Promoting Equitable Learning Outcomes for Underserved Students in Open-Ended Learning Environments

Joyce Horn Fonteles
joyce.h.fonteles@Vanderbilt.edu
Vanderbilt University
Nashville, Tennessee, USA

Celestine Akpanoko
celestine.e.akpanoko@Vanderbilt.edu
Vanderbilt University
Nashville, Tennessee, USA

Pamela Wisniewski
pamela.wisniewski@Vanderbilt.edu
Vanderbilt University
Nashville, Tennessee, USA

Gautam Biswas
gautam.biswas@Vanderbilt.edu
Vanderbilt University
Nashville, Tennessee, USA

ABSTRACT

Computer-Based Open-Ended Learning Environments (OELEs) are designed to challenge learners to become proficient problem-solvers and develop the ability to independently solve complex problems. However, the traditional focus of OELE research has been on demonstrating overall learning gains, potentially overlooking students who struggle in these environments. To address this gap, we take a social justice-based approach by studying 99 sixth-grade students who participated in a week-long classroom study. We first assessed learning outcomes across all then identified 20 students who failed to do well. We qualitatively analyzed video recordings of their interactions with the OELE to understand why they struggled and to determine if interface issues inhibited their learning. Five themes emerged: (1) challenges in knowledge acquisition; (2) challenges in scaffolding learning; (3) disregarding system guidance, (4) not leveraging supporting tools; (5) and getting discouraged by incorrect answers. Based on our findings, we make design recommendations for OELEs to better support underserved learners, recognizing that failure is an important catalyst for motivating improvements in child-centered design.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Applied computing** → **Interactive learning environments**.

KEYWORDS

Undeserved students, social justice design, self-regulated learning, open-ended learning environments

ACM Reference Format:
Joyce Horn Fonteles, Celestine Akpanoko, Pamela Wisniewski, and Gautam Biswas. 2024. Promoting Equitable Learning Outcomes for Underserved Students in Open-Ended Learning Environments. In *Interaction Design and Children (IDC '24), June 17–20, 2024, Delft, Netherlands.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3628516.3655753

IDC '24, June 17–20, 2024, Delft, Netherlands

ACM ISBN 979-8-4007-0442-0/24/06
https://doi.org/10.1145/3628516.3655753

1 INTRODUCTION

Self-regulated learning (SRL) is fundamental for children as it empowers them with the skills and strategies that they need to take control of their learning processes and achieve success [14, 51]. Children with strong self-regulation skills can set specific learning goals, create plans, enact their plans, and monitor progress toward achieving these goals. As a result, they foster effective study habits and time management skills, which leads to overall success in their academic careers [64]. Furthermore, SRL skills extend beyond academic success to impact other important facets of life, such as developing effective problem-solving and decision-making skills, and adapting to different environments [63]. In addition, SRL skills help children develop a sense of autonomy and a proactive approach to learning. Children develop metacognitive processes, such as planning, monitoring, and reflecting, learn to identify obstacles and develop strategies to overcome these obstacles by monitoring and reflecting on their evolving solutions [59]. Moreover, instilling SRL into academic curricula equips children with the tools to navigate the complexities of modern education and prepares them for the future 21^{st} century workforce, where continuous learning is integral to personal and professional growth [18].

Computer-based open-ended Learning Environments (OELEs) are designed to engage learners in solving complex problems independently, thus providing powerful opportunities to help them develop and utilize SRL strategies in their problem-solving tasks [31]. Their open-ended nature provides opportunities for connecting learning to real-world problem-solving scenarios, making the learning processes authentic, and therefore, more motivating. By offering choices and opportunities for developing monitoring and decision-making skills, OELEs empower children to make autonomous decisions about their learning [12], enhancing their ability to set goals, plan their approach to achieving these goals, make informed choices in their problem-solving tasks, and monitor progress toward achieving their goals, all of which are key components of SRL. However, not all students benefit equally from OELEs, and a social justice perspective prompts us to delve deeper into the experiences of those who struggle within these environments [7]. Each student is unique, and individual differences in learning styles, preferences, and cognitive abilities can influence how they engage with and benefit from OELEs. Some students thrive in open-ended scenarios, while others find them more challenging due to factors such as unfamiliarity with content, difficulty making sense of it, or a need to improve their self-regulation skills [22].

Our goals in this paper are to shed light on the intricacies of OELE interactions, with a specific focus on underserved students who encounter challenges within these environments [12, 41]. By delving into the behaviors, strategies, and interface design elements that hinder the learning experiences of these students, we aim to contribute valuable insights and offer tangible interface suggestions to the research community [1, 46]. Through this exploration, we hope to advance child-centered design principles, ensuring that OELEs are not only effective for some but are genuinely beneficial for all learners, ultimately fostering equitable learning outcomes for underserved students. In line with our goals, we address OELE design issues by formulating three research questions:

- RQ1: How can we identify classes of underserved students based on their differentiating learning outcomes after working on an OELE?
- RQ2: What are the barriers and unproductive strategies employed by underserved students while working on an OELE?
- RQ3: How do interface design choices in OELEs impact the learning experiences of underserved students?

To investigate these questions, we analyzed data from a comprehensive study involving 99 sixth-grade students engaged in a week-long OELE-based learning experience, specifically focusing on learning science concepts by building a causal model of the human causes for the greenhouse effect and the impact of the greenhouse effect on climate. In addressing potential oversights in OELE design that lead to student difficulties, our study presents three key contributions. First, we introduce a method for identifying and classifying underserved students through the analysis of quantitative data, including assessments of previous knowledge, learning outcomes, and performance during OELE engagement. Second, our qualitative analysis, grounded in video recordings of student interactions, reveals five prominent themes hindering the learning experiences of underserved students: challenges in knowledge acquisition, challenges in scaffolding learning, disregarding system guidance, not leveraging supporting tools, and getting discouraged by the lack of progress in their learning tasks. Last, having explored how interface design influenced students' strategies in navigating these challenges, we provide design recommendations for OELEs, acknowledging the social justice imperative of addressing the needs of underserved learners.

2 RELATED WORK

This section synthesizes the prior literature on self-regulated learning, open-ended learning environments, and the use of pedagogical agents in these environments to support scaffolds, and survivorship bias in learning and design research.

2.1 Self-Regulated Learning

Self-regulated learning (SRL) is characterized by an engaged learner actively overseeing and managing their cognitive and metacognitive processes throughout the learning journey and aligned with their individual goals [38, 66]. The emotional dimensions of learners, such as confusion and curiosity, referred to as 'learning emotions,' are also significant influencers of their overall learning outcomes [39]. Curiosity-driven learning has shown significant positive effects on children's metacognitive efficiency and their ability to express their curiosity through questions [48].

Within this framework, three key dimensions are highlighted. First, there is a dual focus on both self-regulation processes and the strategies employed that target these processes. Second, the importance of continuing feedback emerges as a critical facilitator in enabling the self-regulated learning process. Lastly, SRL emphasizes the interdependence between motivation and self-regulating processes. This interconnected relationship has been extensively explored, with the social cognitive view of SRL emphasizing self-efficacy as a pivotal measure of self-regulation, acting as a driving force behind motivation. In addition, various authors have affirmed the positive relationship between self-efficacy and motivational elements, such as goal-setting and planning [24, 50].

Zimmerman's cyclical phase model for self-regulated learning (2009) [65] outlines three key phases: Forethought, Performance, and Reflection. In the Forethought phase, learners undertake pre-learning behaviors, including goal-setting. During the Performance phase, learners actively employ self-control processes and engage in self-observation to glean internal feedback. Progressing to the Reflection phase, learners evaluate their progress based on self-observation and engage in self-judgment. Notably, learners express emotional responses to this judgment, influencing the input for the subsequent iteration of the self-regulated learning cycle. Winne and Hadwin's SRL model emphasizes a metacognitive perspective, portraying self-regulated students as actively involved in managing their learning using monitoring and metacognitive strategies[60]. The model highlights the goal-driven nature of SRL and the impact of self-regulatory processes on motivation. Studying involves four phases in a loop: task definition, goal setting and planning, enacting study tactics and strategies, and metacognitive adaptation [45]. Though different SRL models evaluate different areas and aspects of the learning processes, they converge to four primary components that govern self-regulation of learning, as shown in Figure 1.

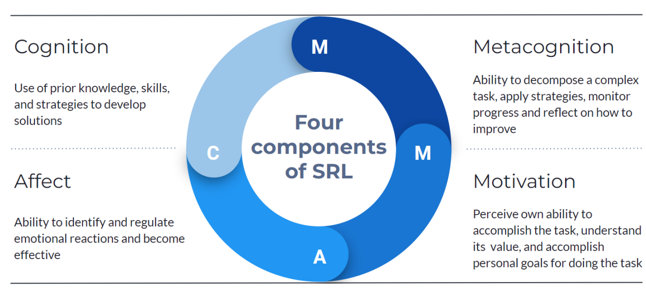


Figure 1: Primary components for self-regulation of learning

2.2 Open-Ended Learning Environments and Pedagogical Agents

Open-Ended Learning Environments (OELEs) have gained significant attention in recent years because of their potential to support student creativity, problem-solving, and self-regulated learning [12].

These environments facilitate a high degree of student autonomy and open-endedness, allowing learners to explore, experiment, design, create, and solve in an unconstrained manner. The current state of the art for OELEs involves a range of technologies, including simulations, virtual and augmented reality, serious games, and pedagogical agents, that aim to support student learning and engagement [27, 53, 56, 57]. According to Land and Jonassen the concept of OELEs is rooted in constructivist theories of learning, which emphasize the importance of learners' active engagement and exploration in the learning process [32]. In this perspective, learners construct their understanding of the world by actively engaging with it, and learning occurs through a process of inquiry, reflection, and iteration. Wang et al.stress that the effectiveness of the OELE systems depends on various factors, such as the quality of the learning content, the pedagogical strategies employed, and the level of support provided [58].

