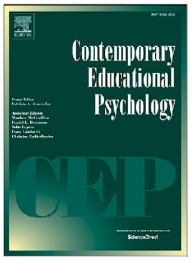
Journal Pre-proof



Not all who avoid help are lost: The relations between help avoidance and performance in a digital learning platform

Natalia S. Martin, Nicholas A. Vest, Avery H. Closser, Anthony F. Botelho

PII: S0361-476X(26)00063-9
DOI: https://doi.org/10.1016/j.cedpsych.2026.102505
Reference: YCEPS 102505

To appear in: *Contemporary Educational Psychology*

Please cite this article as: N.S. Martin, N.A. Vest, A.H. Closser et al., Not all who avoid help are lost: The relations between help avoidance and performance in a digital learning platform. *Contemporary Educational Psychology* (2026), doi: https://doi.org/10.1016/j.cedpsych.2026.102505.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Published by Elsevier Inc.

Journal Pre-proof

Manuscript without author details

Click here to view linked References

Not All Who Avoid Help Are Lost: The Relations Between Help Avoidance and Performance in a Digital Learning Platform

Abstract

Although digital learning platforms (DLPs) provide students with corrective feedback, hints, and scaffolds, learners differ substantially in whether and when they use these supports. While help seeking has been extensively studied, there is little consensus on how to define and operationalize help avoidance (i.e., when students bypass available supports like hints) in online settings, hindering efforts to reliably identify, measure, and address it. This study revisits an established operationalization of help avoidance using fine-grained interaction logs from middle school students (N = 348) in ASSISTments, a digital math learning platform. Building on prior work, we refine the measure by leveraging detailed temporal interaction data to characterize help avoidance as unassisted answer attempts by students with low prior knowledge before accessing available support. We then examine how these behavioral patterns relate to proximal outcomes (i.e., within ASSISTments) and distal math achievement (i.e., on a standardized math test). Our analyses identify distinct subgroups of students classified as help avoidant, suggesting that existing operationalizations combine qualitatively different behaviors, including patterns consistent with productive struggle and unproductive persistence. These findings demonstrate that help avoidance is not a homogeneous construct and suggest that more behaviorally nuanced measures are necessary to differentiate productive struggle from a genuine need for intervention, thus guiding adaptive support in DLPs.

Keywords: Educational Technology; Help Avoidance; Hints; Mathematics; Performance; Learning; Middle School

Journal Pre-proof

Journal Pre-proof

Introduction

While feedback is widely recognized as a driver of learning (Hattie and Timperley, 2007; Tricomi and DePasque, 2016), its effectiveness depends on how learners navigate difficulty and use available supports (Daumiller and Meyer, 2025; Kluger and DeNisi, 1996). Choosing to use these supports is associated with improved student outcomes (Boud and Molloy, 2013; Nelson-Le Gall, 1981). However, many students avoid or delay help even when it is available (Aleven and Koedinger, 2000), a behavior previously linked to poorer learning gains (Ryan and Pintrich, 1997).

The decision to seek or avoid help is grounded in self-regulated learning (SRL), by which students plan, monitor, and evaluate their learning (Zimmerman, 2002). Therefore, help avoidance may serve as a behavioral indicator of differences in SRL processes such as comprehension monitoring and error analysis (Karabenick and Berger, 2013). Digital learning platforms (DLPs) that provide on-demand support (i.e., hints and scaffolds) make these processes visible. Yet, there is no consistent approach to measuring help avoidance as a proxy for SRL in these environments, and the timing of these decisions, including how long students persist and when avoidance becomes detrimental, has received limited attention.

This study addresses this gap by defining and measuring help-avoidance behaviors in ASSISTments (www.assistments.org), a widely used DLP for math education, and examining how the timing of avoidance relates to student performance. We combine fine-grained log data from a large-scale efficacy study (Roschelle et al., 2016) with TerraNova scores (i.e., a standardized math achievement sub-test) to analyze temporal patterns of help avoidance in the classroom. The analysis focuses on whether persistence without support reflects productive self-regulation or behavioral patterns associated with lower performance.

Journal Pre-proof

Journal Pre-proof

Literature Review

Help Seeking

Help seeking is a social-cognitive process involving a more knowledgeable other, which may be either human or non-human (Lee and List, 2023; Makara and Karabenick, 2013; Zimmerman, 2000; Zusho and Barnett, 2011). It begins when a learner encounters an impasse, recognizes they cannot solve the problem independently, identifies the need for support, seeks assistance, and reflects on the outcome (Karabenick and Berger, 2013; Nelson-Le Gall, 1981; Newman, 2002).

Within SRL, help seeking is understood as a metacognitive strategy that connects internal monitoring to external action (Ryan and Pintrich, 1997; Schunk and Zimmerman, 1998; Zusho and Barnett, 2011). Learners must recognize gaps in understanding and respond by accessing available resources to continue progress (Karabenick and Berger, 2013; Schenke et al., 2015; Zusho and Barnett, 2011). Because this process depends on accurately identifying when assistance is needed, effective help seeking requires active engagement. Prior work has linked help seeking to improved comprehension and academic performance (Ryan and Pintrich, 1997), as well as the development of independent problem-solving skills (Nelson-Le Gall, 1981), transfer (Hattie and Timperley, 2007), and self-confidence (Butler, 1998).

Despite these benefits, some learners avoid available support by bypassing resources such as hints in DLPs or the opportunity to ask questions in class. Decisions to seek or avoid help reflect a combination of individual, environmental, and cognitive factors (Ryan and Pintrich, 1997; Schenke et al., 2015; White and Bembenutty, 2013). Classroom context plays a role in shaping these decisions. When errors are treated as failures, students may avoid help to protect their social standing (Turner et al., 2002; Yang et al., 2025). A parallel pattern appears in feedback research;

Journal Pre-proof

Journal Pre-proof

when feedback draws attention to the self rather than the task, students disengage and learn less, especially after repeated negative feedback (Fyfe et al., 2026). Similarly, studies of medical students interacting with a DLP show that learners may persist with incorrect reasoning rather than use available support, even when they recognize gaps in their understanding (Fischer et al., 2002). In many cases, students with the greatest need for assistance are the least likely to seek it (Karabenick and Knapp, 1991; Ryan and Pintrich, 1997).

Social relationships and motivational factors further influence help-seeking behavior. Students may seek help to strengthen relationships with supportive instructors (Karabenick and Berger, 2013), or avoid it to maintain peer acceptance in contexts where help seeking is stigmatized (Butler, 1998). Motivational orientation also shapes these decisions. Learners with a mastery orientation are more likely to view help seeking as part of the learning process, whereas those with a performance orientation may avoid it to protect self-perception (Zusho and Barnett, 2011). To address these patterns, Aleven et al. (2006) proposed a framework for operationalizing help avoidance as an observable behavior within DLPs.

Help Avoidance

Prior work defines help avoidance as instances in which a student attempts a step without sufficient mastery, makes an error, and does not use available support (Aleven et al., 2003; Ryan et al., 2005). The consequences of this behavior depend on when it occurs, as persistence without support can hinder learning when students internalize misconceptions and fail to correct errors (Butler, 1998; Karabenick and Knapp, 1991).

At the same time, delaying support does not always indicate ineffective behavior. While early intervention can reduce unproductive persistence (Beck and Gong, 2013), there is also evidence

Journal Pre-proof

that delayed support can promote deeper engagement (Clark and Bjork, 2014). Struggling before receiving assistance can lead to improved outcomes compared to immediate help, a pattern described as productive failure (Kapur, 2008). In these cases, learners use errors to identify gaps in understanding and adjust their approach before seeking support (Zimmerman, 2002). This pattern of sustained engagement suggests that some behaviors classified as help avoidance may reflect purposeful persistence rather than ineffective regulation (Mullet et al., 2014).

These contrasting patterns indicate that help avoidance, as currently defined, may capture multiple forms of behavior. Some cases reflect persistence without sufficient knowledge, while others reflect deliberate effort prior to seeking support. This distinction raises concerns about whether existing measures accurately represent the underlying processes they are intended to capture.

