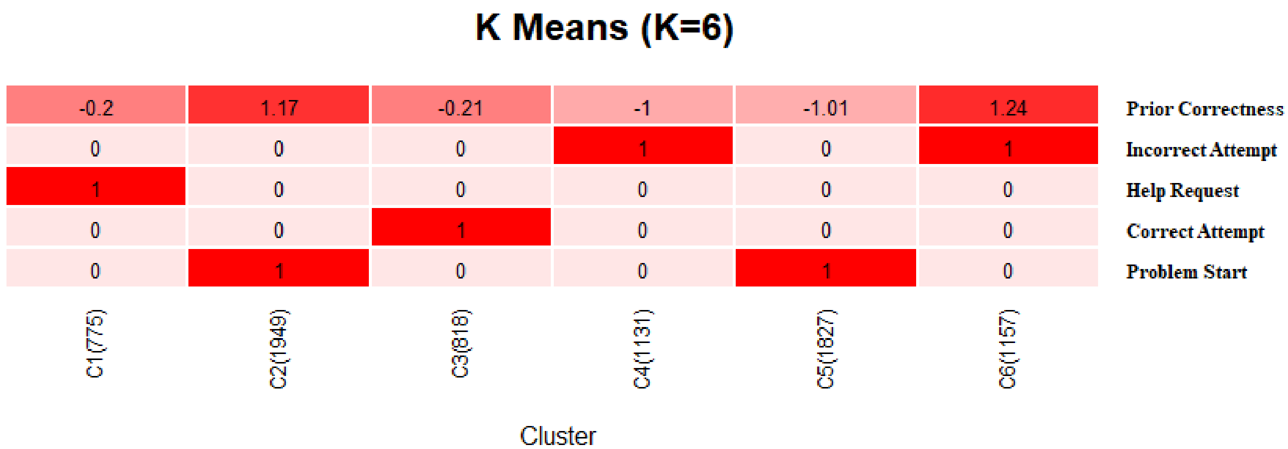

Figure 5: The resulting clusters of student prior correctness and last action pertaining to student stopout.