OELEs can be particularly effective at supporting SRL as they provide learners with the freedom and flexibility to explore and experiment, while also offering guidance and feedback to support self-regulation. They support SRL by providing students with targeted learning goals; a set of tools to facilitate the learning and problem-solving processes; and an open-ended approach that offers choice in how students combine these tools to achieve their learning goals[10, 12]. However, as Munshi et al. observed, novice learners might be unfamiliar with these tools and lacking in SRL processes, therefore, they frequently resort to less-than-optimal strategies when approaching learning and problem-solving tasks, thus increasing the difficulties they face in their learning tasks [42]. The authors have addressed this issue by designing and implementing an adaptive scaffolding framework to help students develop and refine their SRL behaviors while working in an OELE.

Scaffolds have been employed in several prominent OELEs to support SRL. Ecolab is a family of environments for learning ecology and adapts to the student's goal to determine the appropriate form of scaffolding to support metacognitive monitoring and task selection [37]. nStudy is a web-based application that offers a toolkit for students to define and evaluate their learning strategies and link them to their learning artifacts, such as bookmarks and notes [61]. MetaTutor employs four pedagogical agents to scaffold the development and use of specific SRL processes while students learn about topics in biology [3]. Betty's Brain uses the learning-by-teaching paradigm to help students study and construct causal models of scientific processes using a visual representation. Students do this in the guise of teaching a computer agent, Betty, and their interactions with Betty and a Mentor agent, Mr. Davis, help them develop social, cognitive, and metacognitive skills [12, 34].

In traditional adaptive and personalized computer-based learning environments, an individual's preferences, behaviors, and overall learning progress are captured by a user model. The user model provides the basis for the system to carry out tasks related to both adaptation and personalization [24]. Kim et al. studied how pedagogical agents can play a key role in supporting SRL in OELEs by providing learners with feedback, guidance, and prompts for reflection [66]. Fu et al. implemented conversational agents to support children's socioemotional learning through self-talk, which affects cognitive performance, the ability to self-regulate, and problem-solving skills [23].

A pedagogical agent may ask learners to reflect on their goals and progress and offer suggestions on alternative approaches to a problem. The integration of pedagogical agents and other intelligent technologies into OELEs has the potential to support learners' development of metacognitive and self-regulated learning skills, which in turn also helps with academic achievement. Pedagogical agents can play different roles in open-ended learning environments, depending on the specific learning goals and the design of the environment. For example, a pedagogical agent may play the role of a mentor or coach, and provide guidance and support to learners as they engage in self-directed learning activities [8, 11]. Pedagogical agents can also take on the role of a peer or collaborator, who interacts with learners more socially and conversationally [19, 33]. Blair and Schwartz state that by observing and imitating the pedagogical agent's behavior, learners can acquire new skills and knowledge, which can be useful for tasks that require procedural knowledge or domain-specific expertise [13].

2.3 Survivorship Bias

Survivorship bias is the tendency to concentrate on the positive outcomes of a selection process and overlook the results that generate negative outcomes[28]. This has been studied in various disciplines, including finance[17], information retrieval [25], and healthcare[20]. These studies highlight the importance of recognizing and mitigating survivorship bias to ensure a more accurate and comprehensive understanding of outcomes.

In educational research and design, survivorship bias is a crucial phenomenon that demands careful consideration. It arises when disproportionate attention is given to successful outcomes or participants who have completed a specific intervention, resulting in an incomplete understanding of the challenges faced by those who could not. Furthermore, it can significantly impact User Experience design, introducing challenges that may compromise the quality of user interactions. Survivorship bias often shifts the focus to success stories, which can overshadow failures, thus creating a skewed perception of user requirements. This limitation hampers the ability of designers to comprehensively address user needs, ultimately contributing to a suboptimal experience. Recognizing challenges that can disrupt or render studies and experiments ineffective is an important aspect of research, particularly in the context of Human-Computer Interaction. Rukmane, et al. explored failures within the Child-Computer Interaction community and found they are not reported enough in the literature due to the pressure of publishing successful stories[49].

Nevertheless, there has been an increasing amount of research to support underserved children and help equalize opportunities for all. Ruan, et al. have investigated the relationship between emotions experienced during learning and metacognition in typically developing (TD) children and those with autism spectrum disorder (ASD) [47]. They adopt a social justice approach for improving machine learning algorithms that examine the relationship between facial emotion expressions and metacognitive monitoring performance for both TD children and those with ASD. Antle et al. designed a neuro-feedback system and applications that enabled traumatized children living in poverty to learn and practice self-regulation by playing games [2]. The researchers addressed design choices to overcome challenges, such as working with illiterate children, those who did not speak English, and lacked computer experience. McLaren and Antle presented a mixed methods framework for evaluating whether sound can help children with attentional challenges to self-regulate using a neurofeedback system [40].

Technology is not neutral and often not suitable for all children involved in a study. Despite the extensive research on SRL, OELEs, and the role of pedagogical agents, there remains a notable gap in understanding how these environments cater to the needs of underserved students [31, 67]. Existing research often focuses on overall learning gains and successful outcomes, potentially overlooking the challenges faced by students who struggle within these environments [21, 53]. Moreover, the phenomenon of *survivorship bias* in educational research and design further complicates efforts to ensure equity and inclusivity, as it tends to prioritize successful outcomes while neglecting the experiences of students who encounter difficulties, and, therefore, fail to make progress in their assigned learning tasks [26]. Posing questions about the suitability of SRL for all kinds of students, and further investigating how their actions should trigger appropriate scaffolding mechanisms in these environments, especially when students are having difficulties, is of paramount importance.

Our research focuses on expanding knowledge concerning design methods and researcher reflexivity. Much like design itself, which is often generative and forward-looking, social justice endeavors share a similar orientation towards the future. By adopting a social justice-based approach and acknowledging the importance of mitigating survivorship bias, we hope to bridge this gap by examining how OELEs and pedagogical agents can be designed and implemented to support underserved students, investigate how their choices and actions in the OELE might be better understood and addressed, and ultimately contribute to the development of more inclusive and effective learning environments that promote equitable learning outcomes for all students.

3 METHODS

3.1 Study Overview

This paper analyzes data collected from a week-long sixth-grade urban classroom study in the southeastern United States. Students used the Betty's Brain environment to learn about a complex science phenomenon [43]. The analysis focused on students' self-regulation behaviors and the effectiveness of adaptive scaffolds delivered by pedagogical agents to help students develop effective learning and problem-solving strategies [41].

3.2 The Betty's Brain Open-Ended Learning Environment

Betty's Brain is an OELE that utilizes the learning-by-teaching paradigm to engage students in learning about science topics [10, 34]. The system includes hypertext resources that describe the topic under study, and students are expected to read the resources and construct a causal model to teach Betty, the Teachable Agent. By reading and translating the content in the hypertext resources into a correct causal map of science phenomena, the students demonstrate their emerging understanding of the corresponding science topic. The student's overall goal is to learn the scientific topic well enough to teach Betty a complete model of a scientific process. The complete map, generated by the student's teachers or the research team, is used to evaluate the students' model. The topic of study for this project was the human causes of climate change, and the corresponding "expert" causal map is shown in Figure 2.

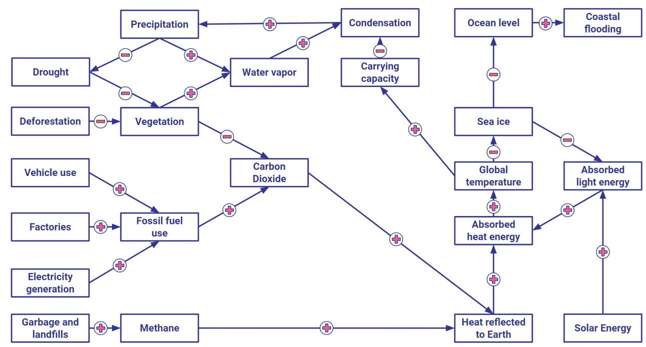


Figure 2: Representation of the expert causal map

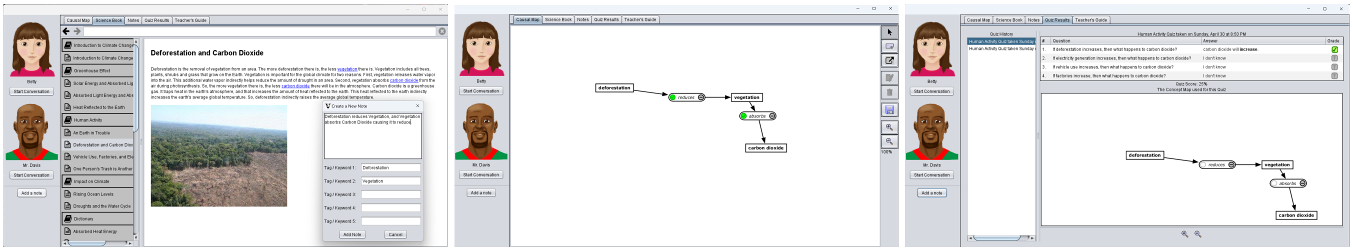
To build the causal map, students can access the science book to acquire knowledge on the three main topics that make up the causal model: Human Activity that causes the Greenhouse Effect, the Greenhouse effect, and the impact of the Greenhouse effect on Climate change (Figure 3a). Students use the drag-and-drop causal map editor to express how concepts are related to each other(Figure 3b). At any time, students can ask Betty to take a quiz and her performance helps them assess the correctness of their maps (Figure 3c). These quizzes are created dynamically by Mr. Davis, the Mentor agent, who also grades them and keeps track of Betty's overall performance. The quizzes are designed to provide feedback on the correctness and completeness of their scientific model, and students can use this information to determine where they have made msitakes. In addition, Mr. Davis monitors the student's actions as they go about their tasks.