Feedback, Additional Supports, and Self-Regulated Learning

Seeking help in a physical classroom differs significantly from seeking help in DLPs. In a classroom setting, seeking support is heavily mediated by social and contextual dynamics (e.g., peer evaluation, classroom culture) (Ryan et al., 2001; Turner et al., 2002). In contrast, DLPs mitigate these social threats by shifting help seeking from a public interpersonal interaction into a private, low-stakes individual decision (Makara and Karabenick, 2013)

Within DLPs, instructional assistance often consists of a combination of automated and elective supports. Narciss and Huth (2004) define feedback as information provided after a response to support the development of cognitive, behavioral, or attitudinal strategies. Some elective supports (e.g., hints, scaffolds) are context-sensitive instructional interventions (Aleven et al., 2006) that learners choose to access (Price et al., 2017). These supports guide students through the problem-solving process by reducing the gap between current understanding and task demands (Hattie and

Journal Pre-proof

Timperley, 2007), while also providing an observable record of when assistance is used.

Automated feedback and elective supports shape how students engage with problem solving. Feedback supports knowledge acquisition through error correction, while hints and scaffolds provide structured guidance that can support metacognitive processes (Puustinen, 1998; Zusho and Barnett, 2011). This distinction allows for the analysis of how students navigate difficulty during problem solving, including cases in which available supports are not used.

Although both feedback and elective supports are widely available in DLPs, they have been studied differently. Feedback has been extensively examined, including work on how its content and timing affects learning (Fong, 2025; Fyfe et al., 2021; Mullet et al., 2014). In contrast, less attention has been given to when and how students choose to use elective supports.

ASSISTments: A Shared Research Infrastructure

ASSISTments (Heffernan and Heffernan, 2014) is a K-12 mathematics digital learning platform used primarily for classwork and homework. It can be used to supplement curricula or integrated into teachers' existing curricula and learning management system. ASSISTments provides students with a variety of practice and assessment activities, with automated scaffolding, grading, and performance reports for teachers and students. Used widely across the United States and internationally, its integration has been associated with improvements in mathematics learning relative to typical classroom practices (Feng et al., 2023; Murphy et al., 2020).

As a shared research infrastructure, ASSISTments records time-stamped student actions during regular classroom use. Routinely available data consist of fine-grained interaction logs (e.g., attempt sequences, hint requests, duration, and correctness), which are de-identified and made available to external researchers through open-access repositories or research collaborations under

Journal Pre-proof

institutional review board protocols. In particular, ASSISTments provides E-TRIALs to support external research studies embedded within the platform (Ostrow et al., 2017). Within the broader educational research ecosystem, ASSISTments data have traditionally been used to study knowledge tracing (Koedinger et al., 2010; Wang and Heffernan, 2011), cognitive modeling (Feng and Heffernan, 2006), pedagogical scaffolding (e.g., immediate vs. delayed feedback) (Smith et al., 2025), behavioral patterns such as wheel spinning (Beck and Gong, 2013), and large-scale efficacy trials (Roschelle et al., 2016). Non-behavioral measures, self-report surveys, or district-provided measures (such as the TerraNova assessment scores used here) are not natively embedded in standard logs and must be merged or collected separately through specific research protocols.

While completing practice problems, students can regularly request hints, attempt solutions, and access scaffolded sub-problems, with a bottom hint providing the correct answer when needed. Mastery-based skill builders (i.e., problem sets designed for targeted skill acquisition) require students to demonstrate proficiency through consecutive correct responses. These features generate consistent records of how students engage with problems, including the timing and sequence of attempts and support use.

Skill builders offer optimal conditions for studying constructs such as help avoidance, where the same observable behavior may reflect different underlying processes. In skill builders, students know that the objective is to answer three or more problems in a row correctly in order to complete (or "master") the assignment and hints are available within most problems. The process-level data from these activities provide a basis for examining how students' choices about when to select (or avoid) help through hints is related to their performance on proximal and distal outcomes. We chose this dataset because its fine-grained granularity allows us to closely examine student interactions with hints and scaffolds, revealing distinct behavioral patterns. Beyond the current

Journal Pre-proof

analytic sample (N = 348), leveraging such an expansive infrastructure creates opportunities to extend this research and evaluate its generalizability across varied educational settings.

Current Study

This study examines how help-avoidance behavior is defined and measured in ASSISTments and how the timing of these behaviors relates to student performance. Using a publicly available dataset from a large-scale efficacy study (Roschelle et al., 2016), we analyze data from 348 middle school mathematics students in the Northeastern United States. The analysis uses problem-level records of student interactions, including time-stamped actions, prior knowledge estimates, and hint request behavior. Prior knowledge is incorporated using a mastery threshold of 0.43 (i.e., the likelihood a student has mastered a skill) motivated by Aleven et al. (2006). Mastery threshold was operationalized as a prior knowledge proxy using Bayesian Knowledge Tracing (BKT). Because BKT estimates latent skill mastery while accounting for guess and slip probabilities, crossing the mastery threshold provides a mathematically reliable indicator of established domain competency. These data allow us to examine how students persist, when they request support, and how these patterns relate to proximal outcomes within the platform and distal outcomes on standardized tests.

The study is guided by the following research questions: **(1)** How can help avoidance be operationalized in online learning environments and how does students' probability-of-knowledge estimate affect this measure? **(2)** How often do students exhibit help avoidance under a probability-of-knowledge threshold-based definition? **(3)** What patterns of time and problem-solving behavior are observed within help-avoidance cases? **(4)** How are help-avoidance behaviors, including their timing and structure, related to proximal and distal performance outcomes?

Prior work suggests help-avoidance behavior may not be fully captured by existing definitions.

Journal Pre-proof

We expect that current measures may group together behaviors that differ in how students engage with problems. Particularly, persistence without support may reflect sustained effort or a failure to recognize the need for help. By incorporating measures of time, attempts, and prior mastery, this study examines whether these differences can be distinguished and how they relate to performance.

Methods

Compliance with Ethical Standards

All procedures were conducted in compliance with relevant laws and institutional guidelines and were approved by [Institution blinded for review] Institutional Review Board on 10/24/2025 (Protocol blinded for review).

Dataset

This study uses ASSISTments log data collected during the 2013–2014 and 2014–2015 academic years from students who participated in the Maine ASSISTments efficacy trial (Roschelle et al., 2016). Students and teachers participated for one academic year within one of two cohorts, each situated in seventh-grade math classrooms. The data reflect regular in-classroom platform use rather than a short-term research intervention. This context is relevant because the present study examines how an operationalization of help avoidance behaves under typical instructional conditions, where students encounter varied problem types and teachers determine assignments.

Teachers used ASSISTments for required classwork and homework alongside their existing curricula, which drew from multiple textbooks. The platform delivered content aligned with these materials, allowing teachers to assign problems corresponding to ongoing instruction. Skill Builders (mastery-focused problem sets requiring a series of consecutive correct answers to prove

Journal Pre-proof

skill mastery) were not required by the study protocol, and teachers varied in how frequently they assigned them. This detail is relevant as the analyses depend on students encountering problems with available hints. Although assignment was discretionary, many teachers used Skill Builders regularly, resulting in a sufficient number of hint-enabled problem interactions for analysis.

Platform data were linked to end-of-year TerraNova math assessment scores, which serve as the distal outcome measure. The TerraNova is a widely recognized, nationally norm-referenced standardized assessment used to evaluate overall student achievement in core subject areas, including middle school math. These scores do not capture immediate problem-level performance or help-seeking behavior directly. Instead, they provide a common external measure of academic performance that can be compared with student-level summaries of platform interactions.

From an initial dataset of 12,857 students with 29,659 completed problems, filtering for matching start-end dates and hint availability yielded 7,985 students and 13,825 problems. A second filter restricted interaction time to under 20 minutes (e.g., attempt or hint-request clicks), targeting actions immediately preceding help requests. This step retained all 7,985 students and 13,825 problems. Finally, merging these platform logs with non-null TerraNova assessment records produced an analytic sample of 348 unique students and 1,802 problems, with interaction times under 20 minutes. TerraNova raw scores had a mean of 692.4 and a median of 698.0, with an interquartile range from 677.8 to 721.0 and a range from 487 to 848. The distribution was approximately normal.