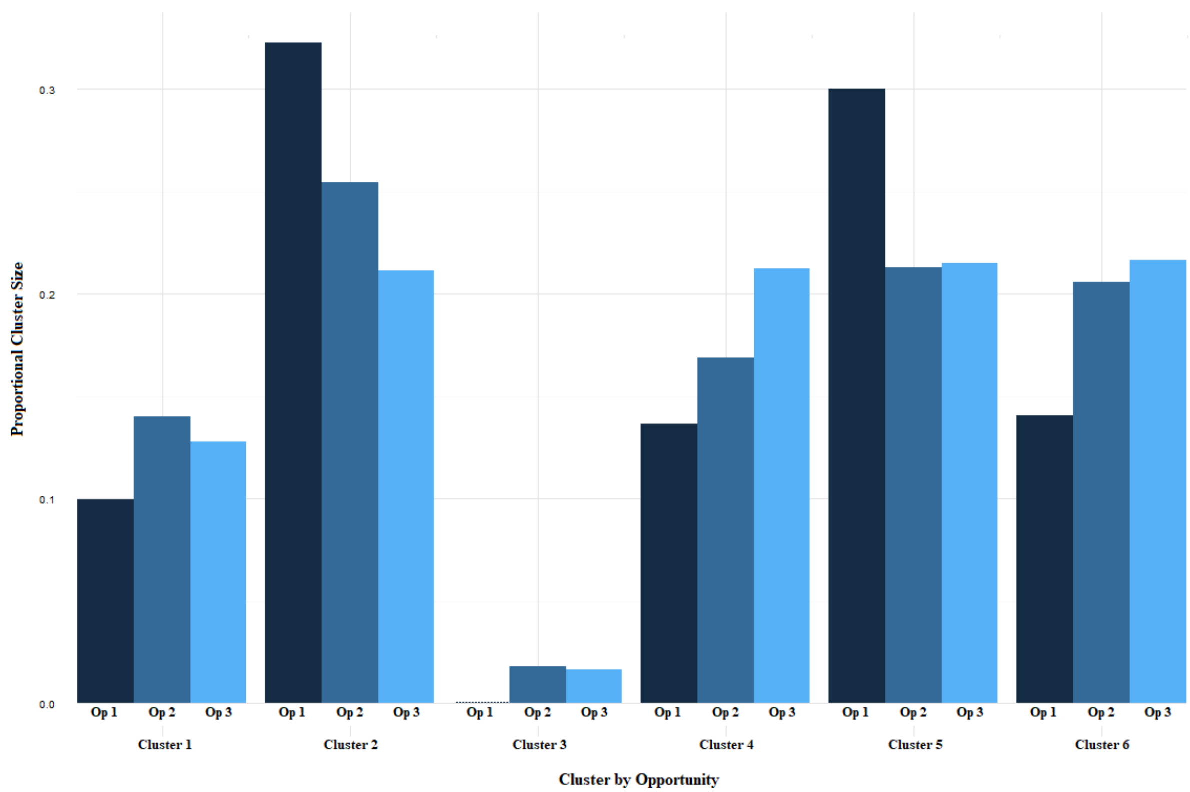

Figure 6: The proportional distribution of samples within each cluster over the first three learning opportunities.

observing all student assignments, this suggests that this behavior is exhibited more on later opportunities. It is also the case, due to our definition of stopout, that no student can stopout on the first opportunity following a correct response. Aside from this, cluster 1 similarly contains the fewest number of students that also appears to be less affected by opportunity as no clear trend emerges within this cluster.

The remaining four clusters, however, do exhibit interesting trends over the first three opportunities. Clusters 4 and 6 exhibit increasing numbers of students stopping out following incorrect

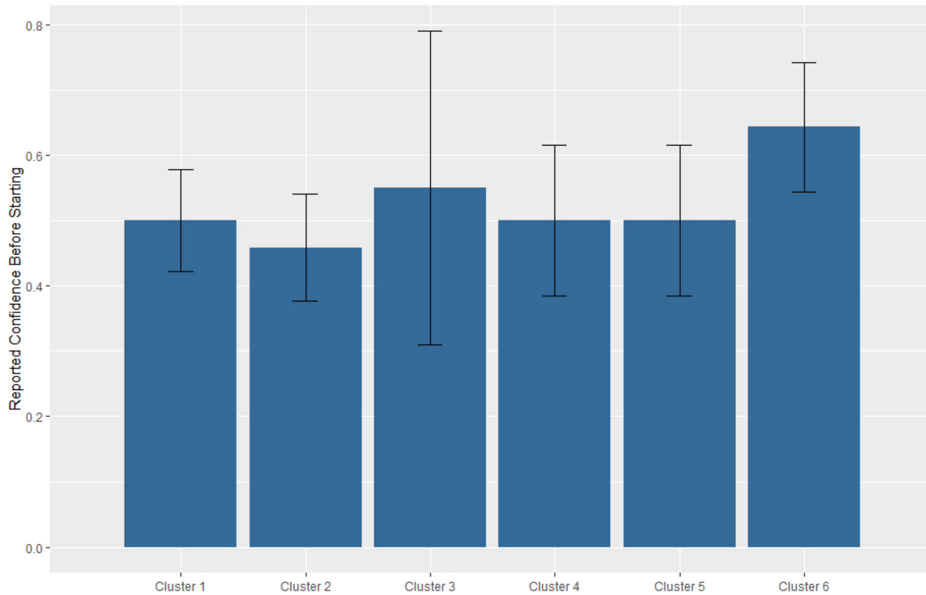

Figure 7: The reported confidence of students within each cluster with associated 95% confidence intervals.