3.3 Participants

The study reported in this paper took place in December 2018 with 99 consenting sixth-grade students in an urban public school in the southeastern US. Overall, the school's population is 60% White, 25% Black, 9% Asian, and 5% Hispanic, with 8% enrolled in the free/ reduced-price lunch program, which reflects the demographics of our study population. Unfortunately, individual classroom demographics were not collected for this study.

3.4 Study Design and Data Collection

The study spanned five days. On day 1, students took a paper-based pre-test that used multiple-choice and short-answer questions, to assess their comprehension of the domain and proficiency in causal reasoning skills. The tests were designed in consultation with middle school educators, keeping the sixth-grade public school curriculum in mind. Day 2 was dedicated to students engaging with a practice unit to acquaint themselves with the Betty's Brain environment. On Days 3–5, students actively built causal models of climate change in the Betty's Brain OELE. On the final day, the students also completed a post-test identical to the pre-test.



(a) Science Book **(b) Causal Map** **(c) Quiz Results**

Figure 3: (a) The student access the science book to read and deduce concepts and links while taking notes, (b) the student builds causal links from concepts learned in the science book, and (c) the student gets Betty to take a quiz and assess the results.

Data collected from consenting students included screen recordings, webcam recordings, eye-tracking data, system logs, and the paper-based pre-and post-tests. These tests have previously been used in multiple Betty's Brain classroom studies[43, 53].

3.5 Data Analysis

3.5.1 RQ1: Identification of Underserved Students. In exploring the first research question (RQ1), we used two different data sources to understand how students benefited from the intervention. We started with pre-tests to assess what students already knew about climate change. Then, by comparing pre- and post-test results, we measured how much each student learned overall through our intervention. Additionally, system logs documented the sequence of actions performed by the students,i.e., reading, note-taking, map-building, and quiz-taking activities with corresponding timestamps. The final map scores, a composite measure derived from the sum of correct links minus incorrect links on the students' maps, were instrumental in gauging not only what students learned but also their potential misconceptions.

In previous studies that evaluated Betty's Brain data, students were grouped into High Learning Gain (High) and Low Learning Gain (Low) categories based on the differences in their normalized learning gains in domain knowledge and causal reasoning skills [42]. However, categorizing students solely based on pre-to-post test differences has its limitations. The scenario where a student with a high pre-test score maintains or slightly raises their score in the post-test may result in a categorization of Low Learning Gain, despite the limited room for improvement (ceiling effect). Conversely, a student with a low pre-test score might have more opportunities for growth in the post-test, but this doesn't guarantee the student will have positive learning gains.

The intervention is expected to yield significant overall learning gains when aggregating students' learning gains. Past studies have shown high aggregated learning gains across all participants in a study [12, 34, 43]. However, to determine if these are significant for all of the members in a study, we further delve into the data to identify classes of students based on their differing learning outcomes. We looked for the significance of these gains for all members of the group individually. To do this, we created a quadrant-based categorization by performing a median split on both pre-test and post-test scores for each student, and categorized them before and after the intervention using the following labels: low-low, low-high, high-low, and high-high. As a clarification, a student with a low-low label had a pre-test score that was below the median score for the class and a post-test score that was also below the median score for the class. This approach provides us with a more nuanced assessment of the role of previous knowledge in students' knowledge acquisition and their relative ranking in the class before and after the intervention. Additionally, this quadrant-based approach also allows us to investigate situations where learning gains and performance within the system, measured by map scores, may diverge, providing a comprehensive understanding of the intervention's impact on different student groups.

3.5.2 RQ2 and RQ3: Unproductive Strategies and the Role of the Interface. After categorizing students into quadrants based on pre-test and post-test scores in RQ1, our attention turned to two specific groups for RQ2 and RQ3: low-low and high-low. These groups represent underserved students who encountered challenges in deriving comparable benefits from the learning environment as their peers. To delve into a detailed analysis of students' behaviors, we used the following data sources: screen recordings depicting student's work on the OELE with an added overlay of their gaze paths (captured using a calibrated Tobi 4C eye tracker) as well as webcam footage showing their interactions with others, reactions, and facial expressions. This multimodal approach enabled a more nuanced analysis not only of their engagement with the OELE and the task, but also their attentional focus, emotional responses, and social dynamics. This provided us with deeper insights into their learning processes, motivations, and emotional responses. We selected 20 students (12 low-low and 8 high-low) who had complete data spanning all four days of working on the climate change unit, totalling 80 hours of videos. This approach enabled us to observe the persistence of specific challenges the students had and their inability to overcome them throughout the intervention.

As a next step, we employed a *thematic analysis* approach [16], which allowed us to analyze the data and uncover underlying patterns and themes systematically. The analysis was conducted collaboratively by the first and second co-authors, following the phases of thematic analysis. Initially, we familiarized ourselves with the data and then proceeded to generate initial codes to encapsulate key ideas. Through an iterative process, these codes were conceptually grouped to form overarching themes. All through this process, the last two authors provided valuable input, offering high-level feedback on the coding process and codebook. After refining the codebook, we identified a final subset of emergent themes that aligned closely with our research questions. We provide details of the codebook in the Appendix. This in-depth exploration not only shed light on the barriers and unproductive strategies employed by underserved students (RQ2) but also delved into the influence of interface design choices on their learning experiences (RQ3).

4 RESULTS

4.1 Data Scoping by Identifying Learning Quadrants

Existing approaches usually categorize students into two groups by overall performance – High and Low—based solely on differences in their normalized learning gains from pre-to-post tests. However, this classification overlooks the nuances in students' interactions with the system and their performance, particularly in scenarios where students with high pre-test scores might have limited room for improvement, which could potentially lead to misclassifications. Conversely, students with low pre-test scores might have more opportunities for growth in the post-test, but this doesn't guarantee substantial improvement in understanding of the learning materials. Therefore, The research team decided to conduct a more in-depth analysis to gain a deeper understanding of the impact of the intervention on different student subgroups. To achieve this, we introduced a quadrant-based categorization scheme that considered median splits on both pre-test and post-test scores independently. This approach, outlined in Table 1, allowed the researchers to more accurately identify and target underserved student populations. Identifying students' prior knowledge of the subject materials allows us to better analyze if and when the system's design choices and content might have been skewed toward students who start with a good grasp of the subject material. Furthermore, by assessing where students fit in the median split of the post-test, we can focus our efforts on the students who did not benefit from the intervention with the system.

Quadrant	Description
High-High (H-H)	39 students classified as high on both pre and post tests
Low-High (L-H)	15 students classified as low on pre, high on post tests
Low-Low (L-L)	31 students classified as low on both pre and post tests
High-Low (H-L)	14 students classified as high on pre, low on post tests

Table 1: Quadrant-based classification of students

The quadrant analysis, defined by the four quadrants outlined in Table 1, showed that a majority of students who scored high on the pre-test also did so on the post-test, confirming how prior knowledge is almost always a pre-requisite for learning and applying SRL skills for complex tasks [54, 55]. In addition, students who transitioned from a low score on the pre-test to a high score on the post-test demonstrated significant improvement, indicating that the OELE provided them with productive learning experiences by correctly attending to their specific needs and skills, which resulted in high gains. However, our goal was to delve deeper into the experiences of the students who did not benefit much from the intervention. Therefore, we studied two groups in greater detail: (1) 31 students who started low on the pre-test and remained low on the post-test, i.e., the Low-Low (L-L) group; and (2) 14 students who started with high scores on the pre-test but ended with low scores on the post-test, i.e., the High-Low (H-L) group.

Furthermore, we compared our quadrant-based classification to two other metrics to better circumscribe the underserved students we selected for a deeper analysis: (1) the final scores that students achieved on their causal maps in the Betty's Brain environment; and (2) their pre-to-post normalized learning gains. Out of the 31 students classified as low-low in our quadrant, 28 of them also had low map scores at the end of the Betty's Brain intervention and low learning gains, reinforcing their status as underserved students in need of assistance. For the 14 students who were in the high-low quadrant, a significant proportion had high map scores in the Betty's Brain environment, but they still exhibited low learning gains. This finding led us to hypothesize that some students with sufficient previous knowledge may have gotten confused when working with the Betty's Brain environment.

4.2 Exploring Challenges to Learning and Interface Impact

To answer RQ2 we focused on the students from the Low-Low (L-L) and High-Low (H-L) quadrants that we discussed in RQ1. This categorization helped us identify underserved students who may not have benefited from the learning environment in the same way as their peers. Out of the 20 students selected for in-depth qualitative analysis, 12 belonged to the Low-Low group, while 8 belonged to the High-Low group. We eliminated students for whom we did not have data for all four days of the intervention.

The qualitative analysis involved video data acquired from three sources: (1) screen recordings showing students' interactions with the learning environment; (2) webcam recordings capturing facial expressions and reactions; and (3) eye-tracking overlays providing insights into attention and focus on specific visual elements. Thematic analysis was employed to identify patterns and themes in students' behaviors. Similarly, we focused on the challenges students faced in their interactions with the Betty's Brain environment. We created the code book using the Betty's Brain task model developed in previous work [30, 53]. The task model breaks down the primary OELE tasks into three cognitive processes that support metacognition and self-regulation behaviors, i.e. Information Seeking/Acquisition, Solution Construction/Refinement, and Solution Assessment. The cognitive processes are directly linked to observable actions students perform on the system's interface, such as reading, taking notes, editing the causal map, asking for the agents' assistance, taking quizzes and evaluating answers.

Thematic analysis of the data resulted in identifying five major themes that represent the progression of behaviors and strategies of the underserved students, from their getting started with the system until they disengage from it:

(1) Challenges in knowledge acquisition,
(2) Challenges in adopting scaffolds for learning,
(3) Disregarding system guidance,
(4) Not leveraging the support tools provided, and
(5) Getting discouraged because they could not figure out how to correct incorrect answers in the quizzes.