Problem-Level Measures and Preprocessing

The data consist of action-level ASSISTments logs, where each record represents a single student action on a problem. These actions include answer attempts and requests for help in the form

Journal Pre-proof

Journal Pre-proof

of on-demand hints. Because the operationalizations examined in this study depend on whether a student attempts a problem before requesting help, the action-level records provide the temporal detail needed to recover this sequence. The data were aggregated to the student–problem level, which serves as the unit of analysis for all subsequent measures.

To study help avoidance, the dataset was restricted to problems with hints available. This restriction ensures that help-seeking is possible and that comparisons can be made between instances where students request support and those where they do not. The retained observations therefore correspond to student–problem instances in which help could be accessed.

Aggregation to the student–problem level produced measures of how a student engaged with a problem prior to, or in the absence of, a help request. For each instance, we retained whether a hint was requested, the number of attempts made before the first hint request, the elapsed time before the first hint request, and total time on the problem. For problems in which no hint was requested, the attempt and time variables correspond to total attempts and total elapsed time. This approach preserves information about persistence and latency even when students do not seek help, which is central to evaluating whether similar classifications reflect different patterns of behavior.

Each student–problem instance was associated with a probability-of-knowledge estimate, denoted $P(\mathrm{Know})$. This measure was derived from a Bayesian Knowledge Tracing model (Corbett and Anderson, 1994) fit in prior work and applied to the present data (Botelho, 2019). Conceptually, $P(\mathrm{Know})$ represents the estimated likelihood that a student has mastered the relevant skill prior to beginning the problem. This allows for distinction between cases in which a student attempts before seeking help while likely lacking knowledge and those in which similar behavior may reflect greater preparedness. Because $P(\mathrm{Know})$ is continuous, identifying instances of "low knowledge" depends on the choice of a threshold. Prior work has shown that this choice affects

Journal Pre-proof

how help avoidance is measured (Almeda et al., 2017), which motivates our approach.

Replication of Threshold-Based Help Avoidance

Aleven et al. (2006) introduced a problem-level operationalization of help avoidance intended to capture cases in which a student who likely needs help does not seek it before attempting a problem. Under this definition, help avoidance occurs when a student's estimated probability of knowledge falls below a specified threshold and the student makes an attempt before requesting help. In Aleven et al., the threshold was set at $P(\text{Know}) < 0.6$. This operationalization is simple and interpretable, but the choice of threshold is consequential because it determines which student–problem instances are treated as cases where help *should* have been sought.

Almeda et al. (2017) revisited this by treating the threshold as an empirical choice rather than a fixed constant. Using the same general problem-level rule, they varied the threshold across various candidate values. For each, they aggregated problem-level labels to the student level as the proportion of problems on which each student was classified as exhibiting help avoidance, and then examined how that student-level measure correlated with several outcome variables. Their interpretation focused not on the threshold yielding the largest correlation, but on how the correlation pattern changed across thresholds. They observed that correlations appeared more variable at lower thresholds and more stable above approximately 0.6. On that basis, they argued a threshold near 0.6 was a reasonable choice.

This progression from Aleven et al. to Almeda et al. is central here as it identifies the threshold as a substantive component of the operationalization rather than an arbitrary implementation detail; also suggesting a particular selection logic. Rather than identifying the threshold that maximizes predictive association with an outcome, the goal is to identify a threshold above which the

Journal Pre-proof

resulting measure behaves consistently. This ensures threshold changes no longer materially alter the observed relationship with outcomes. That logic motivated our replication analysis.

To operationalize help avoidance, we evaluated candidate knowledge thresholds across the probability scale from 0.05 to 1.00 in increments of 0.01. For each candidate threshold, problem-level help avoidance was defined following Aleven et al. (2017), and we aggregated to the student level as the proportion of hint-enabled problems on which the student exhibited help avoidance. We then computed the Pearson correlation between these student-level proportions and TerraNova scores across candidate thresholds, producing a correlation profile. A complete summary of all variables, log metrics, and operational decision rules is provided in Table 1.

To select the threshold, we formalized Almeda et al.'s (2017) stability logic by identifying the cut point in the correlation profile that best separated a variable lower region from a stable upper region. Specifically, candidate thresholds were scored by evaluating the ratio of variance in correlations below versus above the cut point, combined with a balance penalty to avoid extreme edge partitions. The selected threshold was the candidate value that maximized this stability criterion, yielding an optimal threshold of $\tau = 0.43$, which was then used to instantiate the help-avoidance measure in subsequent analyses. Full formal definitions, aggregation equations, and scoring criteria are provided in Appendix A.

Application of the Derived Measure and Analytic Strategy

Following the replication analysis, the threshold of $\tau = 0.43$ was applied to the ASSISTments data to generate a problem-level help-avoidance label for each student–problem instance. TerraNova scores were standardized using a z-score transformation to place the outcome on a common scale. This facilitates comparison of coefficients across models and simplifies interpretation of

Journal Pre-proof

Journal Pre-proof

direction and relative magnitude.

To examine the behaviors surrounding this measure of help avoidance in the context of student performance we conduct two exploratory analyses. The first evaluates whether the derived measure is associated with student performance. This assesses whether the measure captures information beyond general patterns of engagement or correctness. The second examines whether behaviors classified as help avoidance form a homogeneous population. This follows from the concern that the operationalization may group together student actions that differ in how learners engage with problems prior to or instead of requesting support.

Analysis 1: Outcome Relevance at the Student Level

The first analysis examined the association between student-level help avoidance and standardized TerraNova performance. For each student, we computed the proportion of problems classified as help avoidance, along with total problems, total attempts, and total log-transformed time. From these, we derived average attempts and average log time per problem. These measures were combined with TerraNova scores and additional behavioral summaries, including average hint usage and correctness. These variables serve as alternative indicators of student engagement and performance, allowing the analysis to assess whether the help-avoidance proportion captures information beyond broader patterns of platform use.

We estimated a sequence of linear regression models predicting standardized TerraNova scores. The first model included only the help-avoidance proportion. The second model added average hint usage and average correctness. The third model further included average log time per problem, average attempts per problem, and the total number of problems completed. This staged approach allows for evaluation of whether the association between help avoidance and TerraNova perfor-

Journal Pre-proof

Journal Pre-proof

mance persists after accounting for more general measures of engagement and performance. In this context, regression is used to assess the explanatory contribution of the derived measure rather than to optimize prediction.

Analysis 2: Behavioral Heterogeneity Within Help Avoidance

The second analysis examined whether observations classified as help avoidance form a behaviorally consistent category. This follows from the concern that a threshold-based measure may group together students whose behaviors differ in how they engage with problems prior to or instead of requesting support. To examine this possibility, we restricted the dataset to problem instances labeled as help avoidance and summarized each student's behavior within that subset.

For each student, we computed the mean and variance of $P(\mathrm{Know})$, number of attempts, and time spent on help-avoidance problems. We also calculated the proportion of problems falling into coarse categories, such as multiple attempts or longer time on task. These summaries capture both typical behavior and variation within help-avoidance cases, allowing for assessment of whether these instances are consistent or reflect mixed patterns of engagement.

The resulting student-level summaries were merged with standardized TerraNova scores and used to fit a regression tree. The tree was used as a subgroup discovery tool rather than a predictive model. This approach allows for identification of behavioral configurations that may arise through interactions among time, attempts, and estimated knowledge without requiring specification of these interactions in advance.

After identifying tree-defined subgroups, we examined them in two ways. First, we inspected the partition structure to characterize the behavioral profiles present among students whose cases were classified under the same rule. Second, we estimated linear regression models that included

Journal Pre-proof

subgroup indicators along with selected continuous behavioral predictors. This step evaluates whether the identified subgroups differ in their association with TerraNova performance. Together, these analyses assess whether the threshold-based operationalization captures a single form of behavior or a set of distinct patterns with different outcome relationships.

Results

Replication of threshold-based help avoidance

Applying the replication procedure described in Replication of Threshold-Based Help Avoidance, the variance-based selection criterion identified $P(\text{Know}) = 0.43$ as the threshold in the present dataset, which resulted in 128 identified help-avoiders within the analytic sample (N = 348). Figure 2 shows the correlation between the student-level proportion of help avoidance and TerraNova score across candidate thresholds, along with the variance-balance score used to select the threshold. The score reaches its maximum at 0.43, indicating that this value separates a lower-threshold region with greater variation in correlations from a higher-threshold region with more stable patterns.