responses, though distinguishable by the estimated knowledge level of students found within these clusters. Cluster 2 conversely exhibits a decreasing number of high knowledge students exhibiting stopout at the start of a problem before taking any further action. Finally, cluster 5 contains a notable trend in that the number of low knowledge students stopping out on the first item before taking action is noticeably higher than subsequent opportunities and exhibits no increasing or decreasing trend beyond this point within the observed opportunities. For this reason, it is likely that the cause for the disproportionate stopout on the first learning opportunity is largely due to students within clusters 2 and 5; these, again, are the students exhibiting refusal by our definition. Furthermore, the number of students who fall within clusters 2 and 5 on the first learning opportunity are 1,025 and 954, respectively, which, when subtracted from the total number of 3,076 students who exhibited stopout on the first opportunity as illustrated in Figure 2, the resulting 1,097 falls much closer to the expected 1,371 students as determined by our fit exponential model described in Section 4.1. We are not attempting to claim, of course, that this simple comparison of sample sizes fully explains the observed disproportionate stopout exhibited on the first learning opportunity, but the results of our analyses coupled with these comparisons do suggest that refusal behavior accounts for a majority of the phenomenon.

It is found, comparing the results of both the clustering analysis and comparison of cluster distributions across learning opportunities, that the disproportionate stopout tends to occur regardless of knowledge level, at the beginning of the problem before taking any action. This problem becomes more perplexing considering the effort to remove optional assignments using a completion threshold during data collection and filtering. Assuming that at least a majority of optional assignments and outlier cases are removed during that cleaning process, the fact that the two largest clusters are still comprised of those students who stopout without taking action further stresses the need to understand the definitive causes of such behavior.

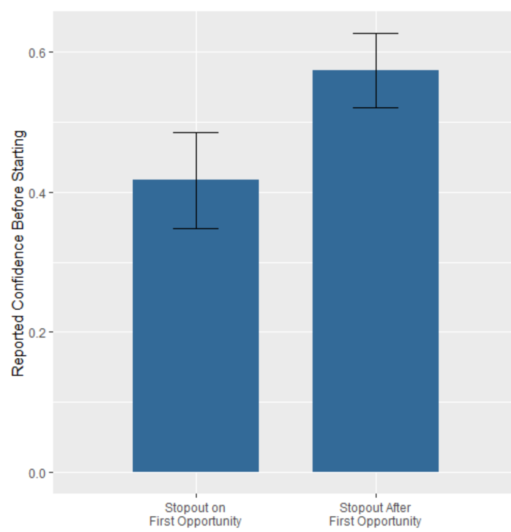

Figure 8: The reported confidence of students who stopout on the first learning opportunity as compared with students who stopout after the first learning opportunity with associated 95% confidence intervals.

The results of our final analyses are depicted in Figures 7 and 8, comparing the reported confidence measures of students by both cluster (Figure 7) and first opportunity versus subsequent opportunities (Figure 8). As the number of students who exhibited stopout in this supplementary dataset is significantly less than that observed in our earlier analyses, the 95% confidence intervals vary greatly. In observing Figure 7, for example, the majority of intervals overlap making us unable to claim reliable differences between many of the clusters. However, two clusters, 2 and 6, do emerge as significantly different with regard to the level of reported confidence. These two clusters represent the highest performing students compared to other clusters and yet exhibit vastly different levels of confidence, with the lower confident students being those who stopout without making any action in the problem. It is important to clarify that this figure includes students who stopout across all opportunities and not specifically those who stopout on the first opportunity (e.g. Cluster 2 here is not specifically students exhibiting refusal). It is also important to recognize that all reports of student confidence are reliably smaller than 0.8 (and several being even lower), suggesting that a large number of students who exhibited stopout, unsurprisingly, were not confident in their ability to successfully complete the assignment.