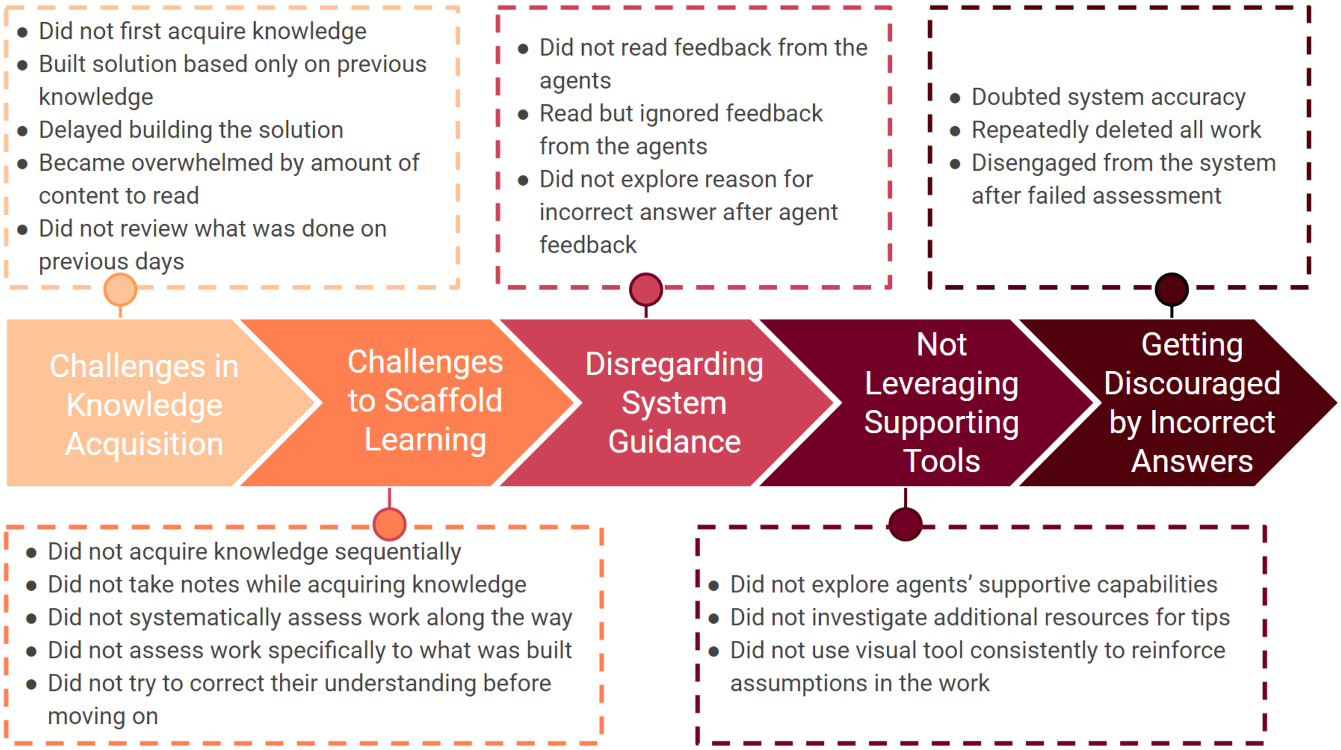


Figure 4: Thematic analysis on the progression of challenges faced by underserved students

Figure 4 depicts the thematic analysis results on the progression of challenges faced by underserved students. This provides us a visual roadmap illustrating the interconnected nature of the identified themes and codes. It reveals a sequential ordering relation among the problems, where one problem leads to another, reflecting the cascading effects of student interactions with the learning environment and the need for rethinking the structure of our interventions. For instance, consider a student who initially spends a lot of time reading the science book without beginning to translate their understanding to causal links and start building the causal map. This to delayed knowledge application can lead to further problems such as the inability to remember material that was read earlier. As Theme 1 - Challenges in Knowledge Acquisition progresses into Theme 2 - Challenges in Adopting Scaffolds for Learning, the student's disorganized approach hampers effective knowledge acquisition and map construction. Despite the system's attempts to provide guidance through the Mentor Agent and Betty in Theme 3 - Disregarding System Guidance, the student may ignore feedback due to intense focus, inability to interpret the guidance, and even unclear instructions in the guidance. This pattern continues with Theme 4 - Not Leveraging Supporting Tools, as the student neglects to explore available resources provided in the system. Finally, in Theme 5 - Getting Discouraged by Incorrect Answers, the student's frustration peaks when quiz results do not improve in spite of their efforts to add and correct links. The lack of clear returns for their efforts, leads to disengagement [5, 36]. This sequence was observed multiple times in the first two days of the intervention, influencing students' dispositions to continue putting effort until the end of the activity. This figure elucidates the sequential progression of challenges faced by underserved students, highlighting critical points where intervention and system improvements are needed.

Building on these findings, we delved into a deeper exploration of the impact of the interface design on students' behaviors and choices, to answer RQ3. Specifically, we focused on understanding how the interface played a role in hindering students' performance, failing to support them, and negatively influencing their learning experiences. For each code and theme, we also looked for distinctions between the High-Low and Low-Low groups, to further our understanding of how the difference in their previous knowledge and learning skills also influenced their actions and how they perceived the system.

Theme 1 - Challenges in Knowledge Acquisition

The students' knowledge acquisition challenges are primarily influenced by the information seeking task in the OELE. Information seeking covers reading the science book to learn about science concepts and relations and note-taking serves as a memory aid to capture information that is relevant to building the causal map. Fifteen students, 8 from the low-low group and 7 from the high-low group encountered challenges related to Theme 1. Across most codes, there was no significant disparity between the two groups. However, it is noteworthy that the code about feeling overwhelmed by the amount of content was exclusive to the L-L group. Five students (3 L-L and 2 H-L) faced challenges in initially acquiring knowledge from the Science Book. This implies that the system was not successful in helping the students engage in the reading task so they could find new information and apply it in an effective manner to build the causal map. Six students (3 L-L and 3 H-L) relied solely on prior knowledge rather than acquiring new information from the science book, opening up the possibilities of prior misconceptions being used to construct the causal map. This challenge highlights how the system failed to help students adapt and become engaged in new learning experiences. Eight students (4 L-L and 4 H-L) had delays in starting their map building task. This is because they focused solely on knowledge acquisition or continued to explore of the system (perhaps, because they did not understand their learning task). Two students from the L-L group showed signs, either through facial expressions or speech, of being overwhelmed by the amount of content presented in the Science Book. In other words, they were overwhelmed because of cognitive load. Furthermore, since the intervention lasted four consecutive days, we observed how not reviewing work from previous days hindered the metacognitive process of reflection for eight students (5 L-L and 3 H-L).

- **Interface Impact - Navigational Complexity and Information Overload**
  The organization of the science book was a major cause for many of the difficulties students faced when acquiring knowledge. Its lack of hierachy of content and dense combination of both text and images proved to be overwhelming for students, causing them to bypass the first process of acquiring knowledge and instead depend on their current knowledge or prematurely create causal maps (refer to Figure 5 for the structure of the interface). Furthermore, certain students felt obligated to thoroughly study all pages before trying to create their maps, which caused a delay in knowledge application (i.e. building their solution). In addition, the organization of the science book includes hyperlinked texts that redirects students to a different page on the dictionary. This causes confusion and prevents them from following a clear and logical reading path where they can use the structure of the science book to build a chain of causal links connecting related concepts. The interface's design unintentionally distracts students from their main reading objective, complicating the process of acquiring knowledge.
  The combined influence of these interface design elements affected the students' capacity to acquire and employ information in an effective way for map building, highlighting the crucial function of intuitive interface design in supporting students' productive knowledge acquisition. In addition, at the beginning of the study, the researchers suggested that students should build their models sequentially and in parts, however this group of students either ignored or did not understand this suggestion and the system itself failed to reinforce it as a strategy.

Theme 2 - Challenges in Adopting Scaffolds for Learning

All 20 students faced challenges in systematically approaching their causal model construction and debugging tasks. This happened in relation to all of their primary cognitive processes: information seeking and acquisition, solution construction and refinement, and solution assessment. Seventeen students (12 L-L and 5 H-L) underutilized the note-taking feature. Thirteen students (10 L-L and 3 H-L) encountered difficulties in acquiring knowledge sequentially or understanding how concepts identified in one part of the Science Book might relate to others. This points to issues in understanding the hierarchical structure of the science book and the inability to use the hyperlinks to navigate material in the book to find relations between concepts. As a result, the students had difficulties in finding an effective approach to developing and advancing their causal maps. Sixteen students (8 L-L and 8 H-L) faced challenges in systematically evaluating their work by taking quizzes. Unable to interpret the quiz results, i.e., the implications of right and wrong answers, they resorted to continuing to read the science book in the hope that may help them find erroneous links, or they just continued to tweak their maps for a long time (a pure trial and error approach) [62]. Eleven students (7 L-L and 4 H-L) struggled to assess their work about the current state of their causal models, taking quizzes on topics and concepts that they had not added to their causal maps. Six students (4 L-L and 2 H-L) often moved on to add additional links to their maps without attempting to correct errors shown in the quiz results. This resulted in the number of errors continuing to grow in their map, which made it even harder to debug their maps [29].