The correlation profile in the top panel of Figure 2 is consistent with the pattern described by Almeda et al. (2017). Correlations to the right of the selected threshold vary within a narrow range and remain close to zero, suggesting that increases beyond 0.43 produce little change in the relationship between help avoidance and TerraNova score. The selected threshold therefore corresponds to a transition between a region of higher variability and one of relative stability.

The lower portion of the threshold range exhibits two additional patterns. First, the strongest negative correlations occur at the smallest threshold values. This pattern is consistent with the interpretation that very low thresholds isolate students with particularly low estimated knowledge,

Journal Pre-proof

where help avoidance is more likely to reflect unmet need for support and is associated with lower performance. Second, correlations become positive over a portion of the range immediately below the selected threshold. This pattern should be interpreted cautiously, but it suggests that behaviors captured in this region may differ from those implied by the original operationalization. One possible interpretation is that some students delay help while remaining engaged with the problem. Because these results are correlational, they do not support causal claims about the benefits of avoiding help, but they indicate that the meaning of the measure varies across the $P(\mathrm{Know})$ range.

In this dataset, 0.43 provides a separation between a region in which the measure is sensitive to threshold choice and one in which it behaves more consistently. This threshold was used in all subsequent analyses.

Results of Analyses Examining Help Avoidance Behavior

Applying the threshold from the replication analysis, we computed the student-level proportion of help avoidance across hint-enabled problems and examined its association with standardized TerraNova performance.

Table 2 presents the regression models described in Section Replication of Threshold-Based Help Avoidance. In the first model, the help-avoidance proportion was not significantly associated with TerraNova score, $b = -0.024$, $p = .685$, with an effect close to zero. Adding average hint usage and average problem correctness in Model 2 did not change this result. The coefficient for help avoidance changed sign but remained small and non-significant, $b = 0.043$, $p = .457$. In the fully specified model 3, the coefficient remained non-significant, $b = 0.040$, $p = .503$.

Across models, average problem correctness was the strongest and most consistent predictor of TerraNova scores, and average hint usage was positively associated with performance in Models 2

Journal Pre-proof

and 3. The help-avoidance measure did not account for additional variation in the outcome beyond these variables. The full model explained approximately 12% of the variance in TerraNova scores, with an adjusted R^2 of approximately .11, indicating that most variation remains unexplained by the behavioral summaries included here. These results do not provide evidence that the threshold-based help-avoidance measure is associated with student-level TerraNova performance.

This null result can be interpreted in two ways. One possibility is that the behavior captured by this measure is not strongly related to distal performance in this dataset. Another possibility, which aligns with the motivating argument of the study, is that the measure combines different behavioral patterns whose associations with performance differ and may offset one another when aggregated. The following analysis examines this possibility by evaluating behavioral heterogeneity within help-avoidance cases.

To examine whether students classified as help avoidant formed a behaviorally consistent group, we restricted the dataset (N = 12,857) to help-avoidance cases (n = 128) under the derived threshold and summarized each student's behavior within that subset. Figure 1 shows the regression tree fitted to these student-level summaries. The tree partitions students primarily based on mean $P(\mathrm{Know})$, with additional splits on average time and number of attempts. Even under a binary label, students differ along these dimensions, indicating variation in how they engage with problems prior to or instead of requesting support.

The regression tree identified nine terminal groups representing distinct combinations of mean knowledge, time, and number of attempts. Most groups are associated with negative predicted TerraNova values relative to the standardized mean, indicating that many patterns captured by the help-avoidance label are associated with lower performance. This pattern is most evident on the left side of the tree, where students with higher mean $P(\mathrm{Know})$ within their help-avoidance cases

Journal Pre-proof

(n = 24) consistently show below average predicted outcomes, a trend that is most severe for those who also spend longer time on problem (n = 7).

The right side of the tree shows a smaller set of groups with positive predicted outcomes. These groups are concentrated among students with mean $P(\mathrm{Know})$ between approximately 0.29 and 0.38 and fewer than 3.3 attempts per problem (n = 56). Within these groups, longer time on problem is associated with higher predicted performance (n = 42). This pattern suggests that some behaviors classified as help avoidance differ from those typically associated with ineffective persistence. In particular, a subset of students combine moderate estimated knowledge, fewer attempts, and sustained time on task in ways associated with stronger outcomes (n = 73).

To examine whether these subgroup differences persist after accounting for continuous behavioral measures, we estimated a regression model including group indicators along with mean attempts and mean time, shown in Table 3. The model was statistically significant, $F(10, 117) = 5.24$, $p < .001$, and explained approximately 31% of the variance in TerraNova scores, with an adjusted R^2 of .25. Mean attempts was not significantly associated with the outcome, while mean time showed a positive trend.

These results indicate that the threshold-based operationalization does not identify a single behavioral pattern. Most groups classified as help avoidant are associated with lower performance, but a subset shows more positive relationships with the outcome. These groups may include students whose behavior is more consistent with sustained engagement prior to seeking support. These students appear to spend more time on problems while making relatively few attempts, within a range of estimated knowledge that meets the threshold criterion but does not reflect the lowest levels of proficiency. The findings indicate that the operationalization aggregates multiple behavioral patterns with different relationships to performance.

Journal Pre-proof

Journal Pre-proof

Discussion

This study examined how help-avoidance behavior is operationalized in ASSISTments, a widely used digital learning platform (DLP) for middle school mathematics, and how these definitions relate to performance. Building on the threshold-based framework of Almeda et al. (2017), we introduced an empirical procedure to select probability-of-knowledge threshold. Across two analyses, the results indicate that current operationalizations capture a set of behaviors that exhibit inconsistent associations with distal outcomes.

The replication analysis identified $P(\text{Know}) = 0.43$ as the point that separates regions of greater sensitivity from those in which the correlation profile is more stable. This value is lower than the 0.60 threshold used in prior work (Aleven et al., 2006; Almeda et al., 2017). While Almeda et al. (2017) relied on visual inspection of correlation patterns, the variance-based criterion applied here produces a comparable but more explicit selection rule. Differences in the resulting threshold likely reflect differences in the datasets, including population, domain, and model parameters. Prior studies examined university students in a genetics context, whereas the present data come from middle school mathematics classrooms. These differences support treating the threshold as an empirical quantity that may vary across contexts. In addressing **RQ1**, these findings show that the measurement of help avoidance is sensitive to the choice of probability-of-knowledge threshold and that different values can produce substantively different classifications.

Accordingly, $P(\text{Know}) = 0.43$ should be interpreted as a dataset-specific estimate rather than a general recommendation. The variance-based approach formalizes the identification of stable regions in the correlation profile, but it also depends on analytic choices, such as the form of the balance term. These choices influence the selected value, which should therefore be viewed as one

Journal Pre-proof

estimate within a broader range of plausible thresholds.

Help Avoidance and Distal Performance

The absence of a significant association between the student-level help-avoidance proportion and standardized TerraNova scores across all three regression models is an important null result. In addressing **RQ2**, the student-level proportions indicate that help-avoidance behavior occurs with measurable frequency across learners, although its prevalence varies by student and does not, on its own, distinguish differences in distal performance. This relationship remained non-significant after optimizing the threshold and controlling for correctness, hint usage, time, and number of problems completed. Average problem correctness was the strongest and most consistent TerraNova score predictor, followed by average hint usage, which was positively associated with distal performance. This pattern is consistent with prior work showing that effective use of available supports is associated with improved outcomes (Aleven et al., 2006; Baker et al., 2011). The contrast between the positive association for hint usage and the null result for help avoidance suggests that engaging with support may be more relevant to performance than persistence without it.

There are at least two interpretations of this result. One possibility is that help-avoidance behavior may not have a strong relationship with distal performance in this dataset. TerraNova scores reflect cumulative performance over an academic year and are influenced by factors beyond platform interactions. This is consistent with prior work showing that fine-grained behavioral measures often account for a limited portion of variance in standardized assessments (Baker et al., 2011). A second interpretation, which aligns with the theoretical motivation of this study, is that the null result reflects the aggregation of different behavioral patterns within a single measure. Students who persist without seeking help may differ in whether that persistence reflects sustained

Journal Pre-proof

Journal Pre-proof

engagement or a failure to recognize the need for support. When these patterns are combined into a single proportion, their associations with performance may offset one another. In addressing **RQ4**, these results indicate that help-avoidance behavior, when aggregated, does not show a consistent relationship with distal performance. The heterogeneity analysis was designed to examine this possibility directly.