Figure 8 illustrates a significant difference found between the reported confidence of students who exhibit stopout on the first opportunity as compared to students who stopout on subsequent opportunities. It is important to clarify, however, that this comparison includes all students who stopout on the first opportunity in a single group as opposed to comparing students specifically exhibiting refusal (i.e. stopping on the first opportunity after taking no action) as it was found that very few students exhibited refusal in the supplementary dataset (only 4 students were found). This is contrary to the proportion that was found in other skill builder dataset, but may be attributable to the context of the study; we believe refusal may occur as students realize that they are not confident in their ability to successfully complete the assignment, and as their confidence is revealed by the survey item, it is likely that students who would have exhibited refusal simply never began the assignment and subsequently would not exist in our dataset (as they saw no learning opportunities of the assignment). Despite this, we still see a significant difference between students who stopout on the first opportunity when compared to stopout on subsequent opportunities, suggesting that confidence, perhaps even more so than knowledge (in considering clusters 2 and 6 in Figure 5), is associated with refusal and early stopout behavior in student assignments.

6 CONTRIBUTIONS AND FUTURE WORK

The current work represents an initial step toward better understanding the causes of student stopout in K-12 classrooms by exploring the student actions and attributes associated with such behavior. With this in mind, this work can act as a foundation for future research aimed at finding more causal links between behavior and stopout as. A simple approach, as the students do not drop out of the respective courses, would be to survey students to determine the reasons for stopping out of an assignment.

There are several limitations to the current work that can be addressed with further research as well. The first is in the scope of the behaviors considered for grouping student assignments. In the analyses presented in this work, only the last action taken by the student was considered within the clustering. This feature can be vastly improved by generating more descriptive features of student activity or even by utilizing earlier information pertaining to each student. Another limitation of the current work is the lack of contextual information pertaining to each assignment. The clustering is performed observing only student attributes as it is believed that this is most important to understand the behaviors associated with stopout, but understanding how these attributes interact with assignment-level features, such as the difficulty of the subject matter, may be helpful to understanding the concept as well.

Another limitation of the current work is the lack of causality of our analyses. While it is among the goals of this work to identify potential causes of stopout and refusal behavior, all analyses conducted are limited to correlation rather than causal claims. Future work may be able to address this by conducting randomized controlled trials aimed at identifying and deploying interventions to prevent potential stopout and refusal behaviors.

The contributions of the current work are 3-fold toward understanding the behaviors and actions associated with student assignment-level attrition in K-12 classrooms. First, the current work identified a disproportionate stopout on the first opportunity as compared with subsequent opportunities. While stopout tends to follow an exponential decay, this does not extend to the first learning opportunity. This highlights a need to research this phenomenon further to direct the development of learning interventions aimed at deterring students from giving up to early or too easily when faced with difficult content. We show in this work that a large proportion of this early stopout is likely attributable to a behavior we have identified as refusal.

The second contribution is in the exploration of student actions associated with stopout. With the 6 groups of student knowledge-action interactions that emerged from the analysis, these clusters form the basis to conduct further research exploring their predictive power in other aspects of student learning. These groups of students highlight that low persistence, as defined by student stopout, is not exhibited in the same way across all students or even across students of similar prior knowledge. Furthermore, the actions associated with stopout behavior are found to change over each learning opportunity, suggesting that, unsurprisingly, the reason for stopout is dependent on where the behavior occurs within each assignment.

Finally, it is clear from this work, as well as the work of Lang et al. [13], that confidence is strongly related to student assignment-level attrition, perhaps even more so than gaps in student knowledge, supporting the need for learning interventions to address this factor to promote more productive learning practices. This confidence level, while comparatively low for all students who exhibited stopout in our analyses, appeared lowest for students who exhibited stopout behavior on the first learning opportunity. Similarly, the level of confidence for high knowledge students was divided between two of the identified clusters of students, suggesting that confidence is not directly dependent on prior knowledge.

ACKNOWLEDGMENTS

We want to thank multiple NSF grants (e.g., ACI-1440753, DRL-1252297, DRL-1109483, DRL-1316736, DGE-1535428, OAC-1724889,OAC-1636782, 1822830 & DRL-1031398), the US Department of Education Institute for Education Sciences (e.g., IES R305A180401, R305A120125, R305A180401 & R305C100024) and the Graduate Assistance in Areas of National Need program (e.g., P200A180088 & P200A150306), and the Office of Naval Research (N00014-18-1-2768).