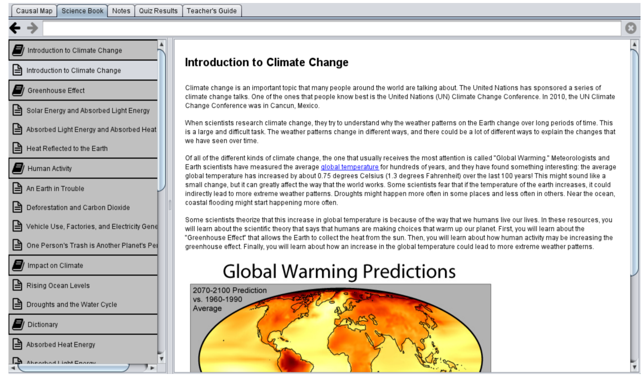


Figure 5: Interface impact for theme 1 - dense combination of text and images in science book

- **Interface Impact - Ineffective Feedback Mechanisms and User Engagement**
  The challenges to scaffold learning are intricately connected to the design of the learning environment. The note-taking tool, although designed to facilitate systematic knowledge acquisition, poses a navigational obstacle. The procedure necessitates that students transition to a distinct 'Notes' page to write and access their notes, causing interruptions in their reviewing of the science book and difficulty in linking the notes back to science book pages, thus reducing the effectiveness of note-taking as a mechanism for collecting linked information as a reference and organizer for subsequent map building. This design issue frequently leads to students not using the note-taking feature, as the notes are not easily accessible in conjunction with the knowledge acquisition and map-building tasks.

The causal map workspace, in conjunction with the agents and quiz results page, is specifically designed to streamline the process of constructing, debugging, and assessing students' work. Nevertheless, the interface fails to sufficiently motivate students to interact with these aspects systematically and effectively. Students frequently prioritize building or correcting their maps in the causal map workspace to an extreme degree, disregarding the crucial procedures of evaluating their map using the quiz results, pinpointing errors and missing links, and then trying to update their models to correct their errors. in other words, students often resort to trial and error approaches to debugging their map instead of using a systematic assess, locate, and correct process using the quiz results [29]. This disparity suggests that the interface should stimulate and reinforce the significance of regular and specific checking of the causal map to improve the overall effectiveness of the causal map building (solution construction) process. The Mentor agent and Betty could be more effective in reminding students and facilitating these important SRL processes.

Theme 3 - Disregarding System Guidance

This theme describes the gap in effective student-agent communication. The agents monitor students' actions to provide adaptive scaffolds and encouragement, offering tips on specific learning or debugging strategies like periodically taking quizzes to check their map, and cautioning against adding concepts to the map without prior reading. Fourteen students were coded for this theme and two codes revealed a significant disparity, with a higher count of L-L students neglecting to read feedback or utilize quiz feedback to correct misconceptions in the map. Having the eye-tracking gaze overlaid on the screen recording provided a mechanism for more detailed analysis. On multiple occasions, seven students (6 L-L and 1 H-L) did not engage with or read the feedback provided by the agents, often ignoring them for long periods or just clicking on the dialogue until the agent went away. Furthermore, eleven students (6 L-L and 5 H-L) read agent feedback that provided information on effective information acquisition and solution construction strategies, but chose not to act on the suggestions provided. This included suggestions for reading specific pages that contained information that was relevant to the part of the causal map the student was constructing and tips on how to go about debugging the students' current causal map.

- **Interface Impact - Inadequate Feedback Delivery**
  The system utilizes pedagogical agents to motivate and guide the students toward their learning goals, especially the ones who face difficulties. However, the efficacy of this assistance is dependent on how well the interface provides the feedback. The feedback, designed to steer students away from unproductive sequences of activities, is shown as conversational prompts at the top of the screen (see Figure 6 for an example of conversational feedback). Nevertheless, this design decision poses two main concerns. First, the positioning and format of the feedback require students to pause their tasks, which may interrupt their learning and thinking processes. Furthermore, the repetitious nature of the feedback, which lacks precise direction tailored to the situation, causes children to either ignore the information or mechanically proceed without actively interacting with the subject on hand. In addition, the system's inability to track and address whether the children take action based on the feedback exacerbates this problem. This is apparent in cases where students fail to inquire into the causes of their incorrect answers, even after getting feedback on their errors. The interface design choices unintentionally lead to learners ignoring system instructions, highlighting the urgent requirement for better user-friendly and prompt feedback systems in educational interfaces.

Theme 4 - Not Leveraging Supporting Tools

This theme encompasses various ways students could have been supported by the system, including initiating conversations with agents and exploring functionalities for understanding systematic map construction and map debugging processes. Thirteen out of 20 students struggled with this theme, and across all codes, the L-L group exhibited a higher count than the H-L group. Six students (4 L-L and 2 H-L) did not initiate dialogues with the agents to take advantage of the supportive information provided by them, eleven students (8 L-L and 3 H-L) did not actively seek explore the resources in the Teacher's Guide, which provides valuable tips, and ten students (7 L-L and 3 H-L) were not consistent or did not use the “Mark as Right" tool, a visual tool to both visually reinforce correct links while also differentiating it from the portions of the map that still require assessment and revision.

- **Interface Impact - Feedback Deficiency and Student Disengagement**
  The theme of Not Leveraging Supporting Tools is closely linked to how these tools are presented and made accessible through the interface, rather than their complete absence. The essential supporting tools, such as agent feedback, the teacher's guide, and the link-marking feature for the causal map, suffer from challenges linked to visibility and user engagement. The teacher's guide closely mirrors the structure of the science book, providing a dense combination of pertinent tips and supplementary information. The intricate nature of this resource poses difficulty for children in efficiently extracting relevant information, potentially resulting in its disregard. This is especially pertinent for younger children, who may not be good readers and may not like to read long passages. Furthermore, the link-marking feature, which is essential for accurately measuring causal map accuracy and simplifying the error finding and correction process, is not as prominent on the UI as other features that have dedicated buttons. The lack of clarity leads to the underutilization of an essential instrument.
  Furthermore, the agents' conversational feedback, although intended to provide help, does not successfully inform or remind students about the accessible capabilities, such as the ability to examine particular causal relationships. The deficiency in the agents' guiding system exacerbates the underutilization of supporting tools. Together, these interface design elements have a considerable influence on students'

level of involvement with and use of the materials that are accessible to them.

Theme 5 - Getting Discouraged by Incorrect Answers

Students constantly got frustrated and discouraged as a result of receiving negative quiz assessments [6, 12]. We noticed how this theme was heavily influenced by challenges students faced previously in other themes. For instance, a student who struggled with acquiring knowledge initially, and then began building the map based solely on previous knowledge (Theme 1) was more likely to doubt system accuracy during the quiz assessments. In all of our codes, the L-L group had only a slightly higher count than the H-L group. Thirteen out of 20 students were coded for this theme, with four (3 L-L and 1 H-L) having openly expressed disbelief about the system being correct, a potential breakdown in trust between the student and the learning environment. Eleven students (7 L-L and 4 H-L) repeatedly deleted their work, either displaying a lack of confidence while building their causal maps or after receiving a 0% quiz assessment. Interestingly, some students that chose to delete their existing work and start anew, chose to do so on a different topic, without attempting to correct their previous errors. Six students (4 L-L and 2 H-L) gave up on the learning task and disengaged from the system for a considerable period of time after failing a quiz.

- **Interface Impact - Interface Design Hurdles in Learning Scaffolding**
  The interface primarily communicates correct and incorrect links to the student indirectly via the quiz results (see Figure 3c). Additional information is provided through the agents' feedback. The quiz results page classifies each answer to a quiz question as either 'correct', 'incorrect', or 'I don't know', depending on the student's constructed causal map. Additionally, it offers a partial or complete view of the student's causal map in connection to every quiz question. Nevertheless, the interface's approach to simply highlighting incorrect answers without providing thorough explanations may result in student dissatisfaction. This dissatisfaction is evident through behaviors such as completely deleting their map because of low quiz scores and the inability to find and fix errors or growing doubt about the system's accuracy and applicability to the realm of knowledge.
  Furthermore, the inability of the interface, and by extension, the agents, to clarify why an answer may be incorrect and to initially provide a step-by-step process to help students find errors in their map may contribute to student disengagement. This can become worse if the situation recurs frequently. The absence of comprehensive feedback and adaptable support may be a primary cause for student disengagement following many failures to understand the reasoning behind graded answers. To summarize, the interface design's constraints in delivering unambiguous, productive feedback and its inability to dynamically interact with students' learning processes greatly contribute to student discouragement.

5 DISCUSSION

Our utilization of the quadrant-based approach to characterize and study children's difficulties in using an OELE distinguished this work from previous research in the field. Our results provide a more nuanced understanding of students' learning trajectories from start to finish. By categorizing students into L-L and H-L groups, we investigated the actions and strategies employed by the underserved students. A number of these children started with low prior knowledge but some did not; they started with high prior knowledge and ended up with low post-test scores. We systematically investigated how the system's design might have contributed to students' inability to progress leading to confusion and frustrations. This approach allowed us to analyze the design with a more critical eye, specifically in terms of individual student's needs.

Each theme identified in our analysis represents an important aspect of design for OELEs in supporting underserved students who usually struggle in such open-ended environments. The progression of themes, from challenges in knowledge acquisition to disregarding system guidance and getting discouraged by incorrect answers, underscores their interconnected nature and impact on student engagement and learning outcomes. Our findings strongly suggest that underserved students require more support during their initial interactions with the system, which should be gradually faded (i.e., reduced) to provide them with opportunities to develop their own SRL skills [44]. A general issue with providing adaptive scaffolding in OELE environments is that the scaffold triggering mechanisms fire more frequently for students who engage more frequently with the system, resulting in underserved students receiving less actionable feedback [35, 41]. Monitoring the actions represented in our code book and addressing these issues could enhance the effectiveness of OELEs in supporting all students' learning journeys.

5.1 Drawbacks of Betty's Brain Scaffolding Feedback Triggers

The scaffolding agent feedback framework of Betty's Brain, as described by Munshi, et al [41], has two essential elements: a learner model for tracking trigger conditions and a conversational tree structure for providing feedback when these conditions are met. This paradigm primarily depends on the learner model to capture the context of the present task, the efficacy of recent actions, and unique problems associated with continuing tasks. The significance of basing scaffolds on the specific tasks and model-building efforts of students offers concrete indicators for strategic feedback [52]. Nevertheless, this feedback method has several constraints, especially for underserved students whose learning experiences may not correspond to the pre-established learner model in the system.