Behavioral Heterogeneity Among Help Avoiders

The regression tree analysis demonstrates that help-avoidant students do not form a behaviorally uniform group. The tree partitions students primarily by mean $P(Know)$ within help-avoidance cases, with secondary splits on time on problem and attempt count. These distinctions confirm that students matching the same problem-level rule differ in their problem-solving behavior. In addressing **RQ3**, these results indicate that help avoidance encompasses multiple behavioral patterns rather than a single, consistent form of engagement.

Most terminal nodes were associated with below-average predicted TerraNova scores, consistent with prior work characterizing help avoidance as maladaptive (Almeda et al., 2017; Baker et al., 2011). This pattern is most evident among students with higher mean $P(\text{Know})$ within their help-avoidance cases, particularly when combined with shorter time on problem. This profile suggests a pattern of limited engagement, where students make rapid attempts without sufficient time or knowledge to make progress. Such behavior is consistent with prior descriptions of disengaged or superficial interaction (Baker et al., 2010) rather than deliberate avoidance of support.

A smaller set of groups shows a different pattern. Students with mean $P(\text{Know})$ between approximately 0.29 and 0.38, fewer than 3.3 attempts, and longer time on problem were associated with more positive predicted outcomes. The follow-up regression indicates that several of these

Journal Pre-proof

subgroups differ from the reference group even after accounting for time and attempts. One interpretation is that these students engage in sustained effort prior to seeking help (Kapur, 2008). In this case, behavior classified as help avoidance may reflect continued engagement with the problem rather than a failure to seek support. Because the analysis is observational and the regression tree is used for subgroup discovery, this interpretation should be treated with caution. However, the pattern is consistent with the possibility that the current operationalization groups together behaviors that differ in how students respond to difficulty.

This finding presents a broader challenge in measuring student behavior using interaction data. Prior work shows that students may appear to be engaged from observable actions while making limited progress (Baker et al., 2011). Our results suggest that behaviors associated with persistence and disengagement may not be easily distinguished within the current operationalization.

Toward a More Granular Operationalization

These findings indicate a need to reconsider how help avoidance is defined and measured in DLP research. Even with threshold refinement, the current operationalization remains limited in capturing differences in how students engage with problems and decide whether to seek support.

A more granular operationalization may incorporate additional dimensions of student behavior. First, time on problem prior to a help request, or in the absence of a request, provides information about the level of engagement before deciding not to seek help. Second, the number of attempts made before a hint request reflects how many opportunities a student has had to adjust their approach before bypassing available support. Third, consistency in behavior across help-avoidance cases may distinguish students who delay help in some instances from those who exhibit more consistent patterns of avoidance. In addressing **RQ3** and **RQ4**, these dimensions help explain the

Journal Pre-proof

varying relationships with performance.

Each of these dimensions was reflected in the regression tree analysis, where differences in time, attempts, and estimated knowledge were associated with different outcome patterns. These findings suggest that incorporating multiple behavioral features may provide a more accurate representation of help-avoidance behavior than a single threshold-based measure.

Limitations

Several limitations should be considered. First, the distal nature of the TerraNova assessment introduces variation that the behavioral summaries are unlikely to fully capture. End-of-year standardized tests reflect a broad range of competencies influenced by factors beyond students' interactions within ASSISTments, including instructional context, motivation, and test-taking strategies. The weak association between the help-avoidance proportion and TerraNova scores may therefore reflect the difficulty of linking fine-grained behavioral measures to aggregate outcomes.

Second, the analytic sample was limited to 348 students with linked TerraNova records. This restriction reduces statistical power and may introduce selection bias if these students differ systematically from those without linked outcomes. Future work should examine whether the observed patterns replicate in samples with more complete outcome data.

Third, the $P(\text{Know})$ estimates used in this study were derived from a larger auxiliary dataset and applied to the present sample. The extent to which these estimates reflect student knowledge in this context is uncertain, and any inaccuracies will affect the resulting help-avoidance classifications. This limitation is common to approaches that rely on Bayesian Knowledge Tracing (Corbett and Anderson, 1994) and applies to prior work in this area.

Finally, the structure of ASSISTments shapes how students interact with problems. Features

Journal Pre-proof

Journal Pre-proof

such as on-demand hints, scaffolded sub-problems, and immediate feedback influence when and how students seek support. As a result, patterns of help avoidance observed in this setting may differ from those in other DLPs with different support structures. Generalization beyond this context should therefore be made with caution. Similarly, information regarding how teachers implement ASSISTments within their instructional routines is unavailable. Classroom use may vary, such as the relative proportion of classwork to homework, introducing unobserved behavioral differences across the dataset. Contextual factor details could provide crucial insights into how instructional settings mitigate help avoidance and encourage timely help-seeking.

Implications for Research and Practice

One of the most notable findings of this study is that help avoidance is not a homogeneous trait. Our analysis reveals that within the same aggregate classification, delaying assistance may manifest as productive struggle for some learners and unproductive persistence for others. In traditional physical classrooms, help-seeking is heavily influenced by social dynamics; students may avoid asking questions to protect their social standing, avoid negative peer evaluation, or manage how their errors are perceived (Ryan et al., 2001; White and Bembenutty, 2013). Yet, our findings hint that in DLPs, where these social threats are minimized, avoiding immediate support may reflect a strategic effort to understand the problem rather than a failure to engage. Fostering true self-regulated learning requires giving students the space to navigate difficulty independently. Therefore, teachers should interpret persistence cautiously, recognizing that a student delaying help-seeking might be engaging in a productive, self-regulated choice.

Similarly, DLPs that rely on binary detection of help avoidance based on a knowledge threshold risk failing to distinguish between students who need immediate support and those engaged in pro-

Journal Pre-proof

Journal Pre-proof

ductive struggle. Incorporating additional behavioral indicators, such as time on task and attempt sequences, can better identify when intervention is needed without disrupting cognitive processing. Building on this, our study shows that how help avoidance is operationalized, specifically the threshold used to define when help should be sought, directly affects the behavioral categories identified and their relationship with outcomes. Using a data-driven threshold selection procedure, we identified a threshold that differs from prior work and yields a help-avoidance measure not associated with distal performance at the aggregate level. However, heterogeneity analysis reveals that this null result conceals underlying variations in metacognitive processes among students classified as help avoiders.

For researchers, these findings highlight the critical need to move away from treating help avoidance as a single, homogeneous behavioral construct. As DLPs become increasingly central to educational practice, the validity of the behavioral measures derived from their interaction logs becomes increasingly important. More behaviorally nuanced operationalizations of help avoidance have the potential to improve learning analytics, inform adaptive instructional systems, and advance theories of self-regulated learning by more accurately identifying when students are engaging productively and when they require additional support. To achieve this, researchers can leverage ASSISTments and similar digital learning platforms with shared research infrastructures to reproduce our operationalization across different samples. Doing so establishes a foundation for testing theoretically meaningful questions about persistence and the underlying drivers of help avoidance. Rather than asking whether students avoid help, future research should focus on understanding how, when, and why students choose to persist independently before seeking assistance. That shift in perspective may ultimately lead to more effective educational interventions and a richer understanding of how learners regulate their own learning.

Journal Pre-proof

Journal Pre-proof

Conclusions

This study revisited the operationalization of help avoidance in a DLP by examining how prior knowledge thresholds and fine-grained behavioral data influence the identification of help-avoidant behavior. Although we identified an empirical threshold that differed from previous work, refining the threshold alone did not produce a measure that consistently related to distal mathematics achievement. Instead, the findings indicate that help avoidance, as currently operationalized, encompasses multiple behavioral patterns that differ in their relationships with student performance. Notably, they suggest that help avoidance should be understood as a heterogeneous construct rather than a single maladaptive learning behavior. Observable actions, such as delaying or foregoing available support, may reflect different forms of self-regulation, ranging from productive persistence to unproductive struggle. Distinguishing among these behaviors is essential for developing measures that more accurately represent how students engage with challenging learning tasks.