REFERENCES

[1] Angela L Duckworth, Christopher Peterson, Michael D Matthews, and Dennis R Kelly. Grit: perseverance and passion for long-term goals. *Journal of personality and social psychology*, 92(6):1087, 2007.

[2] Christopher Peterson, Martin EP Seligman, et al. *Character strengths and virtues: A handbook and classification*, volume 1. Oxford University Press, 2004.

[3] Carol S Dweck, Gregory M Walton, and Geoffrey L Cohen. Academic tenacity: Mindsets and skills that promote long-term learning. *Bill & Melinda Gates Foundation*, 2014.

[4] Hiroko Kawaguchi Warshauer. Productive struggle in middle school mathematics classrooms. *Journal of Mathematics Teacher Education*, 18(4):375–400, 2015.

[5] Manu Kapur. Productive failure. *Cognition and instruction*, 26(3):379–424, 2008.

[6] Joseph E Beck and Yue Gong. Wheel-spinning: Students who fail to master a skill. In *International Conference on Artificial Intelligence in Education*, pages 431–440. Springer, 2013.

[7] Devendra Singh Chaplot, Eunhee Rhim, and Jihie Kim. Predicting student attrition in moocs using sentiment analysis and neural networks. In *AIED Workshops*, 2015.

[8] Wanli Xing, Xin Chen, Jared Stein, and Michael Marcinkowski. Temporal predication of dropouts in moocs: Reaching the low hanging fruit through stacking generalization. *Computers in Human Behavior*, 58:119–129, 2016.

[9] Diyi Yang, Tanmay Sinha, David Adamson, and Carolyn Penstein Rosé. Turn on, tune in, drop out: Anticipating student dropouts in massive open online courses. In *Proceedings of the 2013 NIPS Data-driven education workshop*, volume 11, page 14, 2013.

[10] Carolyn Penstein Rosé, Ryan Carlson, Diyi Yang, Miaomiao Wen, Lauren Resnick, Pam Goldman, and Jennifer Sherer. Social factors that contribute to attrition in moocs. In *Proceedings of the first ACM conference on Learning@ scale conference*, pages 197–198. ACM, 2014.

[11] Anne Lamb, Jascha Smilack, Andrew Ho, and Justin Reich. Addressing common analytic challenges to randomized experiments in moocs: Attrition and zero-inflation. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pages 21–30. ACM, 2015.

[12] GJ McLachlan and DC McGiffin. On the role of finite mixture models in survival analysis. *Statistical Methods in Medical Research*, 3(3):211–226, 1994. PMID: 7820292.

[13] Charles Lang, Neil Heffernan, Korinn Ostrow, and Yutao Wang. The impact of incorporating student confidence items into an intelligent tutor: A randomized controlled trial. *International Educational Data Mining Society*, 2015.

[14] René F Kizilcec and Sherif Halawa. Attrition and achievement gaps in online learning. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pages 57–66. ACM, 2015.

[15] Pedro A Willging and Scott D Johnson. Factors that influence students' decision to dropout of online courses. *Journal of Asynchronous Learning Networks*, 13(3):115–127, 2009.

[16] Mike Sharkey and Robert Sanders. A process for predicting mooc attrition. In *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, pages 50–54, 2014.

[17] Yue Gong and Joseph E Beck. Towards detecting wheel-spinning: Future failure in mastery learning. In *Proceedings of the Second (2015) ACM Conference on Learning@ Scale*, pages 67–74. ACM, 2015.

[18] Neil T Heffernan and Cristina Lindquist Heffernan. The assistments ecosystem: Building a platform that brings scientists and teachers together for minimally invasive research on human learning and teaching. *International Journal of Artificial Intelligence in Education*, 24(4):470–497, 2014.

[19] Jeremy Roschelle, Mingyu Feng, Robert F Murphy, and Craig A Mason. Online mathematics homework increases student achievement. *AERA Open*, 2(4):2332858416673968, 2016.