A previous empirical study found that there was a discrepancy in the frequency of activities between those who performed well and those who performed poorly, with the high performers participating in a greater number of actions [43]. Hence, the existing scaffolding agent feedback architecture tends to prioritize those who excel in terms of the frequency of actions, inadvertently neglecting the needs of those who struggle. This was further emphasized in a subsequent study, revealing that feedback was found to be more beneficial to high performers than their peers [42]. Ultimately, the scaffolding agent feedback architecture in Betty's Brain is a complex tool that supports students. However, its effectiveness varies across different student groups, especially underserved students. This discrepancy highlights the need for more investigation and advancement in this field. We strive to overcome these restrictions by offering customized design recommendations that cater to the needs of underserved students, therefore promoting a fair and efficient learning experience for all system users.

5.2 Implications for Self-Regulated Learning frameworks and OELEs

This work underscores the critical role of first supporting cognitive abilities to foster effective SRL within OELEs, particularly for underserved students. Our findings indicate that SRL frameworks may not effectively support these students if they lack well-developed cognitive abilities, which generalizes beyond the OELE context [4, 9]. In Betty's Brain, for instance, such cognitive abilities encompass reading and comprehension skills, understanding what a causal map is and how it supports the understanding of relationships between parts of a model, and the ability to reason with correct and incorrect quiz answers and how they relate back to the causal map [12]. Therefore, OELEs must prioritize supporting students' cognitive development early in their engagement with the system. By providing scaffolding and assistance tailored to these cognitive abilities, OELEs can empower students to navigate the learning environment more effectively. This initial support lays the groundwork for students to gradually develop SRL, specially their metacognitive and self-regulation processes to gradually become effective learners [15]. Ultimately, this approach enables students to benefit from the open-ended nature of OELEs, fostering cognitive and metacognitive skills, and then self-regulation processes that promote independent learning and autonomy as they progress in their educational journey.

5.3 Design Implications for Learning Environments

Importance of Content Hierarchy and Specificity Theme 1 emerged because of the different struggles underserved students went through in the process of acquiring knowledge. When they first access the Science Book to begin their learning tasks, they are presented with a long list of topics, pages, and dictionary items, all on a scrollable panel, as seen in Figure 5. There are over 30 items in this list, but only 10 pages have concepts that relate to map construction, of which students are unaware. Below is a dialogue between two students who were coded as "Overwhelmed by the amount of content to read":

Student A: I don't know what to do... What do I do?
Student B scrolls over all the reading material available in the Science Book
Student B: Oh my God! That looks so long!
Student A: Yeah...

Content pages should not be represented the same way as dictionary pages, teacher's guides, or tips. Pages with different purposes should be visually distinct to avoid confusing the students or overwhelming them with long lists.

Constant Changes in Context Add To Cognitive Load

Support material that might help in the acquisition, application, and assessment of knowledge should be readily available on any screen, independent of which tab the student is working on. Within Betty's Brain, this principle holds particularly true for tips. These tips are concise and practical suggestions located at the end of a series of pages in the Teacher's Guide. Unfortunately, many students fail to take advantage of this valuable source of knowledge. The agent has received feedback indicating that students are reluctant to modify the context and conduct a search on a separate page, despite being advised to read a specific tip.

The note-taking feature is a good example of a feature that is readily accessible at all times, on any screen, as are the agents, however even though it is possible to take notes on every screen, the students cannot read them at all times. There is a specific tab listing all the notes previously taken and if the student wants to use them to assist in the knowledge construction on the map, the system requires constant changes of page context back and forth, increasing the students' cognitive load.

Tiered Agent Feedback

The thematic analysis showed that a lot of the struggles underserved students encountered were in failing to scaffold learning. The system's mechanism of adaptive scaffolds relies on the students engaging with the agents, which was also shown to be a struggle that students, especially those in the Low-Low group, faced. The length of the content delivered, its perceived utility, and how many times the same content has been delivered before, all these elements play a part in students' choices to avoid agent feedback or disregard their advice. Figure 6 shows Student C's screen and gaze, after receiving feedback from the mentor agent after a failed quiz assessment. The following dialogue took place while the student was engaging with this.

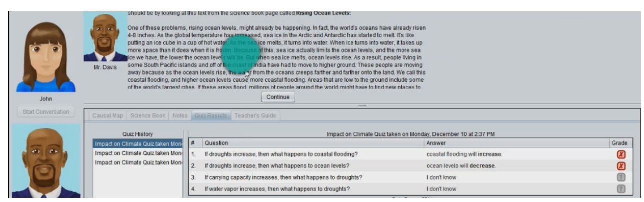


Figure 6: Student received agent feedback after failed quiz assessment

**While seeing a lengthy answer from the agent, the student stares in disbelief and says: **
Student C: Oh my gosh! I have to read all of this?
Student closed the feedback without reading it.
Student C: I really don't care.
Student disengaged from the system.

We suggest that agent feedback should have tiers, feedback should not prevent the student from completing the task they are working on, and feedback must be addressed by the student at that time. Choosing how and when to classify feedback as such should be contingent on:

(1) The type of content that is being delivered by the agent. Timely content that serves to course-correct a student before they disengage is an example of that.
(2) The number of times a student has ignored specific feedback and continued to struggle with the same problem. The system should be able to change the phrasing on the text to convey

its importance to the student before changing its tier, but after multiple failed attempts, make it so the student cannot ignore it and had to take action.

Furthermore, instead of agent conversations just having suggestions on actions to take, could also include actionable tasks that the students must perform before continuing the conversation. The agent would not only suggest the student take an action but walk the student through steps required to complete the action, as a mentor might in real life when a student is struggling to make progress.

5.4 Limitations and Future Work

The in-depth qualitative findings constitute a strength of this research, though the small sample size of evaluated students poses a limitation in terms of generalizability. To address this, future studies could leverage our social justice-based framework with a larger and more diverse sample to enhance the applicability of our findings. Furthermore, though demographic information about students was not collected for this study, future research should prioritize this to contextualize findings and explore potential disparities among student groups. Additionally, the nature of students' participation in the study, which took place during regular class hours and included parental consent, may have influenced their motivation and learning outcomes. Investigating the impact of research context on student engagement is essential; hence, future studies should delve into factors such as parental involvement and extrinsic incentives. Lastly, considering the rising prominence of Large Language Models in educational technology, integrating them into OELEs could offer enhanced adaptive learning experiences and provide valuable insights for the design of educational technologies and instructional practices. Addressing these limitations and exploring future research will contribute to advancing the field of educational technology and promoting equitable learning opportunities.

6 CONCLUSION

In this study, we took a social justice-based approach to investigate the experiences of underserved students in OELEs, aiming to identify classes of underserved students, understand barriers, and explore the impact of interface design choices. Through a week-long classroom study with 99 sixth-grade students, we developed a method for classifying underserved students, utilizing a quadrant approach based on assessments of pre-test and post-tests individually, and validated against learning gains and scored performance within the system. Complementing this, qualitative analysis of video recordings revealed five prominent themes hindering the learning experiences of underserved students, encompassing challenges in knowledge acquisition, scaffolding learning, disregarding system guidance, not leveraging supporting tools, and getting discouraged by incorrect answers. Additionally, our investigation extended to interface design choices, providing insights into their impact on students' strategies. Drawing from these findings, we offered design recommendations aligning with child-centered design principles to foster more inclusive and equitable learning outcomes for all students.

7 SELECTION AND PARTICIPATION OF CHILDREN

The study received approval from an Institutional Review Board at a University in the southeastern United States. It took place during regular class hours at a school where children were recruited by their science teachers, who had been collaborating with the research team on multiple studies since 2010. Teachers received consent letters providing detailed information about the research's purpose, activities, data collection, storage, duration, and confidentiality measures. Similar consent letters were sent to guardians, emphasizing the voluntary nature of participation, benefits of the intervention based on previous studies, and assurances about data privacy. Students received assent letters explaining the research in language suitable for their age and encouraging discussions with their guardians to make a joint decision about participation. All letters highlighted voluntary participation. Researchers worked with teachers to conduct interventions during regular class hours, minimizing disruptions to students' schedules. Consent and assent letters addressed ethical considerations, including data storage, access, and publication anonymization, and emphasized the voluntary nature of participation. Additionally, it was explicitly stated that students could withdraw from the study at any time.

ACKNOWLEDGMENTS

This work is supported in part by the National Science Foundation (NSF) under Grant No. DRL-2112635. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of our sponsors. We thank the teachers and children who participated in this study.

REFERENCES

[1] Vincent Aleven, Elmar Stahl, Silke Schworm, Frank Fischer, and Raven Wallace. 2003. Help seeking and help design in interactive learning environments. *Review of educational research* 73, 3 (2003), 277–320.
[2] Alissa N. Antle, Leslie Chesick, Aaron Levisohn, Srilekha Kirshnamachari Sridharan, and Perry Tan. 2015. Using neurofeedback to teach self-regulation to children living in poverty. In *Proceedings of the 14th International Conference on Interaction Design and Children* (Boston, Massachusetts) *(IDC '15)*. Association for Computing Machinery, New York, NY, USA, 119–128. https://doi.org/10.1145/2771839.2771852
[3] Roger Azevedo, François Bouchet, Melissa Duffy, Jason Harley, Michelle Taub, Gregory Trevors, Elizabeth Cloude, Daryn Dever, Megan Wiedbusch, Franz Wortha, et al. 2022. Lessons learned and future directions of metatutor: Leveraging multichannel data to scaffold self-regulated learning with an intelligent tutoring system. *Frontiers in Psychology* 13 (2022), 813632.
[4] Diana Baas, Jos Castelijns, Marjan Vermeulen, Rob Martens, and Mien Segers. 2015. The relation between Assessment for Learning and elementary students' cognitive and metacognitive strategy use. *British Journal of Educational Psychology* 85, 1 (2015), 33–46.
[5] Ryan SJd Baker, Sidney K D'Mello, Ma Mercedes T Rodrigo, and Arthur C Graesser. 2010. Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies* 68, 4 (2010), 223–241.
[6] Ryan S Baker, Nidhi Nasiar, Jaclyn L Ocumpaugh, Stephen Hutt, Juliana MAL Andres, Stefan Slater, Matthew Schofield, Allison Moore, Luc Paquette, Anabil Munshi, and Gautam Biswas. 2021. Affect-targeted interviews for understanding student frustration. In *International Conference on Artificial Intelligence in Education*. Springer, 52–63.
[7] Satabdi Basu, Gautam Biswas, Pratim Sengupta, Amanda Dickes, John S Kinnebrew, and Douglas Clark. 2016. Identifying middle school students' challenges in computational thinking-based science learning. *Research and practice in technology enhanced learning* 11, 1 (2016), 1–35.