Acknowledging the complexity of help avoidance is an important step toward building more responsive educational practices. By continuing to refine how we measure and interpret pauses in help-seeking, DLPs can better support the independence and strategic problem-solving they aim to cultivate. This ensures that adaptive environments respect productive persistence, intervening primarily when a learner's struggle is no longer beneficial. Ultimately, this nuanced approach has the potential to foster deeper understanding and stronger self-regulatory skills.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process During the preparation of this manuscript, the authors used generative AI tools (specifically Gemini and ChatGPT) to improve the clarity and conciseness of the writing. The authors reviewed and edited all content and take full responsibility for the final version of the manuscript.

Journal Pre-proof

Journal Pre-proof

References

Aleven, V., Koedinger, K.R., 2000. Limitations of student control: Do students know when they need help?, in: International conference on intelligent tutoring systems, Springer. pp. 292–303.

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, 101–128.

Aleven, V., Stahl, E., Schworm, S., Fischer, F., Wallace, R., 2003. Help seeking and help design in interactive learning environments. Review of educational research 73, 277–320.

Almeda, V., Baker, R., Corbett, A., 2017. Help avoidance: When students should seek help, and the consequences of failing to do so. Teachers College Record 119, 1–24.

Baker, R.S.d., Corbett, A.T., Gowda, S.M., Wagner, A.Z., MacLaren, B.A., Kauffman, L.R., Mitchell, A.P., Giguere, S., 2010. Contextual slip and prediction of student performance after use of an intelligent tutor, in: International conference on user modeling, adaptation, and personalization, Springer. pp. 52–63.

Baker, R.S.J.d., Gowda, S.M., Corbett, A.T., 2011. Automatically detecting a student's preparation for future learning: Help use is key, in: Proceedings of the 4th International Conference on Educational Data Mining, pp. 179–188.

Beck, J.E., Gong, Y., 2013. Wheel-spinning: Students who fail to master a skill, in: International conference on artificial intelligence in education, Springer. pp. 431–440.

Journal Pre-proof

Journal Pre-proof

Botelho, A.F., 2019. Characterizing Productive Perseverance Using Sensor-Free Detectors of Student Knowledge, Behavior, and Affect. PhD dissertation. Worcester Polytechnic Institute. Worcester, MA. URL: https://digital.wpi.edu/show/8p58pg32w.

Boud, D., Molloy, E., 2013. Rethinking models of feedback for learning: the challenge of design. Assessment & Evaluation in higher education 38, 698–712.

Butler, R., 1998. Determinants of help seeking: Relations between perceived reasons for classroom help-avoidance and help-seeking behaviors in an experimental context. Journal of educational psychology 90, 630.

Clark, C.M., Bjork, R.A., 2014. When and why introducing difficulties and errors can enhance instruction. Applying science of learning in education: Infusing psychological science into the curriculum , 20–30.

Corbett, A.T., Anderson, J.R., 1994. Knowledge tracing: Modeling the acquisition of procedural knowledge. User modeling and user-adapted interaction 4, 253–278.

Daumiller, M., Meyer, J., 2025. Advancing feedback research in educational psychology: Insights into feedback processes and determinants of effectiveness. Contemporary Educational Psychology , 102390.

Feng, M., Heffernan, N., Collins, K., Heffernan, C., Murphy, R.F., 2023. Implementing and evaluating assistments online math homework support at large scale over two years: Findings and lessons learned, in: International Conference on Artificial Intelligence in Education, Springer. pp. 28–40.

Journal Pre-proof

Journal Pre-proof

Feng, M., Heffernan, N.T., 2006. Informing teachers live about student learning: Reporting in the assistment system. Technology Instruction Cognition and Learning 3, 63.

Fischer, F., Bruhn, J., Gräsel, C., Mandl, H., 2002. Fostering collaborative knowledge construction with visualization tools. Learning and instruction 12, 213–232.

Fong, C.J., 2025. A renaissance in feedback science? reviewing and reimagining feedback research methods. Contemporary Educational Psychology , 102414.

Fyfe, E.R., Grenell, A., Vest, N.A., 2026. The negative effects of self-focused feedback on mathematics problem solving. Applied Cognitive Psychology 40, e70239.

Fyfe, E.R., de Leeuw, J.R., Carvalho, P.F., Goldstone, R.L., Sherman, J., Admiraal, D., Alford, L.K., Bonner, A., Brassil, C.E., Brooks, C.A., et al., 2021. Manyclasses 1: Assessing the generalizable effect of immediate feedback versus delayed feedback across many college classes. Advances in Methods and Practices in Psychological Science 4, 25152459211027575.

Hattie, J., Timperley, H., 2007. The power of feedback. Review of educational research 77, 81–112.

Heffernan, N.T., Heffernan, C.L., 2014. 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, 470–497.

Kapur, M., 2008. Productive failure. Cognition and instruction 26, 379–424.

Karabenick, S.A., Berger, J.L., 2013. Help seeking as a self-regulated learning strategy, in: Bembenutty, H., Cleary, T.J., Kitsantas, A. (Eds.), Applications of Self-Regulated Learning Across

Journal Pre-proof

Diverse Disciplines: A Tribute to Barry J. Zimmerman. Information Age Publishing, pp. 237–268.

Karabenick, S.A., Knapp, J.R., 1991. Relationship of academic help seeking to the use of learning strategies and other instrumental achievement behavior in college students. Journal of educational psychology 83, 221.

Kluger, A.N., DeNisi, A., 1996. The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. Psychological bulletin 119, 254.

Koedinger, K.R., McLaughlin, E.A., Heffernan, N.T., 2010. A quasi-experimental evaluation of an on-line formative assessment and tutoring system. Journal of Educational Computing Research 43, 489–510.

Lee, H.Y., List, A., 2023. Examining students help-seeking when learning from multiple texts. Contemporary Educational Psychology 75, 102232.

Makara, K.A., Karabenick, S.A., 2013. Characterizing sources of academic help in the age of expanding educational technology: A new conceptual framework, in: Karabenick, S.A., Puustinen, M. (Eds.), Advances in Help-Seeking Research and Applications: The Role of Emerging Technologies. Emerald Group Publishing Limited, pp. 37–72. URL: https://www.emerald.com/books/edited-volume/20290/chapter-abstract/105664458/Characterizing-Sources-of-Academic-Help-in-the-Age.

Mullet, H.G., Butler, A.C., Verdin, B., von Borries, R., Marsh, E.J., 2014. Delaying feedback

Journal Pre-proof

promotes transfer of knowledge despite student preferences to receive feedback immediately. Journal of Applied Research in Memory and Cognition 3, 222–229.

Murphy, R., Roschelle, J., Feng, M., Mason, C.A., 2020. Investigating efficacy, moderators and mediators for an online mathematics homework intervention. Journal of Research on Educational Effectiveness 13, 235–270.

Narciss, S., Huth, K., 2004. How to design informative tutoring feedback for multimedia learning. Instructional design for multimedia learning 181195.

Nelson-Le Gall, S., 1981. Help-seeking: An understudied problem-solving skill in children. Developmental review 1, 224–246.

Newman, R.S., 2002. How self-regulated learners cope with academic difficulty: The role of adaptive help seeking. Theory into practice 41, 132–138.

Ostrow, K., Heffernan, N., Williams, J.J., 2017. Tomorrow's edtech today: Establishing a learning platform as a collaborative research tool for sound science. Teachers College Record 119, 1–36.

Price, T.W., Zhi, R., Barnes, T., 2017. Hint generation under uncertainty: The effect of hint quality on help-seeking behavior, in: International conference on artificial intelligence in education, Springer. pp. 311–322.

Puustinen, M., 1998. Help-seeking behavior in a problem-solving situation: Development of self-regulation. European Journal of Psychology of education 13, 271–282.

Roschelle, J., Feng, M., Murphy, R.F., Mason, C.A., 2016. Online mathematics homework increases student achievement. AERA open 2, 2332858416673968.

Journal Pre-proof

Journal Pre-proof

Ryan, A.M., Patrick, H., Shim, S.O., 2005. Differential profiles of students identified by their teacher as having avoidant, appropriate, or dependent help-seeking tendencies in the classroom. Journal of Educational psychology 97, 275.