[8] Amy L Baylor and Yanghee Kim. 2005. Simulating instructional roles through pedagogical agents. *International Journal of Artificial Intelligence in Education* 15, 2 (2005), 95–115.
[9] Kirsten Berthold, Matthias Nückles, and Alexander Renkl. 2007. Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning and Instruction* 17, 5 (2007), 564–577.
[10] Gautam Biswas, Krittaya Leelawong, Daniel Schwartz, Nancy Vye, and The Teachable Agents Group at Vande. 2005. LEARNING BY TEACHING: A NEW AGENT PARADIGM FOR EDUCATIONAL SOFTWARE. *Applied Artificial Intelligence* 19, 3-4 (March 2005), 363–392. https://doi.org/10.1080/08839510590910200
[11] Gautam Biswas, Rod Roscoe, Hogyeong Jeong, and Brian Sulcer. 2009. Promoting self-regulated learning skills in agent-based learning environments. In *Proceedings of the 17th international conference on computers in education*. Asia-Pacific Society for Computers in Education Hong Kong, 67–74.
[12] Gautam Biswas, James R. Segedy, and Kritya Bunchongchit. 2016. From Design to Implementation to Practice a Learning by Teaching System: Betty's Brain. *International Journal of Artificial Intelligence in Education* 26, 1 (July 2016), 350–364. https://doi.org/10.1007/s40593-015-0057-9
[13] Kristen Blair, Daniel L. Schwartz, Gautam Biswas, and Krittaya Leelawong. 2007. Pedagogical Agents for Learning by Teaching: Teachable Agents. *Educational technology* 47, 1 (2007), 56–61.
[14] Monique Boekaerts. 1999. Self-regulated learning: Where we are today. *International journal of educational research* 31, 6 (1999), 445–457.
[15] John D Bransford, Ann L Brown, Rodney R Cocking, et al. 2000. *How people learn*. Vol. 11. Washington, DC: National Academy Press.
[16] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa
[17] Stephen J. Brown, William Goetzmann, Roger G. Ibbotson, and Stephen A. Ross. 2015. Survivorship Bias in Performance Studies. *The Review of Financial Studies* 5, 4 (05 2015), 553–580. https://doi.org/10.1093/rfs/5.4.553
[18] Rodger W Bybee and Bruce Fuchs. 2006. Preparing the 21st century workforce: A new reform in science and technology education. (2006).
[19] Chih-Yueh Chou, Tak-Wai Chan, and Chi-Jen Lin. 2003. Redefining the learning companion: the past, present, and future of educational agents. *Computers & Education* 40, 3 (2003), 255–269.
[20] Dirk M. Elston. 2021. Survivorship bias. *Journal of the American Academy of Dermatology* (June 2021). https://doi.org/10.1016/j.jaad.2021.06.845
[21] Mona Emara, Nicole M Hutchins, Shuchi Grover, Caitlin Snyder, and Gautam Biswas. 2021. Examining Student Regulation of Collaborative, Computational, Problem-Solving Processes in Open-Ended Learning Environments. *Journal of Learning Analytics* 8, 1 (2021), 49–74.
[22] Mary C English and Anastasia Kitsantas. 2013. Supporting student self-regulated learning in problem-and project-based learning. *Interdisciplinary journal of problem-based learning* 7, 2 (2013), 6.
[23] Yue Fu, Mingrui Zhang, Lynn K Nguyen, Yifan Lin, Rebecca Michelson, Tala June Tayebi, and Alexis Hiniker. 2023. Self-Talk with Superhero Zip: Supporting Children's Socioemotional Learning with Conversational Agents. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (Chicago, IL, USA) *(IDC '23)*. Association for Computing Machinery, New York, NY, USA, 173–186. https://doi.org/10.1145/3585088.3589376
[24] Julio Guerra. 2016. Open Social Learner Models for Self-Regulated Learning and Learning Motivation. In *Proceedings of the 2016 Conference on User Modeling Adaptation and Personalization* (Halifax, Nova Scotia, Canada) *(UMAP '16)*. Association for Computing Machinery, New York, NY, USA, 329–332. https://doi.org/10.1145/2930238.2930375
[25] Prashansa Gupta and Sean MacAvaney. 2022. On Survivorship Bias in MS MARCO. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '22)*. Association for Computing Machinery, 2214–2219. https://doi.org/10.1145/3477495.3531832
[26] Robin M Hogarth, Tomás Lejarraga, and Emre Soyer. 2015. The two settings of kind and wicked learning environments. *Current Directions in Psychological Science* 24, 5 (2015), 379–385.
[27] Yuanru Jiang, Luc Paquette, Ryan S J d Baker, and Jeffrey Clarke-Midura. 2015. Comparing novice and experienced students within virtual performance assessments. In *Proceedings of the 8th International Conference on Educational Data Mining*. 562–567.
[28] Patricia Katopol. 2017. Maybe Best Practices Aren't: How Survivorship Bias Skews Information Gathering and Decision-Making. *Library Leadership and Management* 32 (11 2017). https://doi.org/10.5860/llm.v32i1.7287
[29] John S Kinnebrew, James R Segedy, and Gautam Biswas. 2014. Analyzing the temporal evolution of students' behaviors in open-ended learning environments. *Metacognition and learning* 9 (2014), 187–215.
[30] John S Kinnebrew, James R Segedy, and Gautam Biswas. 2017. Integrating model-driven and data-driven techniques for analyzing learning behaviors in open-ended learning environments. *IEEE Transactions on Learning Technologies* 10, 2 (2017), 140–153.
[31] Susan M. Land. 2000. Cognitive requirements for learning with open-ended learning environments. *Educational Technology Research and Development* 48, 3 (Sept. 2000), 61–78. https://doi.org/10.1007/bf02319858
[32] Susan M Land and David H Jonassen. 2012. *Theoretical Foundations of Learning Environments*. Routledge.
[33] Yezdi Lashkari, Max Metral, and Pattie Maes. 1994. Collaborative interface agents. In *AAAI*, Vol. 94. 444–449.
[34] Krittaya Leelawong and Gautam Biswas. 2008. Designing Learning by Teaching Agents: The Betty's Brain System. *Int. J. Artif. Intell. Ed.* 18, 3 (aug 2008), 181–208.
[35] Tongguang Li, Jionghao Lin, Sehrish Iqbal, Zachari Swiecki, Yi-Shan Tsai, Yizhou Fan, and Dragan Gašević. 2023. Do learners appreciate adaptivity? An epistemic network analysis of how learners perceive adaptive scaffolding. In *International Conference on Quantitative Ethnography*. Springer, 3–17.
[36] Jason M Lodge, Gregor Kennedy, Lori Lockyer, Amael Arguel, and Mariya Pachman. 2018. Understanding difficulties and resulting confusion in learning: An integrative review. In *Frontiers in Education*, Vol. 3. Frontiers Media SA, 49.
[37] Rosemary Luckin and Benedict du Boulay. 2016. Reflections on the Ecolab and the Zone of Proximal Development. *International Journal of Artificial Intelligence in Education* 26 (2016), 416–430.
[38] Wannisa Matcha, Nora'ayu Ahmad Uzir, Dragan Gašević, and Abelardo Pardo. 2020. A Systematic Review of Empirical Studies on Learning Analytics Dashboards: A Self-Regulated Learning Perspective. *IEEE Transactions on Learning Technologies* 13, 2 (2020), 226–245. https://doi.org/10.1109/TLT.2019.2916802
[39] Barbara L. McCombs. 1989. *Self-Regulated Learning and Academic Achievement: A Phenomenological View*. Springer New York, 51–82. https://doi.org/10.1007/978-1-4612-3618-4_3
[40] Elgin-Skye McLaren and Alissa N. Antle. 2017. Exploring and Evaluating Sound for Helping Children Self-Regulate with a Brain-Computer Application. In *Proceedings of the 2017 Conference on Interaction Design and Children* (Stanford, California, USA) *(IDC '17)*. Association for Computing Machinery, New York, NY, USA, 393–398. https://doi.org/10.1145/3078072.3084299
[41] Anabil Munshi. 2023. *An Adaptive Scaffolding Framework for Self-Regulated Learning in an Open-Ended Learning Environment*. Ph. D. Dissertation. Vanderbilt University.
[42] Anabil Munshi, Gautam Biswas, Ryan Baker, Jaclyn Ocumpaugh, Stephen Hutt, and Luc Paquette. 2022. Analysing adaptive scaffolds that help students develop self-regulated learning behaviours. *Journal of Computer Assisted Learning* 39, 2 (Dec. 2022), 351–368. https://doi.org/10.1111/jcal.12761
[43] Anabil Munshi, Ramkumar Rajendran, Jaclyn Ocumpaugh, Gautam Biswas, Ryan S. Baker, and Luc Paquette. 2018. Modeling Learners' Cognitive and Affective States to Scaffold SRL in Open-Ended Learning Environments. In *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization*. ACM. https://doi.org/10.1145/3209219.3209241
[44] Omid Noroozi, Paul A Kirschner, Harm JA Biemans, and Martin Mulder. 2018. Promoting argumentation competence: Extending from first-to second-order scaffolding through adaptive fading. *Educational psychology review* 30 (2018), 153–176.
[45] Ernesto Panadero. 2017. A Review of Self-regulated Learning: Six Models and Four Directions for Research. *Frontiers in Psychology* 8 (2017). https://doi.org/10.3389/fpsyg.2017.00422
[46] Dorian Peters. 2014. *Interface design for learning: Design strategies for learning experiences*. Pearson Education.
[47] Xingran Ruan, Charaka Palansuriya, and Aurora Constantin. 2022. Real-time Feedback based on Emotion Recognition for Improving Children's Metacognitive Monitoring Skill. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference* (Braga, Portugal) *(IDC '22)*. Association for Computing Machinery, New York, NY, USA, 672–675. https://doi.org/10.1145/3501712.3538831
[48] Xingran Ruan, Charaka Palansuriya, Aurora Constantin, and Konstantinos Tsiakas. 2023. Supporting Children's Metacognition with a Facial Emotion Recognition based Intelligent Tutor System. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference (IDC '23)*. Association for Computing Machinery, New York, NY, USA, 502–506. https://doi.org/10.1145/3585088.3593882
[49] Lundefinedvija Dana Rukmane, Elisa Rubegni, and Janet Read. 2022. Exploring Failures in Child-Computer Interaction.: Shifting the paradigm from "failures need to be hidden" to "failure is a learning opportunity".. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference* (Braga, Portugal) *(IDC '22)*. Association for Computing Machinery, New York, NY, USA, 492–497. https://doi.org/10.1145/3501712.3535299
[50] Dale H. Schunk. 1990. Goal Setting and Self-Efficacy During Self-Regulated Learning. *Educational Psychologist* 25, 1 (1990), 71–86. https://doi.org/10.1207/s15326985ep2501_6
[51] Dale H Schunk and Peggy A Ertmer. 2000. Self-regulation and academic learning: Self-efficacy enhancing interventions. In *Handbook of self-regulation*. Elsevier, 631–649.
[52] James R Segedy, John S Kinnebrew, and Gautam Biswas. 2013. The effect of contextualized conversational feedback in a complex open-ended learning environment. *Educational Technology Research and Development* 61 (2013), 71–89.