Ryan, A.M., Pintrich, P.R., 1997. " should i ask for help?" the role of motivation and attitudes in adolescents' help seeking in math class. Journal of educational psychology 89, 329.

Ryan, A.M., Pintrich, P.R., Midgley, C., 2001. Avoiding seeking help in the classroom: Who and why? Educational Psychology Review 13, 93–114.

Schenke, K., Lam, A., Conley, A., Karabenick, S., 2015. Adolescents help seeking in mathematics and perceived classroom environmental influences over one school year. Contemporary Educational Psychology 41, 133–146.

Schunk, D.H., Zimmerman, B.J., 1998. Self-regulated learning: From teaching to self-reflective practice. Guilford Press.

Smith, H., Ottmar, E., Ngo, V., Closser, A.H., Yun-Chen Chan, J., Vanacore, K., Sales, A., 2025. Math problem supports in an authentic classroom context: Unpacking the effect of immediate feedback and re-attempts in online problem sets on middle school students math achievement. Journal of Research on Educational Effectiveness , 1–26.

Tricomi, E., DePasque, S., 2016. The role of feedback in learning and motivation, in: Kim, S.i., Reeve, J., Bong, M. (Eds.), Recent Developments in Neuroscience Research on Human Motivation. Emerald Group Publishing Limited. volume 19 of *Advances in Motivation and Achievement*, pp. 175–202. doi:10.1108/S0749-742320160000019015.

Journal Pre-proof

Journal Pre-proof

Turner, J.C., Midgley, C., Meyer, D.K., Gheen, M., Anderman, E.M., Kang, Y., Patrick, H., 2002. The classroom environment and students' reports of avoidance strategies in mathematics: A multimethod study. Journal of educational psychology 94, 88.

Wang, Y., Heffernan, N.T., 2011. The" assistance" model: Leveraging how many hints and attempts a student needs., in: FLAIRS.

White, M.C., Bembenutty, H., 2013. Not all avoidance help seekers are created equal: Individual differences in adaptive and executive help seeking. Sage Open 3, 2158244013484916.

Yang, Y., Liu, R.D., Ding, Y., Wang, J., Chen, Y., Zhou, H., Yang, X., 2025. The influence of competitive and cooperative classroom climate on middle school students' academic help seeking. Learning and Individual Differences 119, 102647. doi:10.1016/j.lindif.2025.102647.

Zimmerman, B.J., 2000. Attaining self-regulation: A social cognitive perspective, in: Handbook of self-regulation. Elsevier, pp. 13–39.

Zimmerman, B.J., 2002. Becoming a self-regulated learner: An overview. Theory into practice 41, 64–70.

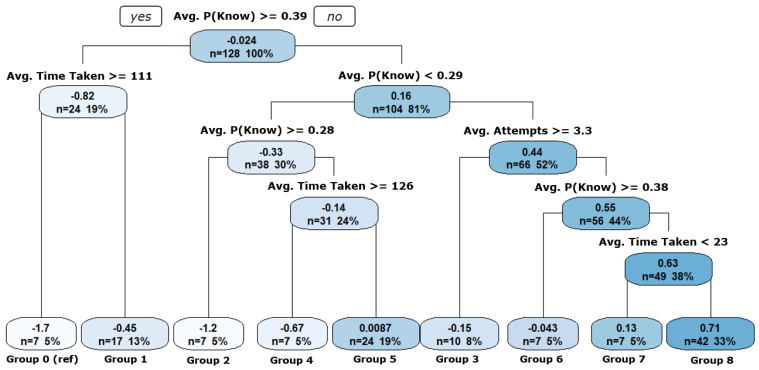
Zusho, A., Barnett, P.A., 2011. Personal and contextual determinants of ethnically diverse female high school students patterns of academic help seeking and help avoidance in english and mathematics. Contemporary Educational Psychology 36, 152–164.

Journal Pre-proof

Journal Pre-proof

Figure 1

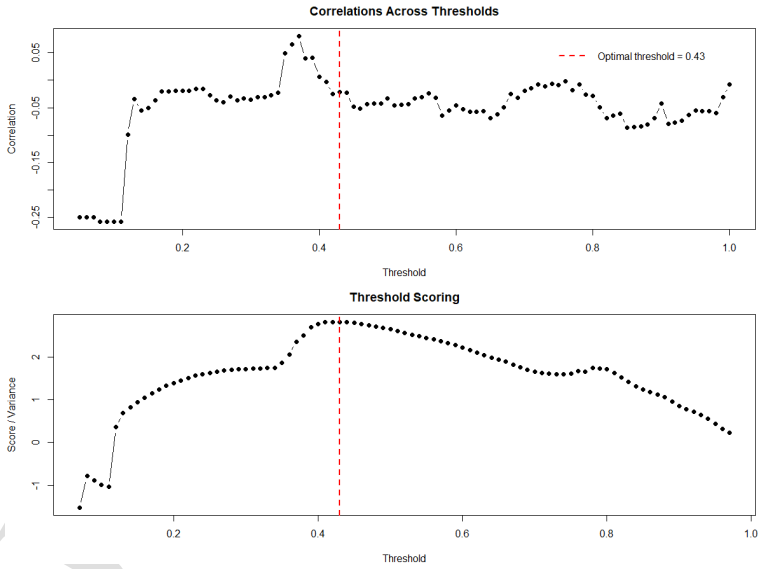
Regression Tree for Students Classified as Help Avoidant



Note. Terminal nodes represent help-avoidant subgroups by student-level behavioral summaries partitioned by mean $P(\text{Know})$, time, and attempts. Node values indicate predicted standardized TerraNova score.

Figure 2

Threshold Replication Analysis



Note. Top: student-level help-avoidance correlation with TerraNova scores across candidate P(Know) thresholds. Bottom: variance-balance score used to select the optimal threshold. The dashed vertical line marks the selected threshold, $P(\text{Know}) = 0.43$.

Journal Pre-proof

Table 1
Replication of Threshold-Based Help Avoidance: Concepts Definitions and Operationalizations

Variable/ Measure	Operational Definition	Source Data / Log Metric	Thresholds/ Decision Rules
Help Avoidance (HA)	A student lacks sufficient knowledge to solve a problem and makes an attempt without asking for help.	ASSISTments logs: accuracy, hint requests, attempts prior to hint, BKT $P(Know)$, TerraNova scores, and action timestamps	Classified as help avoiders based on Almeda et al.'s threshold. Evaluated under baseline benchmark ($\tau = 0.60$) and empirical stability threshold ($\tau = 0.43$).
Probability-of-Knowledge $P(Know)$	The likelihood that a student knows the underlying skill prior to attempting the problem, derived from Bayesian Knowledge Tracing (BKT).	ASSISTments BKT model log outputs	Continuous (0.0 to 1.0). Thresholds tested: $\tau = 0.60$ (Almeda et al. benchmark) and then $\tau = 0.43$ (empirical cut-off).
Hint-Enabled Problem	Problems that offer on-demand hints to support students.	ASSISTments problem configuration logs (problem metadata)	Binary (1 = hints available, 0 = no hints available). Used as a filter: non-hint problems excluded from HA analysis.
Hint Request	Whether a student requested a hint on a problem.	ASSISTments event logs (hint button click)	Binary (1 = Hint requested, 0 = No hint). Aggregated at student level as average hint usage.
Attempts Before First Hint Request	Number of attempts made before seeking a hint.	Timestamped response logs prior to hint request	Ratio variable (0, 1, 2, ...).
Total Attempts	Total number of attempts on a problem when no hint was requested.	ASSISTments response logs	Ratio variable (0, 1, 2, ...). Aggregated at student level as average attempts per problem.
Elapsed Time Before First Hint Request	Time spent before a student requested a hint.	Timestamp during problem (minutes)	Time since last action was 20 minutes or less.
Total Time on Problem	Total time taken to complete the whole problem.	Timestamp during problem (minutes)	Includes problem sessions completed within the same day. Aggregated at student level as average log time per problem.
Correctness	Whether the student answered the question correctly on their first attempt.	ASSISTments item correct/incorrect	Binary (0 = no, 1 = yes). Aggregated at student level as average problem correctness.
Total Number of Problems	Total count of hint-enabled problems completed by the student.	ASSISTments problem logs	Count variable. Used as a control variable (N problems) in regression models.
Distal Math Achievement (TerraNova Scores)	Standardized mathematical achievement score.	TerraNova standardized scores	Standardized scale score.