[53] James R Segedy, John S Kinnebrew, and Gautam Biswas. 2015. Using coherence analysis to characterize self-regulated learning behaviours in open-ended learning environments. *Journal of Learning Analytics* 2, 1 (2015), 13–48.

[54] Bianca A Simonsmeier, Maja Flaig, Anne Deiglmayr, Lennart Schalk, and Michael Schneider. 2022. Domain-specific prior knowledge and learning: A meta-analysis. *Educational psychologist* 57, 1 (2022), 31–54.

[55] Michelle Taub, Roger Azevedo, François Bouchet, and Babak Khosravifar. 2014. Can the use of cognitive and metacognitive self-regulated learning strategies be predicted by learners' levels of prior knowledge in hypermedia-learning environments? *Computers in Human Behavior* 39 (2014), 356–367.

[56] Margaret Taub, Roger Azevedo, Aaron E Bradbury, George C Millar, and James Lester. 2018. Using sequence mining to reveal the efficiency in scientific reasoning during stem learning with a game-based learning environment. *Learning and Instruction* 54 (2018), 93–103.

[57] Camilo Vieira, Yi-Yin Seah, and Alejandra J Magana. 2018. Students' experimentation strategies in design: Is process data enough? *Computer Applications in Engineering Education* 26, 5 (2018), 1903–1914.

[58] Karen D. Wang, Jade Maï Cock, Tanja Käser, and Engin Bumbacher. 2022. A systematic review of empirical studies using log data from open-ended learning environments to measure science and engineering practices. *British Journal of Educational Technology* 54, 1 (Nov. 2022), 192–221. https://doi.org/10.1111/bjet.13289

[59] Philip H Winne. 2013. Self-regulated learning viewed from models of information processing. *Self-regulated learning and academic achievement* (2013), 145–178.

[60] Philip H. Winne and Allyson F. Hadwin. 2008. The weave of motivation and self-regulated learning. In *Motivation and self-regulated learning: Theory, research, and applications*. Lawrence Erlbaum Associates Publishers, 297–314.

[61] Philip H Winne and Allyson F Hadwin. 2013. nStudy: Tracing and supporting self-regulated learning in the Internet. In *International handbook of metacognition and learning technologies*. Springer, 293–308.

[62] Zhang Yingbin, Luc Paquette, Ryan S Baker, Jaclyn Ocumpaugh, Nigel Bosch, Gautam Biswas, and Anabil Munshi. 2021. Can Strategic Behaviour Facilitate Confusion Resolution? The Interplay between Confusion and Metacognitive Strategies in Betty's Brain. *Journal of Learning Analytics* 8, 3 (2021), 28–44.

[63] Barry J Zimmerman. 2002. Becoming a self-regulated learner: An overview. *Theory into practice* 41, 2 (2002), 64–70.

[64] Barry J Zimmerman. 2013. Theories of self-regulated learning and academic achievement: An overview and analysis. *Self-regulated learning and academic achievement* (2013), 1–36.

[65] Barry J. Zimmerman and Adam R. Moylan. 2009. Self-Regulation: Where Metacognition and Motivation Intersect. In *Handbook of Metacognition in Education*. Routledge.

[66] Barry J. Zimmerman and Dale H. Schunk. 1989. *Models of Self-Regulated Learning and Academic Achievement*. Springer New York, New York, NY, 1–25. https://doi.org/10.1007/978-1-4612-3618-4_1

[67] Michal Zion and Ruthy Mendelovici. 2012. Moving from structured to open inquiry: challenges and limits. *Science education international* 23, 4 (2012), 383–399.

Figure 7: Code book - Themes for why students were underserved

Themes for Why They Were Underserved	Codes	Interface Role
Challenges in Knowledge Acquisition **(n=15)** 87.5% H-L (n=7) 66.64% L-L (n=8)	**Did not first acquire knowledge** 25% H-L (n=2) 25% L-L (n=3)	System lacks more active support on how struggling students should work to perform information identification and acquisition
	Built solution based only on previous knowledge 37.5% H-L (n=3) 25% L-L (n=3)	System lacks more active support on how struggling students should work to perform information identification and acquisition
	Delayed building the solution 50% H-L (n=4) 33.32% L-L (n=4)	The system does provide feedback, but students can ignore it
	Became overwhelmed by amount of content to read 0% H-L (n=0) 16.66% L-L (n=2)	List of sections, pages and dictionary has no distinguishable hierarchy so the students get confused about the scope of the learning material
	Did not review what was done on previous days 37.5% H-L (n=3) 41.65% L-L (n=5)	System doesn't track that student has been away for a considerable amount of time to suggest a review of what was learned thus far
Challenges to Scaffold Learning **(n=20)** 100% H-L (n=8) 100% L-L (n=12)	**Did not acquire knowledge sequentially** 37.5% H-L (n=3) 83.33% L-L (n=10)	List of sections, pages and dictionary has no distinguishable hierarchy so the students get confused about the scope of the learning material
	Did not take notes while acquiring knowledge 62.5% H-L (n=5) 100% L-L (n=12)	The system does not reinforce the importance of note-taking nor allows the students to check their notes without changing context to another tab
	Did not systematically assess work along the way 100% H-L (n=8) 66.64% L-L (n=8)	System lacks more active support on when/how struggling students should work to do information assessment periodically
	Did not assess work specifically to what was built 50% H-L (n=4) 58.31% L-L (n=7)	There is no system warning advising the student about a possible mistake in assessment choice
	Did not try to correct their understanding before moving on 25% H-L (n=2) 33.32% L-L (n=4)	Even though the system has content to support student's debugging skills, it is not delivered in a timely manner, the student must search for it in another tab
Disregarding System Guidance **(n=13)** 62.5% H-L (n=5) 66.64% L-L (n=8)	**Did not read feedback from the agents** 12.5% H-L (n=1) 50% L-L (n=6)	Oftentimes when a student received feedback from the agent, they didn't shift their gaze to focus on it, suggesting the need for a change in design. The interaction between agent and student is limited to a predefined conversation tree that might not satisfy the student's needs, and there is no mechanism for student and agent to work together and take turns both in dialogue and in action
	Read but ignored feedback from the agents 62.5% H-L (n=5) 50% L-L (n=6)	
	Did not explore reason for incorrect answer after agent feedback 0% H-L (n=0) 50% L-L (n=6)	
Not Leveraging Supporting Tools **(n=13)** 50% H-L (n=4) 75% L-L (n=9)	**Did not explore agents' supportive capabilities** 25% H-L (n=2) 33.32% L-L (n=4)	The system doesn't showcase the assistive capabilities of the agents, the students need to take initiative and start dialogues with the agents to explore and understand how to benefit from them
	Did not investigate additional resources for tips 37.5% H-L (n=3) 66.64% L-L (n=8)	Tips with actionable advice are not easily found by the students because they are at the bottom of a list of pages on a tab that is commonly ignored by students
	Did not use visual tool consistently to reinforce assumptions in the work 37.5% H-L (n=3) 58.31% L-L (n=7)	After the student assess their work, the system only suggests they use the tool once, instead of constantly reminding them
Getting discouraged by incorrect answers **(n=13)** 62.5% H-L (n=5) 66.64% L-L (n=8)	**Doubted system accuracy** 12.5% H-L (n=1) 25% L-L (n=3)	An incomplete understanding from the student's work (shortcut between two concepts) is assessed as wrong by the system
	Repeatedly deleted all work 50% H-L (n=4) 58.31% L-L (n=7)	Even though the system has content to support student's debugging skills, they are not delivered in a timely manner, the student must search for
	Disengaged from the system after failed assessment 25% H-L (n=2) 33.32% L-L (n=4)	The system does not track student's affect to be able to intervene before confusion turns to frustration and eventually to disengagement