Note. τ denotes the candidate prior knowledge cutoff threshold ($P(Know)$). Students with $P(Know) < \tau$ on a given item are classified as lacking foundational knowledge for that skill.

Journal Pre-proof

Journal Pre-proof

Table 2
Linear Regression Models Predicting TerraNova Score from the Student-Level Proportion of Help Avoidance and Additional Behavioral Covariates

Predictor	Model 1	Model 2	Model 3
Intercept	$-0.014(0.058)$	$-0.018(0.055)$	$-0.019(0.055)$
Help Avoidance	$-0.024(0.058)$	$0.043(0.058)$	$0.040(0.059)$
Avg hint usage		$0.230^{***}(0.055)$	$0.182^{**}(0.060)$
Avg correctness		$0.297^{***}(0.058)$	$0.301^{***}(0.058)$
Avg log time			$-0.034(0.060)$
Avg attempts			$-0.088(0.059)$
N problems			$0.104^{\dagger}(0.060)$
R^2	0.000	0.105	0.122
Adjusted R^2	-0.002	0.097	0.106
N	348	348	348

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Note. Outcome variable is standardized TerraNova score. All predictors were standardized prior to estimation; coefficients are therefore unstandardized in scale but comparable across predictors. Values are reported as β (SE).

Table 3
Regression of Standardized TerraNova Scores on Behavioral Subgroup Membership Among Help-Avoidant Students

Predictor	β	SE
Intercept (reference group)	-1.88^{**}	(0.57)
Mean attempts	-0.11	(0.10)
Mean time	$0.002^{\dagger}$	(0.001)
Group 1	1.59^{**}	(0.54)
Group 2	0.81	(0.62)
Group 3	1.94^{**}	(0.59)
Group 4	$1.06^{\dagger}$	(0.59)
Group 5	2.00^{***}	(0.52)
Group 6	1.79^{**}	(0.62)
Group 7	2.18^{**}	(0.65)
Group 8	2.59^{***}	(0.49)
R^2	0.31	
Adjusted R^2	0.25	
N	128	

$^{\dagger}p < .10$. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Note. Outcome variable is standardized TerraNova score. All predictors are computed at the student level within help-avoidance cases. Group indicators represent terminal nodes from the regression tree.

Journal Pre-proof

Journal Pre-proof

Appendix A

Replication of Threshold-Based Help Avoidance For a candidate threshold τ, we defined problem-level help avoidance as

$$HA_{ij}(\tau) = \mathbb{I}\left(P(\text{Know})_{ij} < \tau \ \wedge \ \text{attempt before hint}_{ij}\right),$$

where i indexes students and j indexes problems. This preserves the structure of the Aleven definition while allowing the threshold to vary. The threshold was evaluated over the interval from 0.05 to 1.00 in increments of 0.01. This grid spans nearly the full support of the probability scale while providing sufficient resolution to detect changes in the behavior of the measure across nearby threshold values.

For each threshold, the problem-level labels were aggregated to the student level as

$$\text{PropHA}_i(\tau) = \frac{1}{n_i} \sum_{j=1}^{n_i} HA_{ij}(\tau),$$

where n_i is the number of retained hint-enabled problems for student i. This aggregation follows Almeda et al. and is necessary because the outcome examined in the replication, TerraNova score, is defined at the student level. The resulting quantity represents the proportion of a student's observed problems classified as help avoidance by the threshold-based rule.

For each threshold τ, we then computed the Pearson correlation between $\text{PropHA}_i(\tau)$ and TerraNova score across students with non-missing TerraNova data. This produced a sequence of correlations indexed by threshold. Reproducing this correlation profile was important for two reasons. First, it allowed us to examine whether our dataset exhibited the same broad pattern described by Almeda et al. Second, it provided the empirical object on which threshold selection would be based in the present study.

Journal Pre-proof

Journal Pre-proof

Almeda et al. argue that threshold choice should be informed by correlation profile stability, not merely the magnitude of a single correlation. More explicitly: if small threshold shifts produce substantial correlation changes, the measure is sensitive and difficult to interpret as its operational meaning shifts appreciably. Conversely, if the correlation remains consistent across nearby thresholds, the measure behaves more stably. A threshold near the transition from instability to stability is therefore of particular interest.

To formalize this logic, we treated the threshold-selection problem as identifying a cut point in the correlation profile that separates a variable region from a stable one. Each candidate threshold partitioned the correlations into lower and higher threshold segments, from which we computed the variance within each segment. Variance directly captures how the help-avoidance relationship fluctuates across thresholds. A candidate threshold aligns with Almeda et al.'s qualitative logic when the correlations below it are relatively variable and the correlations above it are relatively stable.

This comparison requires more than maximizing the ratio of lower-to-upper threshold variance. A threshold near the search ranges edge could produce an extreme ratio from very few points, failing to represent a meaningful transition. To address this, we introduced a balance term penalizing thresholds that split the profile too unevenly. This term peaks when the threshold divides candidates into similarly sized segments, ensuring the criterion favors cut points supported by non-trivial information on both sides.

For each candidate threshold, we therefore computed a score that combined three ideas: high variability below the threshold, low variability above the threshold, and adequate representation of thresholds on both sides of the cut. Let $V_L(\tau)$ denote the variance of correlations for thresholds less than or equal to τ, let $V_R(\tau)$ denote the variance of correlations for thresholds greater than τ,

Journal Pre-proof

Journal Pre-proof

and let $B(\tau)$ denote the balance term derived from the proportion of evaluated thresholds on the lower side of the cut. The score was defined as

$$S(\tau) = \log\left(\frac{V_L(\tau) + \epsilon}{V_R(\tau) + \epsilon}\right) \cdot B(\tau)^{\alpha},$$

where ϵ is a small constant included to avoid numerical instability and α controls the strength of the balance penalty. In the present study, α was set to 1. The logarithmic ratio ensures that the score increases when variance below the threshold is larger than variance above it, while the balance term discounts thresholds that produce trivial partitions.

The selected threshold was the value of τ that maximized $S(\tau)$. This procedure retains the empirical logic underlying Almeda et al.'s analysis, but expresses it in a form that can be reproduced, inspected, and applied consistently. The resulting threshold was then used to instantiate the help-avoidance measure in the subsequent analyses.

Journal Pre-proof

Journal Pre-proof

Not All Who Avoid Help Are Lost: The Relations Between Help Avoidance and Performance in a Digital Learning Platform

Natalia S. Martin*, Nicholas A. Vest, Avery H. Closser, Anthony F. Botelho

University of Florida, Gainesville, FL 32601

Author Note:

*Corresponding Author; Email: nataliamartin@ufl.edu

All authors contributed to this manuscript.

Natalia S. Martin: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – Original Draft.

Nicholas A. Vest: Software, Formal analysis, Writing – Review & Editing.

Avery H. Closser: Writing – Original Draft, Writing – Review & Editing.

Anthony F. Botelho: Conceptualization, Methodology, Software, Formal analysis, Supervision, Writing – Review & Editing.

Compliance with Ethical Standards

All procedures were conducted in compliance with relevant laws and institutional guidelines and were approved by the University of Florida Institutional Review Board on 10/24/2025 (Protocol # NH00048495).

Declaration of Generative AI and AI-assisted Technologies

A declaration is included in the manuscript at the end (above References) and reads as follows:

During the preparation of this manuscript, the authors used generative AI tools (specifically Gemini and ChatGPT) to improve the clarity and conciseness of the writing. The authors reviewed and edited all content and take full responsibility for the final version of the manuscript.

Journal Pre-proof

Journal Pre-proof

Highlights

- Threshold-based metrics show weak associations with distal outcomes.
- Results suggest behavioral diversity within help-avoidance categories.
- Low latency and high mastery may indicate passive disengagement.
- Longer latencies and fewer attempts signal productive struggle.
- Strategic persistence patterns predict higher standardized scores.

Journal Pre-proof

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof