

# Towards Automatic Motivator Selection for Autism Behavior Intervention Therapy

Nur Siyam (✉ [nur.siyam@gmail.com](mailto:nur.siyam@gmail.com))

British University in Dubai <https://orcid.org/0000-0003-1625-4892>

Sherief Abdallah

British University in Dubai Faculty of Engineering and IT

---

## Research Article

**Keywords:** Special Education, Autism, Markov Decision Processes, Reinforcement Learning, Behavior Intervention, Intervention Therapy

**Posted Date:** December 29th, 2021

**DOI:** <https://doi.org/10.21203/rs.3.rs-320352/v1>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

Children with autism spectrum disorder (ASD) usually show little interest in academic activities and may display disruptive behavior when presented with assignments. Research indicates that incorporating motivational variables during interventions results in improvements in behavior and academic performance. However, the impact of such motivational variables varies between children. In this paper, we aim to solve the problem of selecting the right motivator for children with ASD using Reinforcement Learning by adapting to the most influential factors impacting the effectiveness of the contingent motivator used. We model the task of selecting a motivator as a Markov Decision Process problem. The states, actions and rewards design consider the factors that impact the effectiveness of a motivator based on Applied Behavior Analysis as well as learners' individual preferences. We use a Q-Learning algorithm to solve the modelled problem. Our proposed solution is then implemented as a mobile application developed for special education plans coordination. To evaluate the motivator selection feature, we conduct a study involving a group of teachers and therapists and assess how the added feature aids the participants in their decision-making process of selecting a motivator. Preliminary results indicated that the motivator selection feature improved the usability of the mobile app. Analysis of the algorithm performance showed promising results and indicated improvement of the recommendations over time.

# Introduction

Autism spectrum disorder (ASD) is a complex neurodevelopmental disorder with an occurrence rate between 1% and 2% (CDC, 2020). ASD is characterized by challenges in communication and social interactions, repetitive and stereotyped behaviors, and limited interests (Schuetze et al., 2017). While there is no known cure for ASD, early intervention has proven to be effective in improving cognitive abilities, language, and adaptive behavior (Dawson et al., 2012).

One challenging aspect in early intervention is the lack of students' interest in academic activities or homework assignments. Students with ASD may resort to behave in a disruptive manner to avoid the academic tasks (L. Koegel et al., 2010). Such disruptive behaviors are considered major barriers to the attainment of educational goals as described in the Individualized Education Program (IEP) of the student. Disruptive behaviors are likely to worsen if left untreated. However, research indicates that incorporating motivational variables into the intervention of a child with ASD leads to improvements in core symptoms of autism and academic areas (L. Koegel et al., 2010; Schuetze et al., 2017). The use of motivational variables, often referred to as reinforcement learning, has been refined into a structured treatment system called Applied Behavioral Analysis (ABA). ABA-based treatment approaches use reinforcement learning to promote desired behavior (such as eye contact) and diminish atypical behaviors (such as repetitive body movements) (Schuetze et al., 2017). To differentiate between Reinforcement Learning as a machine learning paradigm and reinforcement learning as part of the ABA-based treatment approaches, we refer in this paper to the latter as "the use of motivators".

In this paper, we propose the “motivator selection” function and add it to a mobile app to be used by teachers and therapists. The mobile app, “IEP-Connect”, was developed to enable information sharing between different parties involved in the intervention of special education needs and disabilities (SEND) students. IEP-Connect uses the Individualized Educational Plan (IEP) as the main point of coordination (Siyam & Abdallah, 2021). The app allows teachers, therapists and parents to record details of daily sessions and classes regarding students’ progress towards the IEP goals, including behavior monitoring data and motivators used. This information is shared among participating parties to improve learning and therapy outcomes as suggested by earlier research (Siyam, 2018). While sharing information is crucial to the learning process, teachers and therapists need to navigate through the data to determine the right motivator to use for their student in the current session, which may be time consuming and cause overload (Siyam, 2019). This motivates the proposal of a more data-centric approach that systemically evaluates the space of available motivators in order to determine the best action, which is what we propose in this paper.

The aim of our paper is to improve the learning activities and sessions for students with ASD by providing an adaptive decision support system through a mobile app that recommends motivators for teachers and therapists during a learning session. Using the mobile app, teachers record problematic behavior and the system suggests a motivator to be used with the learner. We consider in this study contingent rewards, which are used to reward children when meeting a specified goal and provide positive reinforcement when a task is well done (L. Koegel et al., 2010). Thus, this study aims to answer the following research question:

RQ1: Does adding the “motivator selection” feature to the IEP-Connect app increase the usability of the app?

RQ2: Does student motivation significantly increase when using the “motivator selection” feature compared to the traditional motivator selection methods?

RQ3: When using the RL algorithm for the “motivator selection”, does reward significantly increase over time?

RQ4: When using the RL algorithm for the “motivator selection”, does the episode length (number of steps) significantly decrease over time?

The contribution of this work can be summarized as following:

1. We are the first to model the selection of a motivator as an MDP problem, considering the factors that impact the effectiveness of a motivator according to ABA-based methods and the tradeoff between different motivators.
2. We attempt to solve the motivator selection problem (MSP) using reinforcement learning. While RL is a widely used method for decision support systems, it was not previously implemented for motivators recommendations.

3. We improve the functionality of the IEP-Connect mobile app by adding the “motivator selection” feature. Results from the usability study indicated that the mobile app has excellent usability rate and user satisfaction. These promising results highlight the capabilities of mobile technologies and artificial intelligence in improving the learning and behavior of children with special educational needs.

In the next section, we formally define the motivator selection problem (MSP) that our proposed system aims to solve. We then present the related work by highlighting the most relevant concepts of behavior intervention for autism and then explore the literature on digitized behavior intervention and the use of Reinforcement Learning in special education. In the following section, we map the motivator selection problem into a Markov Decision Process (MDP) and use Q-Learning to solve the modeled problem. Subsequently, we lay out our methodology. We then present the evaluation results in terms of users’ satisfaction and system performance. We conclude the paper with a discussion and implications for practice, research, and future work.

## Motivator Selection Problem (MSP)

Identifying and assessing potential motivating stimuli for students with SEND, and evaluating the methods for maintaining the effectiveness of these motivators, have received interest in the literature (Mechling et al., 2006). As one of the main issues encountered in the therapy of children with ASD is the lack of motivation, many studies aimed to identify the factors that impact the children’s response to different motivators (Schuetze et al., 2017).

Children with ASD do not typically show preference to social stimuli (such as smiles or praise) or affective stimuli (such as a picture of a crying face), which makes it hard to reinforce new behaviors through a natural environment (Schuetze et al., 2017). That is why parents and teachers tend to rely on edible motivators as an alternative. While various studies showed that many skills can be established while using edible stimuli, there are many downsides for delivering edible motivators (Rincover & Newsom, 1985). For instance, providers tend to use food as a common motivator in ABA-based treatment. Therefore, food becomes a therapeutic tool that influences eating behaviors occurring even when the child is not responding to a biological need, resulting in a tendency to overeat. Evidence suggests that children with ASD are at a higher risk for unhealthy weight compared to other children (Matheson & Douglas, 2017). Despite this fact, therapists, teachers and parents continue to use calorically dense foods as a motivator to influence behavior change. Frequent use of food as a motivator has many undesirable side effects that are mostly not found in other motivators. These side effects include long-term health consequences such as weight gain and dental cavities, interruption of activities to administer the motivator, and satiation (Matheson & Douglas, 2017).

Satiation is a decrease in the effectiveness of a stimuli as a motivator when used repeatedly or for a long period of time (Murphy et al., 2003). Research indicates that children satiate more often on edible

motivators. This resulted in more investigations on the advantages of using other types of motivators such as tokens and sensory stimuli (Matheson & Douglas, 2017; Rincover & Newsom, 1985). However, satiation is considered a negative consequence of using any kind of motivator repeatedly. To solve the issue of satiation, therapists and teachers are required to continually vary stimuli and introduce motivators from different modalities, vary the schedules of reinforcement, provide children with choices, and benefit from the use of technology to access preference items to increase motivation (Mechling et al., 2006; Murphy et al., 2003). Many other factors were considered in the literature such as the benefits of sensory stimuli in promoting greater interaction between the child and the environment (Rincover & Newsom, 1985), the reinforcing value of providing choice (Mechling et al., 2006), and the effects of using preferred (specially assessed) motivators (Schuetze et al., 2017).

One of the important steps for a successful deployment of ABA-based interventions is the identification of motivator preference for each student (L. Koegel et al., 2010). However, evidence suggests the need of conducting frequent assessment to reflect the changes in preferences over time. Motivators for students with autism are sensitive to the changing characteristics across time and children, suggesting that a typical motivator that is powerful in one context might not be effective in other settings or with other children (Dyer, 1987). Thus, there is a need for continuous identification and assessment of effective motivators, which requires time and effort from teachers, therapists and parents. Our research aims to provide an automated system that addresses that need.

The MSP is considered a challenging problem as various factors impact the effectiveness of a motivator, including the child herself, the teacher, the subject being taught, the time of day, the type of disruptive behavior, among others. Therapists who are trained on ABA-based treatment approaches usually choose a specific motivator considering the history of student behavior, current behavior, and some internal approximation of the outcome of possible future therapy decisions. They develop reinforcement sampling menus or lists that can help them identify motivators to each identified problem behavior (e.g., aggression, stereotypy, non-compliance) according to the antecedents of the behavior (e.g., change of activity, denied access to preferred item) context variables (e.g., time of day, subject), and the behavior function (e.g., attention seeking, escape). Additionally, therapists take into consideration the student's preferred motivator which is usually identified by using motivator assessments (National Center on Intensive Intervention, 2016). In academics and behavior, teachers and therapists use these data to individualize instruction on a student-by-student basis according to the student's exceptional learning and behavior needs (Fuchs et al., 2014). The left side of **Fig 1** shows the steps in the behavior intervention strategy (ABA) followed by therapists to devise their plans.

However, therapists and teachers face many challenges in developing such plans. First, devising behavior intervention plans requires training and experience. Studies show that the supply of certified ABA providers is low (Y. X. Zhang & Cummings, 2020). Also, while therapists and interventionists may have the needed knowledge, other teachers, especially those in general education classrooms, rely on behavior therapists to recommend appropriate intervention plans (Hudson, 2020). This requires continuous communication and coordination to prevent the misuse of motivators and to ensure the intervention

works as intended. While effective communication is always sought by all parties, it is often lacking due to time restraints or denied access (Siyam, 2018). Second, as with most clinical treatments, the effect of the motivator used with the student is uncertain (non-deterministic). This uncertainty makes it hard to plan ahead as attempting to predict the effect of a series of treatment over time compounds the uncertainty (Bennett & Hauser, 2013). The process of navigating student-based data and plan for the right intervention is not only time consuming, but requires considering various variables and continuous access to research and heuristic investigations (Bennett & Hauser, 2013). Moreover, there has been a deal of contradicting evidence regarding the use of “extrinsic” motivation to engage and motivate students (Schuetze et al., 2017). We follow in this paper the popular position in this regard in that the application of planned positive motivation is a critical element of teaching children with ASD. Teachers, therefore, are required to sustain students’ motivation using both “intrinsic” and “extrinsic” reinforcement. Motivation is considered an internal “intrinsic” psychological state. Reinforcement, on the other hand, can be intrinsic to the task, extrinsically applied, or both. The challenge is, therefore, in deciding when the use of different types of motivators is effective or even necessary (Healey, 2008).

## **Related Work**

The increased rate of ASD diagnosis in recent years (CDC, 2020) has fueled the research in the area of machine learning with the purpose of improving the learning experience of those affected. The focus of research has been mainly on developing academic or social skills learning applications (Foster et al., 2010; Roman et al., 2018), improving diagnosis efficiency (Kosmicki et al., 2015), and modelling social and behavioral aspects of ASD (Stevens et al., 2017). However, we are not aware of any research that applied reinforcement learning to solve the MSP. The following sections review related work on digitized behavior management and reinforcement learning, in an attempt to identify the gap between available technology tools and the need to provide solutions for therapy recommendations in special education.

## **Digitized Behavior Intervention**

There are many available applications that allow therapists, teachers and parents to monitor the behavior of children with special needs (Marcu et al., 2013; Vannest et al., 2011). These applications allow the people involved with the intervention of children with ASD to track, store and share important information. This information is used to plan for interventions, monitor progress towards IEP objectives, and generate reports. While these applications are very helpful and a good replacement of paper-based data collection, the data collected in special needs settings is usually complex, unstandardized and incomplete (Marcu et al., 2013). Many studies suggested using data mining techniques to support intervention decisions (Thabtah, 2019). For instance, Burns et al. (2015) developed a mobile app that employed association rule mining to reveal pattern in behavior causes and effects to inform the therapists decisions. In the study, parents use a mobile app to collect Antecedent, Behavior and Consequence (ABC) data. The data mining techniques aimed to identify behavior causes and effects patterns to enable therapists to improve intervention. Linstead et al. (2016) introduced the Autism Management Platform (AMP), an integrated

health care information system for managing data related to the diagnosis and treatment of children with ASD. The authors developed a mobile application to facilitate information and multimedia sharing between parents and clinicians. The system also includes a web interface and analytics platforms, allowing specialists to mine patient data in real-time. The analytics platform uses machine learning techniques to provide users with personalized data searching preferences. Bhuyan et al. (2017) studied temporal data to identify factors that aid caregivers in creating an effective intervention plan and predict the right treatment based on the data in other contexts.

Previous studies also focused on using mobile technology to help children with ASD and their caregivers in regulating challenging behaviors. For instance, Crutchfield et al. (2015) evaluated the impact of the I-Connect app on stereotypy in adolescents with ASD in a school setting. Préfontaine et al. (2019) developed the iSTIM app to support parents of younger children with ASD in reducing stereotypy behavior. The app was evaluated and found successful in regulating stereotypy behavior when used by trained researches as well as parents who do not have the required ABA training (Trudel et al., 2020). In another related study, Begoli et al. (2013) aimed to develop a computational representation for ABA to serve as a reasoning foundation for intelligent-agent mediated therapies by formulating the representation of ABA concepts as a process ontology. Concepts that are relevant to the reasoning and operations functions aspects of the agents (e.g., rewarding and prompting) were represented in the ontology and then were formalized as a Belief-Desire-Intention (BDI) reasoning framework. Such formalization is feasible because of the procedural, repetitive and prescriptive nature of ABA (Begoli, 2014).

## Reinforcement Learning (RL)

As a subfield of machine learning, RL has been widely implemented, resulting in its increasing applicability in real-life problems and decision-support systems (Yu et al., 2020). For instance, RL has been used to improve the delivery of personalized care by optimizing medication choices, medicine doses, and intervention timings (S. Liu et al., 2020). In the healthcare and therapy domain, data is characterized by its high dimensionality and complex interdependencies (Gräßler et al., 2017). RL has the potential to automatically explore various treatment options by analyzing patient data to derive a policy and personalized therapy without the need of pre-established rules (S. Liu et al., 2020).

Recommender systems has also been leveraged using RL. Recommender systems based on RL have the advantage of updating the policies during online interaction, which enables the system to generate recommendations that best suit users evolving preferences (Zhao et al., 2019). Examples include news (Zheng et al., 2018), music recommendations (Hong et al., 2020) and personalized learning systems (Shawky & Badawi, 2019).

RL has proven to be an appropriate framework for interaction modeling and optimization of problems that can be formulated as MDPs. The advantage of such methods is the ability to model the stochastic variation of outcomes as transition probabilities between states and action (Tsiakas et al., 2016). RL and

MDPs have been successfully applied to personalized learning systems (Sayed et al., 2020; Shawky & Badawi, 2019), intelligent tutoring systems (Barnes & Stamper, 2008; Stamper et al., 2013), adaptive serious games for ASD (Khabbaz et al., 2017) and robot assisted therapy (Tsiakas et al., 2016). For instance, Bennane (2013) automated the selection of the content of a tutoring system and its pedagogical approach to provide differentiated instruction. Similarly, Shawky and Badawi (2019) used RL to build an intelligent environment to provide learners with suitable content as well as adapt to the learner's evolving states. Khabbaz et al. (2017) proposed an adaptive serious game for rating social ability in children with ASD using RL. The game adapts itself according to the level of the ASD child by adjusting the difficulty level of the activities. In the field of Robot Assisted Therapy, Tsiakas et al. (2016) proposed an Interactive RL framework that adapts to the user's preferences and refine its learned policy when coping with new users.

In this work, we aim to develop an app that can be used by any of the child caregivers in any setting. Moreover, we aim to provide teachers and therapists with a tool that facilitates intervention planning once a problematic behavior was detected through the recommendation of motivators using RL. Unlike previous studies, we rely on online learning rather than on previously collected data. While online learning does not benefit from the offline repetitive training period, it allows the model to adjust the policies to match the non-stationary environment and individuality of each child with SEND (P. Liu & Chen, 2017).

## Solving the MSP

The aim of this work is to leverage the power of RL to solve the problem of selecting the best motivator for each intervention session. We first model the MSP as a Markov Decision Process (MDP) problem. By using MDPs, the proposed model can explicitly model future rewards, which will benefit the motivator recommendation accuracy significantly in the long run. MDPs can address many of the challenges faced in therapy decision-making. We then apply RL by using Q-Learning to solve the modeled problem.

## Markov Decision Processes (MDP)

The MSP can be formulated as an MDP. An MDP is a standard formalization of sequential decision making, which is widely used for applications where an autonomous agent interacts with its surrounding environment through actions. An MDP can be defined as a four-tuple;  $(\delta, A, P, R)$ , where  $\delta$  is a set of states called the state space  $A$ , is a set of actions called the action space,  $P$  is the state transition function, which is the probability of transitioning between every pair of states given an action, and  $R$  is the reward function that assigns an immediate reward after transitioning to a new state due to an action (Sutton & Barto, 2018).

The agent, which is situated in the therapists or teacher mobile application, interacts with the environment at discrete time steps. In our example, a time step is considered each time a therapist records a behavior in the mobile application. At each time, the agent receives a state  $S_t$  from the

environment from a set of possible states,  $\delta$ . Based on this state, the agent selects an action  $A_t$  from a set of valid actions  $A$  in state  $S_t$ . Actions in our example are motivators the therapist can use to motivate the student. Based in part on the agent's action, the agent finds itself in a new state  $S_{t+1}$  one time-step later. The environment also provides the agent a scalar reward  $R_{t+1}$  from a set of possible rewards,  $R$ . The reward in our example will depend on whether the student becomes motivated and to which degree, among other factors that are explained in the next sections. The transition  $(s_t, a_t, r_{t+1}, s_{t+1})$  is stored in memory  $M$ . The ultimate variation of this system aims at enhancing the learning and therapy experience of the child by recommending the right motivator (Sutton & Barto, 2018).

The agent-environment interaction produces a trajectory of experience consisting of state-action-reward tuples. Actions influence immediate rewards as well as future states and, therefore, future rewards. When the agent takes an action in a state, the transition dynamics function  $p(s', r | s, a)$ , formalizes the state transition probability. This produces the probability of transitioning to state  $s'$  with reward  $r$ , from state  $s$  when taking action  $a$ .

## Modelling MSP as an MDP

The research suggests that, in various clinical settings, modeling treatment decisions through MDPs is effective and can yield better results than therapists' intuition alone (Bennett & Hauser, 2013). However, there are no previous attempts to model the MSP and an MDP. Careful formulation of the problem and state/action space is essential to obtain satisfactory results and satisfy the Markov assumption that the current timepoint ( $t$ ) is dependent only on the previous time point ( $t-1$ ) (Sutton & Barto, 2018). **Fig 1** shows how the MSP problem, represented by the ABA intervention, is mapped as an MDP in this work. The following sub-sections describe how each component of the MDP is used to model the MSP problem.

### 1. State

One of the most challenging and critical issues in designing the MDP model is to properly identify the factors that influence the effectiveness of a motivator, especially when these factors may differ from one child to another. The personalization of intervention can be achieved by carefully determining these features that represent the state space (Shawky & Badawi, 2019). Through careful investigation of the research that investigates motivation stimuli for students with ASD, the features outlined in Table 1 were considered:

Table 1: Features representing the state space

Feature	Description	Number of values	Reference
<b>Contextual features</b>			
Antecedent event (trigger)	Event or activity that immediately preceded a problem behavior (alone, given a direction or demand, transitioned to new activity, denied access to an item)	4	(Bhuyan et al., 2017; Stichter et al., 2009)
Time of Day	Time of the day the problem behavior occurred (morning, noon, evening)	3	(Burns et al., 2015)
Subject	The aim of this feature is to account to the place and person the problem behavior occurred with (academic subjects, therapy sessions, home)	8	(Burns et al., 2015)
<b>Behavior</b>			
Behavior	The problem behavior that requires intervention, grouped into seven categories (aggression, self-injury, disruption, elopement, stereotypy, tantrums, non-compliance)	7	(Stevens et al., 2017)
Behavior Function	The reason the behavior is occurring (sensory stimulation, escape, access to attention, access to tangibles)	4	(Alstot & Alstot, 2015)
<b>History</b>			
Last unsuccessful motivator	The ID of the last motivator used that was not successful in motivating the student within an episode, including an option for "none".	7	
Motivator past usage	The number of times each motivator was used within a week grouped in categories of <5, 5-10, >11. This factor is composed of six features according to the number of motivators (actions) available (edibles, sensory, activities, tokens, social, choice).	3 <sup>6</sup>	(Çetin, 2021)

Therapists and teachers aim to identify appropriate intervention for multiple settings. However, these interventions may fail if no attention is given to contextual differences (Stichter et al., 2009). Contextual features such as antecedent events, time of day, and location (where and with whom) all impact the child response to a proposed intervention, therefore informing optimal motivators. Moreover, while interventionists aim to track and remediate problem behaviors, the ability to understand the reason behind the occurrence of a behavior is essential as the behavior itself for creating appropriate behavior plans (Schaeffer, 2018).

Problem behaviors in special education are numerous and diverse. In this study, challenging behaviors are grouped into eight widely observed behaviors (Stevens et al., 2017); aggression (e.g. hitting, biting), self-injury (e.g. head-banging, hitting walls), disruption (e.g. yelling, knocking things over), elopement (e.g. wandering, escaping), stereotypy (e.g. rocking, hand-flapping), tantrums (e.g. crying, screaming), non-compliance (e.g. whining, defying orders), obsession (e.g. constantly talking about same topic).

Keeping track of the last ineffective motivator used is essential in our problem definition to maintain the Markov property where future state and reward depends only on the current state and action (Sutton & Barto, 2018). We consider this feature a part of the state to prevent suggesting the same motivator repeatedly. Moreover, we keep track of the number of times a motivator group was used to prevent satiation (Matheson & Douglas, 2017; Rincover & Newsom, 1985). While studies have shown that extrinsic reward does not directly harm a child's intrinsic motivation (Cameron & Pierce, 1994), we consider the repeated long-use of tangible rewards, such as edibles or tokens, to have a negative impact when not carefully administered, and therefore should be limited (Witzel & Mercer, 2003).

## 2. Actions

There has been a controversy regarding what type of reward best motivate children with SEND to follow routines and complete academic tasks without negatively impacting their future behavior. Nevertheless, there is a strong evidence that rewarded children report higher intrinsic motivation than the nonrewarded ones (Cameron & Pierce, 1994).

However, the dilemma regarding which motivator is best suited for each intervention withstands. There are many factors that impact the choice of the right contingent reward (motivator) during a therapy or academic session. According to ABA techniques, there is a need to address what happens before the behavior, what is the behavior itself, and what is done immediately after the behavior. In this study, the goal is to recommend an action (contingent motivator) that can be given to the student after completing a certain task or complying to a certain command. The teacher or therapist needs to decide on which motivator to use from a list of six motivators categories (see Table 2); edibles, sensory, activities, tokens, social, and choice (Çetin, 2021). For example, if a student is yelling to get the teacher's attention, the teacher may promise the student a favorite food item (edible) if the student stops yelling and completes her task. Alternatively, the teacher may assign a leadership role (social) as a motivator to the student once she is done with the activity. If another student is wandering to escape a task, the teacher may promise extra computer time (activity) once the student completes the task in hand. Therapists also consider the long-term effect of the motivator. For example, edible items, especially unhealthy choices, should be avoided. Repetitive use of the same motivator should be avoided as well to prevent satiation. Experienced interventionists sometimes use the same motivator for a specific period of time to establish a routine but change it later to prevent the student dependency on that particular reward to complete the tasks.

Table 2: Motivators Categories

Motivator	Description
Edible	Food items, such as fruits, snacks, and juice.
Sensory	Items or activities that realizes pleasure to the senses of the child, such as listening to music, sitting in a rocking chair, or playing with sand.
Activity	Activities may include drawing, playing with the computer, or jumping on a trampoline.
Token	Tangible items that the child values, such as stickers, money, or stars on an honor chart.
Social	Attention or interaction with another person, such as high-fives, smiles, and praise.
Choice	Giving the child the chance to choose between two different items or methods, such as asking whether she prefers to use a pencil or crayons to write.

### 3. Rewards

The reward in our problem definition is the measure of student motivation after introducing the motivator. We adopt in this study the subjective measure of responsiveness proposed by Koegel and Egel (1979) as shown in Table 3. The teacher or therapist rates the student’s responsiveness after introducing a motivator and carrying out an activity.

Table 3: Scale of child’s responsiveness (adapted from Koegel and Egel (1979))

Output	Description	Reward
Negative	Child continues problem behavior (tantrums, kicking, screaming) or does not comply with instructions and engages in behavior unrelated to the activity (rocking, yawning, tapping).	-1
Neutral	Complies with instructions but tends to get restless or loses attention.	+2
Positive	Performs task readily. Attends to task quickly, smiles while doing the task, and presents appropriate behavior.	+4
Rejected recommendation	The user rejects the motivator recommendation and does not introduce it to the child.	-0.25
Edible item	The motivator selected was an edible item.	-1
Token item	The motivator selected was a token item.	-0.5

Each student responsiveness category results in the agent receiving a reward, as shown in Table 3. The agent receives a reward of -1 if the motivator did not work or the student response was negative. Alternatively, it receives a +2 if the response was neutral, or +4 if the response was positive. If the

caregiver chooses not to follow the recommendation, the reward is -0.25. In formulating the problem, we also aim to balance two competing objectives; receiving positive responsiveness from the student, and limiting long-term exposure to unhealthy items. The definition of “safe Reinforcement Learning” has been proposed in the literature, especially for recommender systems that aim to balance user’s satisfaction and the avoidance of recommending harmful items like violent movies (Heger, 1994). Therefore, the agent receives a penalty of -1 when recommending edibles and -0.5 for recommending tokens.

## Q-Learning

To solve the proposed MDP problem, a Q-learning algorithm with an epsilon-greedy (*e*-greedy) policy with linearly decreasing exploration rate was used. Q-learning is an off-policy, value-based RL algorithm that aims to find the best action to take according to the current state. Q-learning seeks to learn a policy that maximizes the total reward. Q-learning is considered off-policy as the Q-learning function learns from actions chosen according to a behavior policy that differs from the updated policy. A policy is equivalent to an ABA-based intervention protocol with the advantage of capturing more individualized details of students. In our case, the agent chooses an action according to an *e*-greedy policy, while learning the optimal policy. *e*-greedy is a method used to balance exploration and exploitation, where epsilon (*e*) refers to the probability of choosing to explore (i.e., choosing a random action) rather than exploit (i.e., choosing the optimal action). The policy is represented by a table that maps all possible states with actions. While following the *e*-greedy policy, the agent exploits with a probability of (1-*e*) and with a probability of exploring of (*e*). This probability (*e*) decays over time by some rate as the agent learns more about the environment. The agent will become “greedy” in terms of exploiting and the probability of exploration becomes less. If the agent becomes “well-trained”, it is possible to select the best action given the state. This process is described as acting according to an optimal policy (Sutton & Barto, 2018).

The reward is an estimation of the scores the state *S* receives under the action *a*, which is denoted as *Q* (*s*, *a*) and updated based on Equation 1 (Q-Learning Function), which is based on Bellman’s optimality equation (Bellman, 1966).

$$Q(s, a) \leftarrow Q(s, a) + \alpha (r + \gamma \max_{a'} Q(s', a') - Q(s, a)) \quad (1)$$

where  $\alpha$  is the learning rate,  $r$  is the observed reward,  $s'$  is the new state,  $\gamma < 1$  is the discount factor for the future rewards received, and  $Q(s', a')$  is the estimation of the maximum reward that can be obtained by taking some future action in the state  $s'$ . The learning process can continue for any number of episodes. In our case, we consider the end of the episode when the student becomes motivated. The Q-learning algorithm can be found in Appendix A.

While it seems straightforward to apply standard learning algorithms to learn the agent’s optimal policy and then use it to recommend motivators to the user, this approach cannot be applied in practice to our

problem. Unlike traditional reinforcement tasks such as Atari games (Mnih et al., 2015), therapy recommendation tasks cannot benefit from the possibility of interacting with the user repeatedly to obtain any amount of experience to update the policy towards an optimal one (Lei & Li, 2019). Moreover, there is no previously collected data to train the algorithm offline before the online interaction. Therefore, we do not vary the experimental parameters in this study.

Additionally, this study is considered to have a “cold start” as all the values in the Q-Table were set to “zero” before the deployment phase. A cold start can be considered problematic as it bothers users by requiring too many interactions for collecting enough experience for learning (C. Zhang et al., 2021). On the other hand, online learning is beneficial for therapy recommendations due to the highly dynamic nature of children preferences and responses to intervention. Moreover, online learning allows us to obtain user’s feedback by tracking whether the suggested motivator was used or not (Arzate Cruz & Igarashi, 2020; P. Liu & Chen, 2017).

Each episode starts when a caregiver records a behavior on the mobile app. Then, it is terminated by reaching the final state in which the student becomes motivated. We use a learning rate  $\alpha$  of 0.1 and a discount  $\gamma$  of 0.95. We apply an  $\epsilon$ -greedy policy that starts with a high  $\epsilon$  of 0.9 to encourage state exploration. Then, it decays exponentially with a rate of 0.99 until it reaches 0.05.

As shown in **Fig 3**, on each time step  $t$ , the therapist or teacher records a behavior instance and requests a motivator recommendation. The agent takes the feature representation of the current state and recommends a motivator using  $\epsilon$ -greedy policy. The caregiver then administers the intervention and provides feedback by rating the response of the student. Alternatively, the caregiver can choose not to use the recommended motivator if deemed unappropriated or skip the recommendation if the item is not available (e.g., edible items) or cannot be applied to the current activity (e.g., choice). When the agent chooses to exploit, it selects an action by selecting the highest  $Q(s,a)$  for the observed state from the Q-table. Otherwise, the agent “explores” by selecting a random action.

## Methodology

To evaluate the proposed model, we deployed a mobile app that can be used by caregivers. This mobile app has been developed and tested in our previous study as an app that facilitates communication and coordination between different parties involved in the therapy and learning of students with ASD (Siyam & Abdallah, 2021). In this study, we added to the mobile app the feature of selecting a motivator. This feature can be used by the teacher, the therapist, and the parent during a learning or therapy session.

To test the usability of the app and the efficiency of the proposed RL model, we asked 12 teachers and therapists of students registered in a private school to maintain behavioral monitoring data using the mobile app without the “Motivator Selection” feature for four weeks. In this phase, caregivers decided what motivator to use for each case according to their knowledge and experience and recorded this information along with the success of the motivator in the app (see **Fig 2**). After the four weeks period, we administered the System Usability Scale (SUS) questionnaire, which is a quick measure of the perceived

usability of the system (Brooke, 1996). We then updated the app to include the “Motivator Selection” feature, and asked the caregivers to use it for four weeks as well (see **Fig 3**). When the users (therapists or teachers) use the “Motivator Selection” feature, they first record the problem behaviour, the antecedent of the behaviour, and the behaviour function by choosing from drop-down lists (**Fig 3(a)**). After clicking the “Suggest Motivator” button, the RL algorithm suggests a motivator to be used (**Fig 3(b)**). If the user “Declines” the motivator suggested, the algorithm will suggest a new motivator. If the user “Accepts” the motivator suggested, the mobile app will prompt the user to choose the student’s response to the motivator from a drop-down list (**Fig 3(c)**). Once the response is entered, the app takes the user back to the “Motivator Suggestion” screen (**Fig 3(a)**) to record a new behaviour, choose another student, or exit the app. We then administered the System Usability Scale (SUS) again, and compared the results to the SUS score previously obtained to answer the first research question (Siyam & Abdallah, 2021). To evaluate the performance of the Q-Learning algorithm, we first applied the Chi-square test to answer the second research question. We then applied Correlation and Regression analysis to answer the third and fourth research questions.

The study design was limited by the settings of teachers’ assignment to classes. As most participating teachers and therapist taught all of the participating students, it was hard to conduct control trials and avoid spillovers. Spillovers occur when a treatment affects those in the control group (Baird et al., 2017). In our case, there was no way to assign the participants randomly into two separate groups. First, students who are in a group may have a teacher who is teaching students in both groups, which will prevent us from measuring the true usability value of the feature. On the other hand, if teachers were assigned to either group, they may teach students who have teachers from both groups, which will prevent us from measuring the impact of the RL policy on the student motivation. Therefore, this study follows a one-group pretest-posttest quasi-experimental research design, in which the same variables (i.e. usability of the app and students motivation) are measured in one group of participants before and after using the “Motivator Selection” app feature (Kirk, 2012). The term “quasi” indicates that the design resembles experimental research. However, since participants are not randomly assigned to different groups, quasi-experimental design is considered non-experimental research and lacks the advantages of having a control group. However, an advantage of this method is that it can be used to conveniently assess an intervention on target participants. Moreover, this design allows for statistical analysis of data using recognized methods (Stratton, 2019).

## Participants

Consent was obtained from 12 parents of children with ASD studying at the same private school. The consent was obtained to use the child’s data in the mobile app and allow teachers and therapists to use the mobile app to track academic progress and behavior of the child. The age of the children ranged from 6 to 9 years and were diagnosed with a middle-range (Level 2) to severe (Level 3) form of ASD (see Table 4). All the participating students attended self-contained classrooms for students with SEND in a private K-12 school. Self-contained classrooms are usually separated from general education classrooms but

within the same school building (Spencer, 2013). All participating children met the following criteria: (1) attended self-contained classrooms for students with SEND, (2) had an IEP, and (3) had a comprehensive behavioral plan. These students were taught by special education teachers in addition to specialized behavioral and operational therapists.

Table 4: Learners' demographics

Category	N
<b>Learners' Gender</b>	
Female	1
Male	11
<b>Learner's Age</b>	
6	3
7	5
8	2
9	2
<b>Level of ASD</b>	
Level 2	5
Level 3	7

Teachers and therapists of these twelve students were invited to participate in the study by assessing the quality of the recommendations they received from the app. This resulted in the participation of 10 teachers, and 2 therapists (see Table 5). The participating therapists have extensive experience in the use of behavioral therapy techniques with students with SEND in general, and with students with ASD in particular. On the other hand, participating teachers did not receive any formal training in the use of behavioral therapy techniques but had varied experience dealing with children with ASD in terms of tracking behavior and using motivators inside the class.

Table 5: Teachers and therapists demographics

<b>Category</b>	<b>N</b>
<b>Gender</b>	
Female	11
Male	1
<b>Nationality</b>	
Egyptian	7
Jordanian	3
Syrian	2
<b>Job Description</b>	
Subject Teacher	10
Occupational Therapist	1
Behaviour Specialist	1

## Ethical Considerations

One challenge for applying RL in experimental settings is exploration. In other domains, such as game playing and movies recommendations, experiments can be repeated as many times as needed. In our clinical setting, the RL agent has to learn online with limited previously collected data. Using trial and error to explore all possible states may conflict with therapy and education ethics (S. Liu et al., 2020). However, in this study, the caregiver has the ability to dismiss any suggested motivator, either because of its unavailability, or because of the caregiver belief that the motivator suggested will not be effective. Therefore, the motivator suggestion in this study aims to augment the decision-making process of the therapist, rather than replace it.

Moreover, the process of ABA-based therapy requires that the therapist varies between the motivator choices. This sometimes requires the therapist to try motivators that may not work. As the caregiver maintains the control of what motivators to use, this study does not pose any identifiable or foreseeable risk to any participant outside of normal daily risks.

Finally, the use of the app does not require caregivers to change any of the normal learning activities and environments where participating students usually engage.

## Usability Evaluation

To evaluate the performance of the of the applicability of the application and RL algorithm, and to answer the first research question, we conducted a user study by administering the SUS questionnaire (see Appendix B). The SUS questionnaire was developed in 1996 by Brooke as a quick measure of the perceived usability of a system (Brooke, 1996). Numerous studies confirmed the validity and reliability of the questionnaire for various platforms and with a limited number of participants (Bangor et al., 2008; Lewis & Sauro, 2009). The questionnaire in this study was presented to participants in two forms, an English version and an Arabic version (AlGhannam et al., 2018). To calculate the SUS score, the score contribution for each item was calculated. The average score of all participants responses is the value of the system usability (Brooke, 1996).

## RQ1: Does adding the “motivator selection” feature to the IEP-Connect app increase the usability of the app?

To answer the first research question, we compared the average SUS score for the IEP-Connect before using the “Motivator Selection” feature, which was 80.42, and the average SUS score for the IEP-Connect app after introducing the “Motivator Selection” feature, which was 84.38. The score is considered to have improved, as shown in **Fig 4**.

## Q-Learning Algorithm Performance Evaluation

Compared to other machine learning algorithms, there is an absence of an agreed upon performance evaluation standards for RL (S. Liu et al., 2020). While this problem is not unique to RL, it is harder to address compared to other machine learning algorithms that rely on accuracy and precision recall as a performance indicator. Calculating the precision and accuracy of an algorithm usually requires an offline dataset to be divided into training and testing sets. As this study does not benefit from an offline dataset, the effectiveness of the proposed algorithm is evaluated through statistical and qualitative analysis (Stratton, 2019).

## Descriptive Summary

During the first four weeks, participants manually recorded the use of motivators through the behavior monitoring page in app. In this phase, teachers and therapists presented students with motivators according to the treatment approaches they usually follow. They also recorded whether the motivator was effective or not (**Fig 2**). During this period, teachers and therapists entered 490 behavior monitoring entries. Out of those entries, 223 (45.5%) contained motivators that worked and 267 (54.5%) contained motivators that did not work. Table 6 presents the descriptive summary for the dataset collected before using the RL algorithm.

Table 6: Descriptive summary for "no-algorithm" dataset variables

Variable	Total	Per	Mean	SD	Min	Max
Users	12	Research				
Students	12	Research				
Days	30	Research				
Entries	490	User (Therapists & Teachers)	40.833	40.653	6	141

In the following four weeks, participants were asked to use the “motivator selection” feature in the app to aid them in identifying the right motivator to use rather than manually choose the motivators. In this phase, teachers and therapists recorded the behavior problem, the antecedent, and the behavior function (**Fig 3**). Then, the app would suggest a motivator to be used. The therapist or teacher can decide whether to use the motivator suggested, or reject it and ask for another suggestion. Once using the motivator, the user would record the student response to the monitor. Each entry is an instance where the app suggested a motivator to be used. If the user “declines” the motivator (i.e., choose not to use it), the app will immediately propose a new motivator. The new recommended motivator becomes a new entry. This continues until the user “accepts” a motivator (i.e., uses the suggested motivator). If the motivator accepted is effective (i.e., the student becomes motivated), the episode ends. If the motivator accepted is not effective (i.e., the student does not become motivated), the user can request another motivator to use or record a new behavior problem. An episode constitutes of the steps where a motivator was “declined” or not effective until the motivator used is marked as effective.

Table 7 presents the descriptive summary for the dataset collected after using the RL algorithm. An episode considered all the steps until the student becomes motivated, including the motivators that were dismissed and not used. This resulted in 598 episodes, with an average of 2.06 steps per episode. **Fig 5** shows the percentage of effective, not effective, and declined motivators when using the motivator selection featured compared to when therapists and teachers selected the motivator by themselves

Table 7: Descriptive summary for the "algorithm" dataset variables

Variable	Total	Per	Mean	SD	Min	Max
Users	12	Research				
Students	12	Research				
Days	32	Research				
Entries	1231	User (Therapists & Teachers)	102.58	98.72	8	376
Episodes	598	Day	38.469	26.39	1	91
Steps (Episode Length)		Episode	2.06	2.35	1	22
Reward					-2.00	4.00
Reward Sum		Episode	2.68	1.478	-4.50	4.00

## Algorithm Evaluation

Three statistical techniques were applied to evaluate the performance of the Q-Learning algorithm: Chi-square test, Correlation, and Regression analysis. Significance level was set by the researcher at  $\alpha = 0.05$ .

### RQ2: Does student motivation significantly increase when using the “motivator selection” feature compared to the traditional motivator selection methods?

To compare the effectiveness of selecting motivators with and without using the “motivator selection” feature, the entries that included a motivator that was actually used are considered. This results in 671 entries. Out of those entries, 602 were motivators that worked and 69 were motivators that did not work. To answer the research question, a chi-square test of independence was conducted to test whether or not the application of the algorithm has a significant effect on motivating students. The test revealed that applying the algorithm is significantly associated with more motivated students, *Pearson*  $\chi^2 = 265.16$  and  $p < 0.001$ . The crosstabulation in Table 8 indicates that 45.5% of students were motivated without applying the algorithm, while 89.6% were motivated with algorithm being applied. The contingency coefficient was 0.432,  $p < 0.001$ , indicates that students being motivated is significantly associated with the algorithm application, and that this association is relatively strong.

Table 8. Crosstabulation of Algorithm vs. Status

		Status		Total
		Not Motivated	Motivated	
Algorithm	Without	267	223	490
		54.5%	45.5%	100.0%
	With	69	594	663
		10.4%	89.6%	100.0%
Total		336	817	1153
		29.1%	70.9%	100.0%

Pearson  $\chi^2 = 265.16, p < 0.001$ . Symmetric Measures: Contingency Coefficient = 0.432,  $p < 0.001$ .

cells (0.0%) have expected count less than 5. The minimum expected count is 142.79.

A binomial logistic regression was also performed to ascertain the effect of algorithm application on the likelihood that students would be motivated. The logistic regression model was statistically significant,  $\chi^2(1) = 273.337, p < 0.001$ , as shown in Table 9. The model explained 30.1% (Nagelkerke  $R^2$ ) of the variance in students' motivation and correctly classified 74.7% of cases. Applying algorithm is 10.307 times more likely to motivate students more than not applying it.

Table 9. Summary of Logistic Regression Analysis for Algorithm Application Predicting Students' Motivation

Variable	B	SE	Wald	df	Sig.	Exp(B)
algo(1)	2.333	.156	222.987	1	.000	10.307
Constant	-.180	.091	3.940	1	.047	.835

$-2LL = 1118.132, \text{Nagelkerke } R^2 = .301, \chi^2 = 273.337, df = 1, p < 0.001$ .

Classification accuracy = 74.7%

### RQ3: When using the RL algorithm for the “motivator selection”, does reward significantly increase over time?

The reward data for the applied algorithm was plotted against time so that it would be easy to observe the pattern of reward data, as shown in the scatterplot in Fig 6. Correlation analysis revealed that there is a significant positive association between reward and sequence,  $r = 0.081, p = 0.036$ . Although the association is weak (in magnitude), it is positive and statistically significant, indicating that reward

significantly increases over time. This allows to proceed to regression analysis that shows how reward increases over time.

## Regression Analysis of Reward

One of the aims of the study is to find whether rewards increase over time; i.e., to test the hypothesis: H-0: Time (sequence) has a positive impact on Rewards. Given the dependent variable (Rewards), and the independent variable (time sequence), a simple linear regression analysis was conducted to investigate how reward is predicted over time. The results of regression analysis, reported in Table 10, shows that the regression mode is significant,  $F(1,661) = 4.412$ ,  $p = 0.036$ , and explains 0.7% of the variance in Reward. The t test of the regression coefficient shows a significant predictor,  $t = 2.101$ ,  $p = 0.036$ . That is, time that refers to each step taken to apply the algorithm, is a significant contributor to the positive change in reward. In other words, each one step taken to apply the algorithm would cause an increase in reward by 0.00037 (or from 0.000024 to 0.00071). Using G\*Power 3.1.9.7 (Faul et al., 2009), effect size  $\eta^2$  was estimate to be 0.007 and power  $(1-\beta)$  to be 0.579.

Table 10. Regression Analysis Summary for Sequence Predicting Reward

Variable	B	95% CI	$\beta$	t	p
(Constant)	2.386	[2.125 , 2.647]		17.944	< .001
Sequence	.000367	[.000024 , .000710]	.081	2.101	.036

$R^2 = .007$ ,  $R^2_{adj} = .005$ ,  $CI = Confidence Interval for B$ .

## Deeper Regression Analysis of Reward

Considering that a reward higher than 2 is already high, a simple linear regression analysis was conducted using data for reward 2 or below. The analysis produced a significant regression model,  $F(1,257) = 7.519$ ,  $p = 0.007$ , with  $R^2 = 0.028$ . The correlation coefficient between reward and sequence is equal to 0.169, which is a stronger value than 0.081. Moreover, investigating the regression coefficient of 0.001 indicates that low rewards significantly increase over time; that is, for each new step, reward increases by 0.001. The regression line is shown in **Fig 7**. Using G\*Power 3.1.9.7 (Faul et al., 2009), effect size and test power were estimated. With 259 cases,  $\alpha = 0.05$ , and  $R^2 = 0.028$ , effect size  $\eta^2$  was estimate to be 0.029 and power  $(1-\beta)$  to be 0.777.

Table 11. Regression Analysis Summary for Sequence Predicting Reward (< 3)

Variable	B	95% CI	$\beta$	t	p
(Constant)	.475	[.136 , .813]		2.764	.006
Sequence	.000627	[.000177 , .001077]	.169	2.742	.007

$R^2 = .028$ ,  $R^2_{adj} = .025$ ,  $CI = \text{Confidence Interval for } B$ .

## RQ4: When using the RL algorithm for the “motivator selection”, does the episode length (number of steps) significantly decrease over time?

In order to answer this question, scatterplot of episode length against time is investigated first to check whether there is a negative pattern on length over time (see **Fig 8**). The scatterplot shows a negative pattern of episode lengths over time as the fit line is moving downward. A correlation analysis revealed that there is a significant negative association between episode length and sequence,  $r = -0.161$ ,  $p < 0.001$ . This encourages to run regression analysis to discover how length decreases over time.

### Regression Analysis of Episode Length

A simple linear regression analysis was conducted to investigate how episode length changes based on sequence. A significant regression equation was found,  $F(1,596) = 15.917$ ,  $p < 0.001$ , with an  $R^2$  of 0.026. The predicted episode length is equal to  $2.758 - 0.001(\text{sequence})$  steps for each episode. That is, episode length decreases by 0.001 (or between 0.000518 and 0.001524) steps for each episode. This indicates that an episode would need a smaller number of steps over time. Using G\*Power 3.1.9.7 (Faul et al., 2009), effect size and test power were estimated. With 598 cases,  $\alpha = 0.05$ , and  $R^2 = 0.026$ , effect size  $f^2$  was estimate to be 0.027 and power  $(1-\beta)$  to be 0.979.

Table 12. Regression Analysis Summary for Sequence Predicting Episode Length

Variable	B	95% CI	$\beta$	t	p
(Constant)	2.758	[2.366 , 3.149]		13.841	< .001
Sequence	-.001021	[-.001524 , -.000518]	-.161	-3.990	< .001

$R^2 = .026$ ,  $R^2_{adj} = .024$ ,  $CI = \text{Confidence Interval for } B$ .

### Deeper Regression Analysis of Episode Length

As number of steps per an episode of 1 or 2 steps is already a small number, regression analysis was conducted using data of episodes with a number of steps greater than 2. The analysis revealed that the association between episode length and sequence is stronger. The produced regression model was significant,  $F(1,126) = 4.951$ ,  $p = 0.028$ , with  $R^2 = 0.038$ . Moreover, the regression coefficient  $B = -0.002$  indicates that for each new episode, the number of steps significantly decreases by 0.002. This causal relationship can be seen in **Fig 9**. Using G\*Power 3.1.9.7 (Faul et al., 2009), effect size and test power

were estimated. With 128 cases,  $\alpha = 0.05$ , and  $R^2 = 0.038$ , effect size  $\eta^2$  was estimate to be 0.040 and power ( $1-\beta$ ) to be 0.607.

Table 13. Regression Analysis Summary for Sequence Predicting Episode Length (> 2)

Variable	B	95% CI	$\beta$	t	p
(Constant)	6.482	[5.312 , 7.652]		10.965	< .001
Sequence	-.001903	[-.003596 , -.000210]	-.194	-2.225	.028

$R^2 = .038$ ,  $R^2_{adj} = .030$ ,  $CI = \text{Confidence Interval for } B$ .

## Discussion

This study aimed to address the problem of selecting motivators to be used in a learning setting with learners with ASD. To this aim, the problem of selecting a motivator was modelled as an MDP problem. The factors that impact the effectiveness of a motivator were considered based on Applied Behavior Analysis as well as learners' individual preferences. The states, actions and rewards were designed through careful consideration of the research that investigates motivation stimuli for learners with ASD. The MDP problem was then solved using reinforcement learning and added as a feature to a mobile app (IEP-Connect) to aid therapists and teachers in selecting the right motivator to use.

Prior to adding the "motivator section" feature to the IEP-Connect mobile app, teachers and therapists used the app to record the progress of the learner toward the learning objectives, as well as record the behavior of the learners along the motivators the therapists selected according to their own methods. To test the usability of the mobile app with the added "motivator selection" feature, we compared the SUS scores before and after adding the feature. The SUS can be considered a standard instrument for iterative software usability evaluation (Bangor et al., 2008). In our study, SUS scores were collected as part of each iteration of the mobile app and related to the new features encompassing the software. This provided an effective way of determining if the tool design is becoming more or less usable with each iteration. However, statistical analysis was not performed on the SUS scores to determine whether the change was significant. This is because participants' answers to the SUS questionnaires were anonymous, making the pre/posttest statistical analysis not possible. Another limitation of the method used is that quasi-experimental methods are susceptible to threats to internal validity, especially those related to observing the same participants over time. Without a control group, it cannot be concluded with a certain degree that the treatment is what caused the change in scores (Stratton 2019). On the other hand, the positive trends identified in this study indicate that the app usability is headed in the right direction. This is because the introduction of new features in a technology tool usually results in an initial drop on usability score as users become accustomed to the new introduction (Bangor et al., 2008).

We then investigated whether students' motivation increases when using the RL-model compared to when therapists and teachers selected the motivator without the aid of the "motivator selection" feature. While this study is limited with the number of participants as well as the study period, the entries analysis gave us insights on how well the new feature supported the decision-making process of the mobile app users. When comparing the number of motivators administered that worked when the therapists used their own methods and when they used the mobile app feature, the results indicated that using the "motivator selection" feature significantly increased the times learners were motivated. While these results demonstrate the feasibility of RL to support the decision-making process in regards of the motivator selection problem, the reliability of these results requires further investigation. This is due to the fact that the "motivator selection" feature was used by a homogenous group of users from the same context. Many factors might affect the results if the study were to be conducted in other contexts. One of the main factors that may impact the results is the experience of the users. In this study, ten of the participants did not have experience in ABA-based methods and usually relied on the therapists to recommend motivators to use. As communication between the different people involved with the intervention of a learner with ASD is considered a challenge (Siyam, 2021; Siyam & Abdallah, 2021), teachers usually find themselves using the same motivator repeatedly. The repeated administration of a motivator causes sanitation. On the other hand, the RL-model was designed in a way that prevents suggesting the same motivator repeatedly.

Another point worth discussing is concerned with the "declined" suggestions when using the "motivator selection" feature. The number of entries that were "declined" was (560). The high number of declined suggestions is due to many reasons. First, teachers and therapists declined the suggested motivator when the item suggested was not available. The highest number of motivators declined was for "edible" motivators. Edible motivators might be easy to administer at home. However, in the school, these are not available unless previously planned. On the other hand, other items were declined when the teacher or therapist believed that the suggested motivator will not work or is hard to administer at the moment. Moreover, teachers declined motivators they were not familiar with or did not have experience in administering them.

While the number of declined suggestions can be considered high, they were not considered as "Not Effective" motivators in the analysis for various reasons. First, these motivators were not tested to determine whether they were effective or not in a particular setting. Moreover, we cannot obtain similar data on what motivators that were "thought off" and not used when the therapists and teachers chose the motivators themselves. Additionally, the aim of this study is not to measure the accuracy of the recommendations. Rather, we hypothesize that the use of the motivator selection feature in the mobile app in tandem with the judgment of the therapist will result in the use of motivators that work more often than when the therapist does not use this feature. Thus, we assume that the RL-based model can augment the therapists and teachers' capabilities, rather than replace their role. However, as with any application of machine learning on intervention and therapy, the long-term impact of the use of the proposed model on the motivation of learners requires longer study periods and a larger number of participants.

Finally, the regression analysis of the data obtained when using the RL algorithm aimed to investigate whether the episode reward improves and the episode length declines over time as the model is trained. The results showed that the change in reward and episode length over time is significant. However, this change is considered small. This limitation is a result of the short period of time the algorithm was trained. The low  $R^2$  value indicates that the independent variable (sequence) is not explaining much of the variation in the dependent variable (rewards), regardless of the significance. That is, sequence is not accounting for much of the variance of rewards. The  $R^2$  value can be improved by extending the study period, as we believe that with the continued use of the algorithm through the app, the performance of the algorithm will continue to improve. Moreover, the effect size is considered small. However, even small effect sizes can have scientific clinical significance, depending on the study field (Grace-Martin, 2013).

## Implications

This research is part of continuous work in developing a mobile application that facilitates collaboration and communication between different stakeholders working with students with SEND.

This study took place during the COVID-19 pandemic, which resulted in disruptions such as schools closing or shifting to distance learning. This unique context highlighted the issues in information sharing and collaboration between school and home regarding children's academic and behavioral plans (Spiller, 2020). Moreover, it was reported that during distance learning, students in general, and students with special needs in particular felt less motivated and engaged with their lessons (Beulah, 2020).

Additionally, behavioral plans had to change according to whether the student is in a distance learning, blended learning, or on-site learning program. However, the challenges in behavior management coordination are not exclusive to distance learning settings. Many studies reported issues in coordination between school and home, where the communication is often one-sided (Siyam, 2018). Moreover, teachers and therapists often lack the time and resources needed to train parents on effective behavior management techniques (Spiller, 2020).

With so many challenges in place, it became more apparent that an innovative and dynamic way of communication and progress monitoring should be adopted. Moreover, as parents' responsibility in managing their children behavior increased, the need for specialized decision support systems for parents or caregivers without the needed experience became increasingly apparent. The use of the IEP-Connect app with its "Motivator Selection" feature can address the inconsistency in behavior management techniques between home and school.

The results shown in this study demonstrate the feasibility of RL to support the decision-making process of caregivers of students with ASD in regards of the motivator selection problem. Despite its early stages of development, the proposed "motivator selection" feature performs significantly better than current conventional motivator selection methods. A distinct advantage of RL over other machine learning approaches is that users' preferences guide the choice of the RL algorithm, which in return optimizes the

reward function, resulting in higher reward values over time. The results indicated that the episode rewards increased over time, which indicates that the agent learnt to maximize its total reward earned. However, the episode reward did not level out at a high reward per episode value, meaning that the agent is not yet behaving optimally at every state. This is due to the large state space compared to the number of episodes the agent experienced. We believe that with the continued use of the algorithm through the app, the performance of the algorithm will continue to improve until it acts optimally at every state.

The proposed RL algorithm was based on scientific research which considered ABA practices, health concerns, and satiation. Moreover, the algorithm encouraged caregivers to incorporate motivational components in academic tasks and give children a choice, which resulted in higher motivation rates. These results add to the existing literature regarding the use of motivators for behavior management. Moreover, the results indicate that the introduction of the “Motivator Selection” increased the usability score of the app. The modelling approach using MDP can be further improved to provide more accurate and personalized motivator selections. The continuous collection of data resulting from ABA sessions will provide valuable data for research on children with ASD behavior and how they are motivated in different settings (Burns et al., 2015). Data mining techniques can be used on the collected data to uncover fundamental patterns regarding the optimal motivator magnitude and to predict the likelihood of the child response to intervention. For example, children with inconsistent or reduced response to motivators may require additional training or therapeutic sessions (Bennett & Hauser, 2013; Schuetze et al., 2017). Moreover, patterns can be extracted from such data to understand and explain problematic and reoccurring behavior or antecedents, which will help in return minimize the situations where such behavior occurs. Using data mining techniques can offer capabilities beyond human comprehension, such as considering numerous contextual features (Burns et al., 2015).

While RL has the potential to offer significant contribution in the area of therapy decision-making for special needs, there are certain key issues that need to be addressed, such as clinical implementations and ethics (S. Liu et al., 2020). RL-based policies face many challenges before they can be deployed to inform clinical decision-making. One of the most common limitations is the challenge in obtaining training data, both in terms of time and cost. Thus, the number of trajectories obtained is usually limited compared to those resulting from simulations. Besides the number of trajectories limitation, another factor that impact the estimation of the value function in this type of studies is that the type of data collected on different children can be extremely variable, both across children and within a child over time (Shortreed et al., 2011). Since the proposed RL algorithm is run in a real scenario without the benefit of a training period, the large number of possible state-action pairs has to be considered to avoid coverage problems. The RL models in clinical settings require iterative refining to include new data from various resources as well as longer training periods to increase the system’s knowledge-base. Using data from limited resources may result in biased algorithms that do not apply to all scenarios. Additionally, among the raised ethical concerns of AI algorithms is the ownership of generated data and the right to benefit from them (Shawky & Badawi, 2019).

The number of participants and the study period may be considered limitations. Moreover, this study did not consider a personalized model for each learner due to the large state space compared to the number of participants. Therefore, there is a need for a dynamic adaptation mechanism that allows the agent to efficiently refine its policy towards each learner (Tsiakas et al., 2016).

Moreover, future work should consider the interpretability of produced recommendations. The system should be able not only to suggest suitable motivators, but to explain such choices to the user (Nunes & Jannach, 2017). “Explainable” AI systems have been proven to increase users trust and acceptance of the system (Abdi et al., 2020). Other possible directions for future work include modeling the RL algorithm as a partially observable MDP where data are mapped to some state space that truly represents students’ behavior and context features.

The proposed RL framework can be integrated into a multi-agent ecosystem that aims to improve the coordination among the stakeholders of an IEP. The agents in the system will be able to “learn” over time and adapt to the real-world variation. Such systems can aid in the decision-making process of education providers while developing updated knowledge about effective special education practices (Bennett & Hauser, 2013).

## Declarations

**Funding:** No funding was received for conducting this study.

**Conflicts of interest/Competing interests:** None

**Availability of data and material:** The data collected and analyzed during the current study are available from the corresponding author on request.

**Code availability:** The reinforcement learning algorithm code is available from the corresponding author on request.

**Ethics approval:** Approval was obtained from the Research Ethics Committee in the British University in Dubai.

**Consent to participate:** All participants (parents and teachers) provided written informed consent prior to enrolment in the study.

**Consent for publication:** Not applicable.

## References

1. Abdi, S., Khosravi, H., Sadiq, S., & Gasevic, D. (2020). Complementing educational recommender systems with open learner models. *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, 360–365. <https://doi.org/10.1145/3375462.3375520>

2. AlGhannam, B. A., Albustan, S. A., Al-Hassan, A. A., & Albustan, L. A. (2018). Towards a Standard Arabic System Usability Scale: Psychometric Evaluation using Communication Disorder App. *International Journal of Human–Computer Interaction*, 34(9), 799–804. <https://doi.org/10.1080/10447318.2017.1388099>
3. Alstot, A. E., & Alstot, C. D. (2015). Behavior Management: Examining the Functions of Behavior. *Journal of Physical Education, Recreation & Dance*, 86(2), 22–28. <https://doi.org/10.1080/07303084.2014.988373>
4. Arzate Cruz, C., & Igarashi, T. (2020). A Survey on Interactive Reinforcement Learning: Design Principles and Open Challenges. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference* (pp. 1195–1209). Association for Computing Machinery. <https://doi.org/10.1145/3357236.3395525>
5. Baird, S., Bohren, J. A., McIntosh, C., & Ozler, B. (2017). *Optimal Design of Experiments in the Presence of Interference* (SSRN Scholarly Paper ID 2900967). Social Science Research Network. <https://doi.org/10.2139/ssrn.2900967>
6. Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human–Computer Interaction*, 24(6), 574–594. <https://doi.org/10.1080/10447310802205776>
7. Barnes, T., & Stamper, J. (2008). Toward Automatic Hint Generation for Logic Proof Tutoring Using Historical Student Data. In B. P. Woolf, E. Aïmeur, R. Nkambou, & S. Lajoie (Eds.), *Intelligent Tutoring Systems* (pp. 373–382). Springer. [https://doi.org/10.1007/978-3-540-69132-7\\_41](https://doi.org/10.1007/978-3-540-69132-7_41)
8. Begoli, E. (2014). *Procedural-Reasoning Architecture for Applied Behavior Analysis-based Instructions*. [https://trace.tennessee.edu/utk\\_graddiss/2749](https://trace.tennessee.edu/utk_graddiss/2749)
9. Begoli, E., Ogle, C. L., Cihak, D. F., & MacLennan, B. J. (2013). Towards an Integrative Computational Foundation for Applied Behavior Analysis in Early Autism Interventions. In H. C. Lane, K. Yacef, J. Mostow, & P. Pavlik (Eds.), *Artificial Intelligence in Education* (pp. 888–891). Springer. [https://doi.org/10.1007/978-3-642-39112-5\\_138](https://doi.org/10.1007/978-3-642-39112-5_138)
10. Bellman, R. (1966). Dynamic Programming. *Science*, 153(3731), 34–37. <https://doi.org/10.1126/science.153.3731.34>
11. Bennane, A. (2013). Adaptive Educational Software by Applying Reinforcement Learning. *Informatics in Education - An International Journal*, 12(1), 13–27.
12. Bennett, C. C., & Hauser, K. (2013). Artificial intelligence framework for simulating clinical decision-making: A Markov decision process approach. *Artificial Intelligence in Medicine*, 57(1), 9–19. <https://doi.org/10.1016/j.artmed.2012.12.003>
13. Beulah, J. (2020). *Progress Monitoring Through The Lens of Distance Learning*. <https://red.mnstate.edu/thesis/398>
14. Bhuyan, F., Lu, S., Ahmed, I., & Zhang, J. (2017). Predicting efficacy of therapeutic services for autism spectrum disorder using scientific workflows. *2017 IEEE International Conference on Big Data (Big Data)*, 3847–3856. <https://doi.org/10.1109/BigData.2017.8258388>

15. Brooke, J. (1996). SUS-A quick and dirty usability scale. In *Usability Evaluation In Industry: Vol. 189(194)* (pp. 4–7).
16. Burns, W., Donnelly, M., & Booth, N. (2015). Mining for Patterns of Behaviour in Children with Autism Through Smartphone Technology. In C. Bodine, S. Helal, T. Gu, & M. Mokhtari (Eds.), *Smart Homes and Health Telematics* (pp. 147–154). Springer International Publishing.  
[https://doi.org/10.1007/978-3-319-14424-5\\_16](https://doi.org/10.1007/978-3-319-14424-5_16)
17. Cameron, J., & Pierce, W. D. (1994). Reinforcement, Reward, and Intrinsic Motivation: A Meta-Analysis. *Review of Educational Research*, *64*(3), 363–423.  
<https://doi.org/10.3102/00346543064003363>
18. CDC. (2020). *Data and Statistics on Autism Spectrum Disorder | CDC*. Centers for Disease Control and Prevention. <https://www.cdc.gov/ncbddd/autism/data.html>
19. Çetin, M. E. (2021). Determination of Reinforcement Usage Strategies during Literacy Education of Teachers Working with Students with Multiple Disabilities in Turkey. *International Journal of Education and Literacy Studies*, *9*(1), 25–32. <https://doi.org/10.7575/aiac.ijels.v.9n.1p.25>
20. Crutchfield, S. A., Mason, R. A., Chambers, A., Wills, H. P., & Mason, B. A. (2015). Use of a Self-monitoring Application to Reduce Stereotypic Behavior in Adolescents with Autism: A Preliminary Investigation of I-Connect. *Journal of Autism and Developmental Disorders*, *45*(5), 1146–1155.  
<https://doi.org/10.1007/s10803-014-2272-x>
21. Dawson, G., Jones, E. J. H., Merkle, K., Venema, K., Lowy, R., Faja, S., Kamara, D., Murias, M., Greenson, J., Winter, J., Smith, M., Rogers, S. J., & Webb, S. J. (2012). Early Behavioral Intervention Is Associated With Normalized Brain Activity in Young Children With Autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(11), 1150–1159.  
<https://doi.org/10.1016/j.jaac.2012.08.018>
22. Dyer, K. (1987). The competition of autistic stereotyped behavior with usual and specially assessed reinforcers. *Research in Developmental Disabilities*, *8*(4), 607–626. [https://doi.org/10.1016/0891-4222\(87\)90056-4](https://doi.org/10.1016/0891-4222(87)90056-4)
23. Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G\*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160.  
<https://doi.org/10.3758/BRM.41.4.1149>
24. Foster, M. E., Avramides, K., Bernardini, S., Chen, J., Frauenberger, C., Lemon, O., & Porayska-Pomsta, K. (2010). Supporting children's social communication skills through interactive narratives with virtual characters. *Proceedings of the International Conference on Multimedia - MM '10*, 1111.  
<https://doi.org/10.1145/1873951.1874163>
25. Fuchs, D., Fuchs, L. S., & Vaughn, S. (2014). What Is Intensive Instruction and Why Is It Important? *TEACHING Exceptional Children*, *46*(4), 13–18. <https://doi.org/10.1177/0040059914522966>
26. Gräßer, F., Beckert, S., Küster, D., Schmitt, J., Abraham, S., Malberg, H., & Zaunseder, S. (2017). Therapy Decision Support Based on Recommender System Methods. *Journal of Healthcare Engineering*, *2017*, e8659460. <https://doi.org/10.1155/2017/8659460>

27. Healey, J. B. (2008). Extrinsic Reinforcers as a Critical Component of Education for Motivating Students With Special Needs. *Motivation and Practice for the Classroom*, 143–163.  
[https://doi.org/10.1163/9789087906030\\_009](https://doi.org/10.1163/9789087906030_009)
28. Heger, M. (1994). Consideration of Risk in Reinforcement Learning. In W. W. Cohen & H. Hirsh (Eds.), *Machine Learning Proceedings 1994* (pp. 105–111). Morgan Kaufmann.  
<https://doi.org/10.1016/B978-1-55860-335-6.50021-0>
29. Hong, D., Li, Y., & Dong, Q. (2020). Nonintrusive-Sensing and Reinforcement-Learning Based Adaptive Personalized Music Recommendation. *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, 1721–1724.  
<https://doi.org/10.1145/3397271.3401225>
30. Hudson, H. (2020). *Teaching Students with Autism Spectrum Disorders: Are Teachers Truly Prepared?* [Ed.D., Wilmington University (Delaware)].  
<http://search.proquest.com/pqdtglobal/docview/2425898207/abstract/39201764A57E4FE4PQ/1>
31. Khabbaz, A. H., Pouyan, A. A., Fateh, M., & Abolghasemi, V. (2017). An adaptive RL based fuzzy game for autistic children. *2017 Artificial Intelligence and Signal Processing Conference (AISP)*, 47–52.  
<https://doi.org/10.1109/AISP.2017.8324105>
32. Kirk, R. E. (2012). Experimental Design. In *Handbook of Psychology, Second Edition*. American Cancer Society. <https://doi.org/10.1002/9781118133880.hop202001>
33. Koegel, L., Singh, A. K., Koegel, R. L., & Koegel, L. (2010). Improving Motivation for Academics in Children with Autism. *Journal of Autism and Developmental Disorders*, 40(9), 1057–1066.  
<https://doi.org/10.1007/s10803-010-0962-6>
34. Koegel, R., & Egel, A. (1979). Motivating autistic children. *Journal of Abnormal Psychology*, 88(4), 418–426. <https://doi.org/10.1037/0021-843X.88.4.418>
35. Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Translational Psychiatry*, 5(2), e514. <https://doi.org/10.1038/tp.2015.7>
36. Lei, Y., & Li, W. (2019). Interactive Recommendation with User-Specific Deep Reinforcement Learning. *ACM Transactions on Knowledge Discovery from Data*, 13(6), 61:1-61:15.  
<https://doi.org/10.1145/3359554>
37. Lewis, J. R., & Sauro, J. (2009). The Factor Structure of the System Usability Scale. In M. Kurosu (Ed.), *Human Centered Design* (pp. 94–103). Springer. [https://doi.org/10.1007/978-3-642-02806-9\\_12](https://doi.org/10.1007/978-3-642-02806-9_12)
38. Linstead, E., Burns, R., Nguyen, D., & Tyler, D. (2016). AMP: A platform for managing and mining data in the treatment of Autism Spectrum Disorder. *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2545–2549.  
<https://doi.org/10.1109/EMBC.2016.7591249>
39. Liu, P., & Chen, M. (2017). Performance Evaluation of Recommender Systems. *International Journal of Performability Engineering*, 13(8), 1246. <https://doi.org/10.23940/ijpe.17.08.p7.12461256>

40. Liu, S., See, K. C., Ngiam, K. Y., Celi, L. A., Sun, X., & Feng, M. (2020). Reinforcement Learning for Clinical Decision Support in Critical Care: Comprehensive Review. *Journal of Medical Internet Research, 22*(7), e18477. <https://doi.org/10.2196/18477>
41. Marcu, G., Tassini, K., Carlson, Q., Goodwyn, J., Rivkin, G., Schaefer, K. J., Dey, A. K., & Kiesler, S. (2013). Why do they still use paper? Understanding data collection and use in Autism education. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 3177–3186*. <https://doi.org/10.1145/2470654.2466436>
42. Matheson, B. E., & Douglas, J. M. (2017). Overweight and Obesity in Children with Autism Spectrum Disorder (ASD): A Critical Review Investigating the Etiology, Development, and Maintenance of this Relationship. *Review Journal of Autism and Developmental Disorders, 4*(2), 142–156. <https://doi.org/10.1007/s40489-017-0103-7>
43. Mechling, L. C., Gast, D. L., & Cronin, B. A. (2006). The Effects of Presenting High-Preference Items, Paired With Choice, Via Computer-Based Video Programming on Task Completion of Students With Autism. *Focus on Autism and Other Developmental Disabilities, 21*(1), 7–13. <https://doi.org/10.1177/10883576060210010201>
44. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., Graves, A., Riedmiller, M., Fidjeland, A. K., Ostrovski, G., Petersen, S., Beattie, C., Sadik, A., Antonoglou, I., King, H., Kumaran, D., Wierstra, D., Legg, S., & Hassabis, D. (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533. <https://doi.org/10.1038/nature14236>
45. Murphy, E. S., McSweeney, F. K., Smith, R. G., & McComas, J. J. (2003). Dynamic Changes in Reinforcer Effectiveness: Theoretical, Methodological, and Practical Implications for Applied Research. *Journal of Applied Behavior Analysis, 36*(4), 421–438. <https://doi.org/10.1901/jaba.2003.36-421>
46. National Center on Intensive Intervention. (2016). *Reinforcement Strategies*. U.S. Department of Education, Office of Special Education Programs. <https://intensiveintervention.org/intervention-resources/behavior-strategies-support-intensifying-interventions#reinforcement>
47. Nunes, I., & Jannach, D. (2017). A systematic review and taxonomy of explanations in decision support and recommender systems. *User Modeling and User-Adapted Interaction, 27*(3), 393–444. <https://doi.org/10.1007/s11257-017-9195-0>
48. Préfontaine, I., Lanovaz, M. J., McDuff, E., McHugh, C., & Cook, J. L. (2019). Using Mobile Technology to Reduce Engagement in Stereotypy: A Validation of Decision-Making Algorithms. *Behavior Modification, 43*(2), 222–245. <https://doi.org/10.1177/0145445517748560>
49. Rincover, A., & Newsom, C. D. (1985). The Relative Motivational Properties of Sensory and Edible Reinforcers in Teaching Autistic Children. *Journal of Applied Behavior Analysis, 18*(3), 237–248. <https://doi.org/10.1901/jaba.1985.18-237>
50. Roman, J., Mehta, D. R., & Sajja, P. S. (2018). Multi-agent Simulation Model for Sequence Generation for Specially Abled Learners. In S. C. Satapathy & A. Joshi (Eds.), *Information and Communication*

- Technology for Intelligent Systems (ICTIS 2017)–Volume 1* (pp. 575–580). Springer International Publishing.
51. Sayed, W. S., Gamal, M., Abdelrazek, M., & El-Tantawy, S. (2020). Towards a Learning Style and Knowledge Level-Based Adaptive Personalized Platform for an Effective and Advanced Learning for School Students. In M. H. Farouk & M. A. Hassanein (Eds.), *Recent Advances in Engineering Mathematics and Physics* (pp. 261–273). Springer International Publishing.  
[https://doi.org/10.1007/978-3-030-39847-7\\_22](https://doi.org/10.1007/978-3-030-39847-7_22)
  52. Schaeffer, M. (2018). *What Motivates You: Smiles or Stickers?: Extrinsic Vs. Intrinsic Motivators in a Self-contained Special Education Classroom* [Thesis, Trinity Christian College].  
<https://search.proquest.com/openview/137af1da3c1662e00f9f7ff703a82f67/1?pq-origsite=gscholar&cbl=18750&diss=y>
  53. Schuetze, M., Rohr, C. S., Dewey, D., McCrimmon, A., & Bray, S. (2017). Reinforcement Learning in Autism Spectrum Disorder. *Frontiers in Psychology, 8*. <https://doi.org/10.3389/fpsyg.2017.02035>
  54. Shawky, D., & Badawi, A. (2019). Towards a Personalized Learning Experience Using Reinforcement Learning. In A. E. Hassanien (Ed.), *Machine Learning Paradigms: Theory and Application* (pp. 169–187). Springer International Publishing. [https://doi.org/10.1007/978-3-030-02357-7\\_8](https://doi.org/10.1007/978-3-030-02357-7_8)
  55. Shortreed, S. M., Laber, E., Lizotte, D. J., Stroup, T. S., Pineau, J., & Murphy, S. A. (2011). Informing sequential clinical decision-making through reinforcement learning: An empirical study. *Machine Learning, 84*(1), 109–136. <https://doi.org/10.1007/s10994-010-5229-0>
  56. Siyam, N. (2018). Special Education Teachers' Perceptions on Using Technology for Communication Practices. *Journal for Researching Education Practice and Theory (JREPT), 1*(2), 6–18.  
<https://doi.org/10.5281/zenodo.2537590>
  57. Siyam, N. (2019). Factors impacting special education teachers' acceptance and actual use of technology. *Education and Information Technologies, 24*(3), 2035–2057.  
<https://doi.org/10.1007/s10639-018-09859-y>
  58. Siyam, N. (2021). *Using Mobile Technology for Coordinating Educational Plans and Supporting Decision Making Through Reinforcement Learning in Inclusive Settings* [Thesis, The British University in Dubai (BUiD)]. <https://bspace.buid.ac.ae/handle/1234/1879>
  59. Siyam, N., & Abdallah, S. (2021). A Pilot Study. Investigating the Use of Mobile Technology for Coordinating Educational Plans in Inclusive Settings. *Journal of Special Education Technology, 1*–14. <https://doi.org/10.1177/01626434211033581>
  60. Spencer, T. D. (2013). Self-contained Classroom. In F. R. Volkmar (Ed.), *Encyclopedia of Autism Spectrum Disorders* (pp. 2721–2722). Springer. [https://doi.org/10.1007/978-1-4419-1698-3\\_84](https://doi.org/10.1007/978-1-4419-1698-3_84)
  61. Spiller, A. N. (2020). *Understanding Stakeholder Communication and Coordination for Children with Behavioral Needs* [Thesis, University of Michigan].  
<https://deepblue.lib.umich.edu/handle/2027.42/162555?show=full>
  62. Stamper, J., Eagle, M., Barnes, T., & Croy, M. (2013). Experimental Evaluation of Automatic Hint Generation for a Logic Tutor. *International Journal of Artificial Intelligence in Education, 22*(1–2), 3–

17. <https://doi.org/10.3233/JAI-130029>
63. Stevens, E., Atchison, A., Stevens, L., Hong, E., Granpeesheh, D., Dixon, D., & Linstead, E. (2017). A Cluster Analysis of Challenging Behaviors in Autism Spectrum Disorder. *2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 661–666. <https://doi.org/10.1109/ICMLA.2017.00-85>
64. Stichter, J. P., Randolph, J. K., Kay, D., & Gage, N. (2009). The Use of Structural Analysis to Develop Antecedent-based Interventions for Students with Autism. *Journal of Autism and Developmental Disorders*, *39*(6), 883–896. <https://doi.org/10.1007/s10803-009-0693-8>
65. Stratton, S. J. (2019). Quasi-Experimental Design (Pre-Test and Post-Test Studies) in Prehospital and Disaster Research. *Prehospital and Disaster Medicine*, *34*(6), 573–574. <https://doi.org/10.1017/S1049023X19005053>
66. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction* (2nd edition). MIT Press.
67. Thabtah, F. (2019). Machine learning in autistic spectrum disorder behavioral research: A review and ways forward. *Informatics for Health and Social Care*, *44*(3), 278–297. <https://doi.org/10.1080/17538157.2017.1399132>
68. Trudel, L., Lanovaz, M. J., & Préfontaine, I. (2020). Brief Report: Mobile Technology to Support Parents in Reducing Stereotypy. *Journal of Autism and Developmental Disorders*. <https://doi.org/10.1007/s10803-020-04735-6>
69. Tsiakas, K., Dagioglou, M., Karkaletsis, V., & Makedon, F. (2016). Adaptive Robot Assisted Therapy Using Interactive Reinforcement Learning. In A. Agah, J.-J. Cabibihan, A. M. Howard, M. A. Salichs, & H. He (Eds.), *Social Robotics* (pp. 11–21). Springer International Publishing. [https://doi.org/10.1007/978-3-319-47437-3\\_2](https://doi.org/10.1007/978-3-319-47437-3_2)
70. Vannest, K. J., Burke, M. D., Payne, T. E., Davis, C. R., & Soares, D. A. (2011). Electronic Progress Monitoring of IEP Goals and Objectives. *TEACHING Exceptional Children*, *43*(5), 40–51. <https://doi.org/10.1177/004005991104300504>
71. Witzel, B. S., & Mercer, C. D. (2003). Using Rewards to Teach Students with Disabilities: Implications for Motivation. *Remedial and Special Education*, *24*(2), 88–96. <https://doi.org/10.1177/07419325030240020401>
72. Yu, C., Liu, J., & Nemati, S. (2020). Reinforcement Learning in Healthcare: A Survey. *ArXiv:1908.08796 [Cs]*. <http://arxiv.org/abs/1908.08796>
73. Zhang, C., Wang, S., Aarts, H., & Dastani, M. (2021). Using Cognitive Models to Train Warm Start Reinforcement Learning Agents for Human-Computer Interactions. *ArXiv:2103.06160 [Cs]*. <http://arxiv.org/abs/2103.06160>
74. Zhang, Y. X., & Cummings, J. R. (2020). Supply of Certified Applied Behavior Analysts in the United States: Implications for Service Delivery for Children With Autism. *Psychiatric Services*, *71*(4), 385–388. <https://doi.org/10.1176/appi.ps.201900058>
75. Zhao, X., Zhang, L., Xia, L., Ding, Z., Yin, D., & Tang, J. (2019). Deep Reinforcement Learning for List-wise Recommendations. *ArXiv:1801.00209 [Cs, Stat]*. <http://arxiv.org/abs/1801.00209>

76. Zheng, G., Zhang, F., Zheng, Z., Xiang, Y., Yuan, N. J., Xie, X., & Li, Z. (2018). DRN: A Deep Reinforcement Learning Framework for News Recommendation. *Proceedings of the 2018 World Wide Web Conference*, 167–176. <https://doi.org/10.1145/3178876.3185994>

## Figures

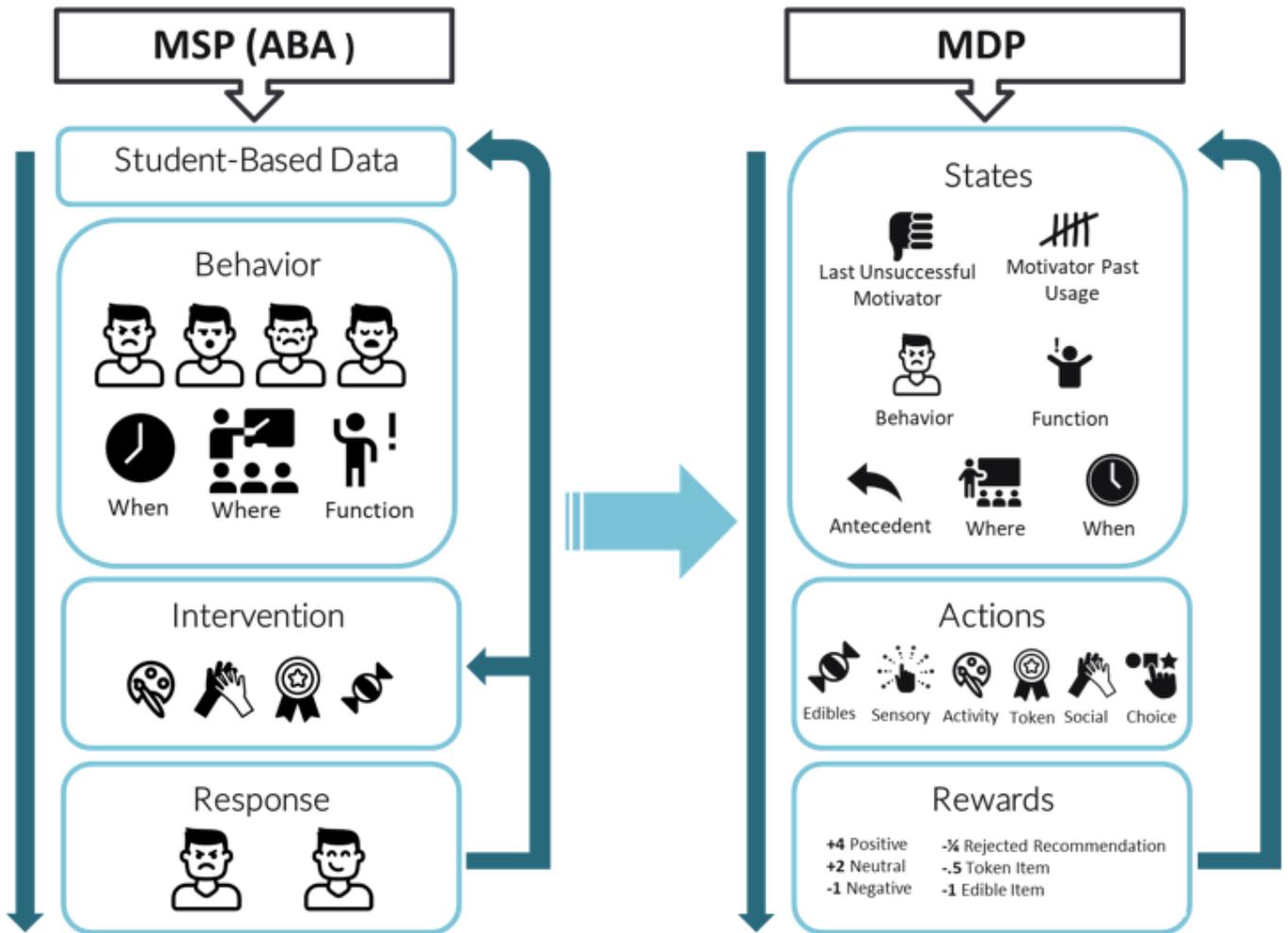


Figure 1

Behaviour Intervention Strategy

The screenshot shows a mobile application interface with a dark purple header containing a menu icon, a brain icon with colorful rays, and a vertical ellipsis. Below the header, the main content area is light gray and titled "Add Objective Update" with a green checkmark icon. It features two dropdown menus for "Date" and "Time", a "Duration" input field, and a toggle for "Objective Achieved" set to "0 %". A tabbed interface at the bottom of this section has "Add note" and "Behavior monitoring" tabs, with "Behavior monitoring" being the active tab. The "Behavior monitoring" section includes a "Behavior:" dropdown menu with "Throwing things" selected, a "Frequency:" input field, a "Motivators used:" input field, and an "Effective Motivator:" toggle. A large purple plus sign icon is centered at the bottom of this section.

**Figure 2**

Behaviour monitoring without the “Motivator Selection” feature

Motivator Suggestion

Behavior\*  
 Disruption (interrupting, yelling, knocking things over, ▾)

Antecedent\*  
 Given direction or cor ▾

Behavior Function\*  
 Escape ▾

**Suggest Motivator**

---

Session History

Obsession	Sensory	Positive
Obsession	Sensory	Positive
Obsession	Edibles	Rejected
Obsession	Edibles	Rejected
Obsession	Social	Neutral
Elopement	Sensory	Neutral
Tantrums	Sensory	Neutral
Stereotypy	Activities	Neutral
Obsession	Activities	Positive
Obsession	Tokens	Rejected
Non-compliance	Activities	Positive
Self-injury	Tokens	Neutral
Obsession	Edibles	Neutral
Obsession	Social	Neutral
Obsession	Choice	Rejected

Motivator Suggestion

Use "Activities" motivator.

Such as coloring, playing with a friend, and going to the play area.

ACCEPT      DECLINE

Motivator Suggestion

(Please Choose Student Response)

Behavior  
 Disruption (interrupting, yelling, knocking things over, ▾)

Antecedent  
 Given direction or cor ▾

Behavior Function  
 Escape ▾

Motivator Used  
 Activities

Student Response\*  
 Positive ▾

**Submit**

---

Session History

Disruption	Tokens	Rejected
Disruption	Social	Rejected
Obsession	Sensory	Positive
Obsession	Sensory	Positive
Obsession	Edibles	Rejected
Obsession	Edibles	Rejected
Obsession	Social	Neutral
Elopement	Sensory	Neutral

(a) Recording behaviour.

(b) Suggested motivator.

(c) Recording student response to motivator.

Figure 3

The "Motivator Selection" Feature

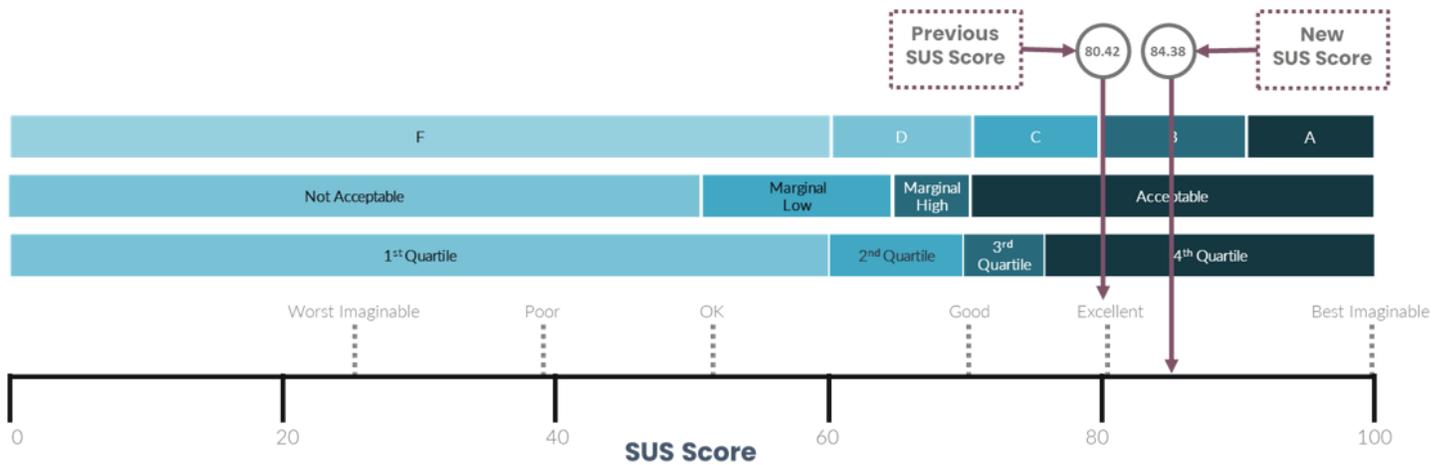
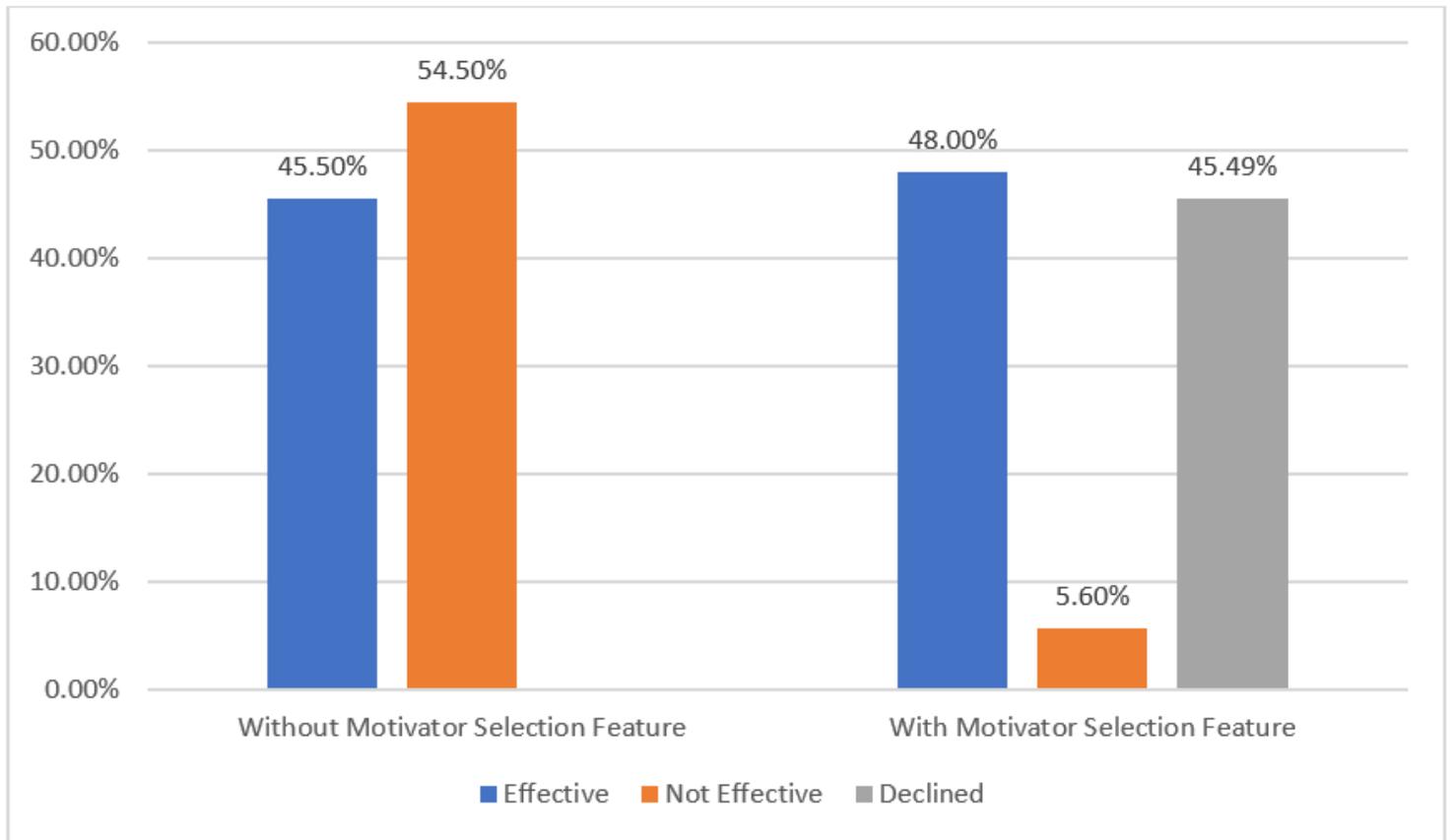


Figure 4

New SUS mean score compared to the SUS mean score of the previous iteration



**Figure 5**

Percentage of effectiveness of motivator with and without using the motivator selection feature

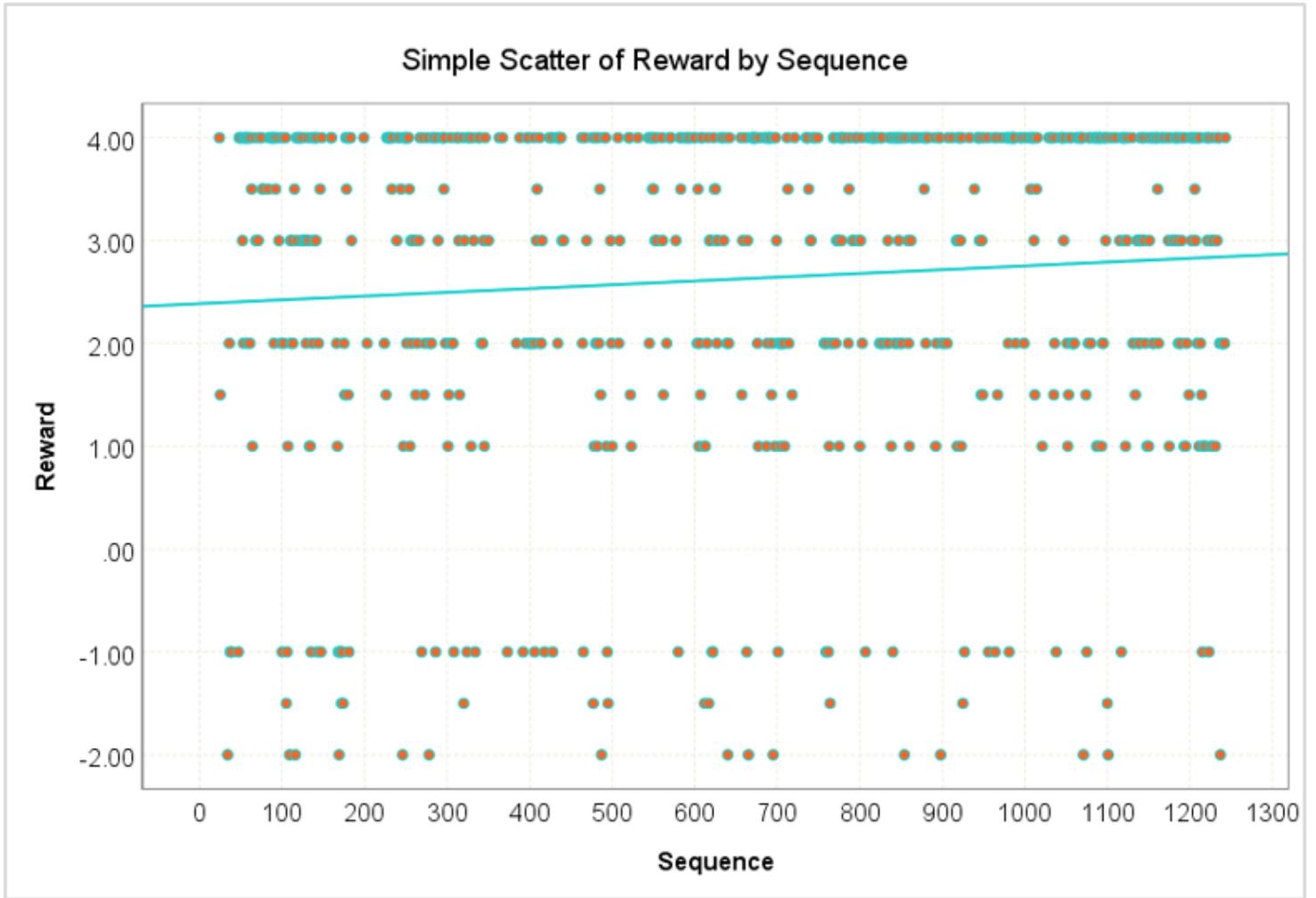


Figure 6

Scatterplot of Reward over Time

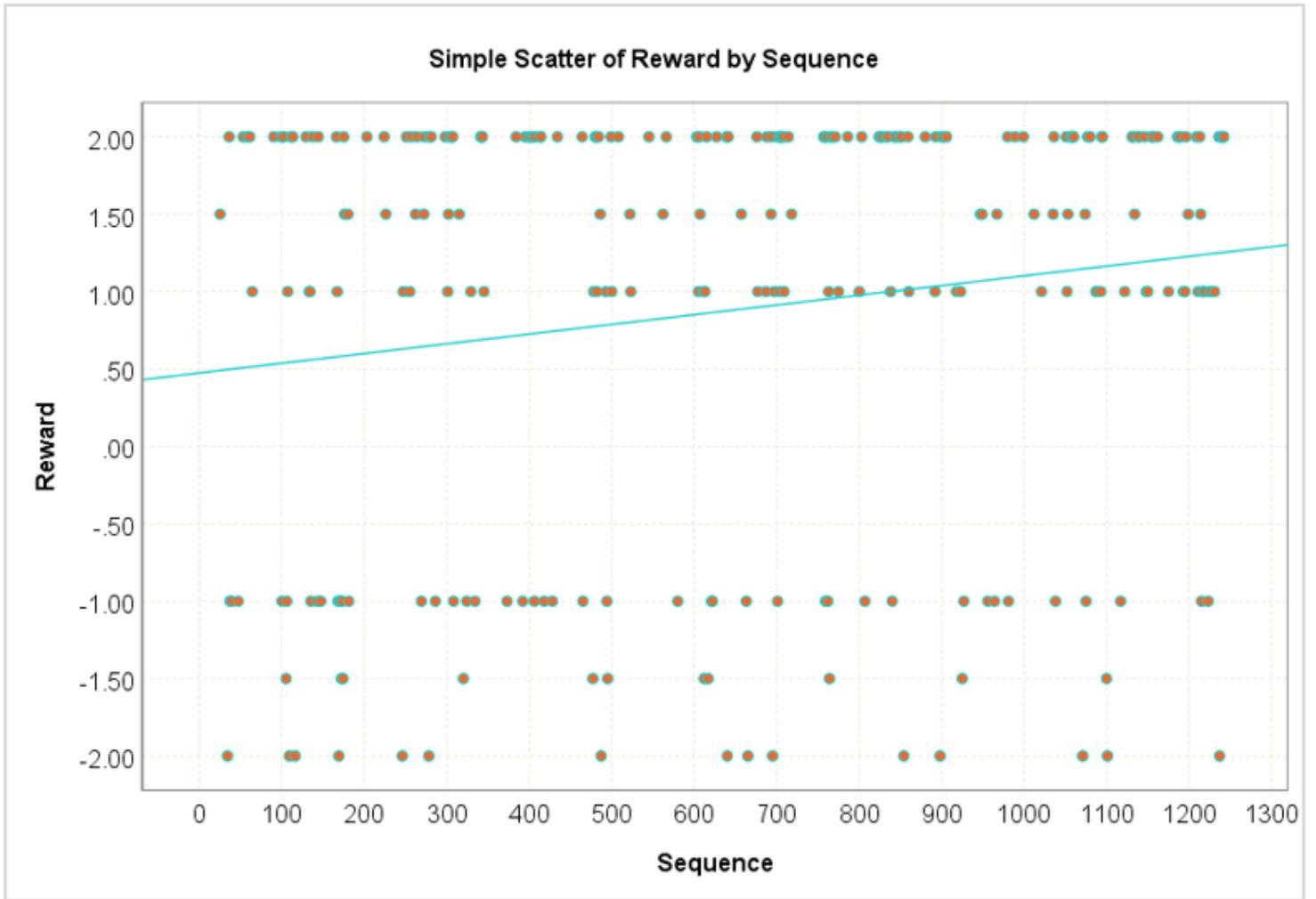


Figure 7

Scatterplot of Reward (< 3) over Time

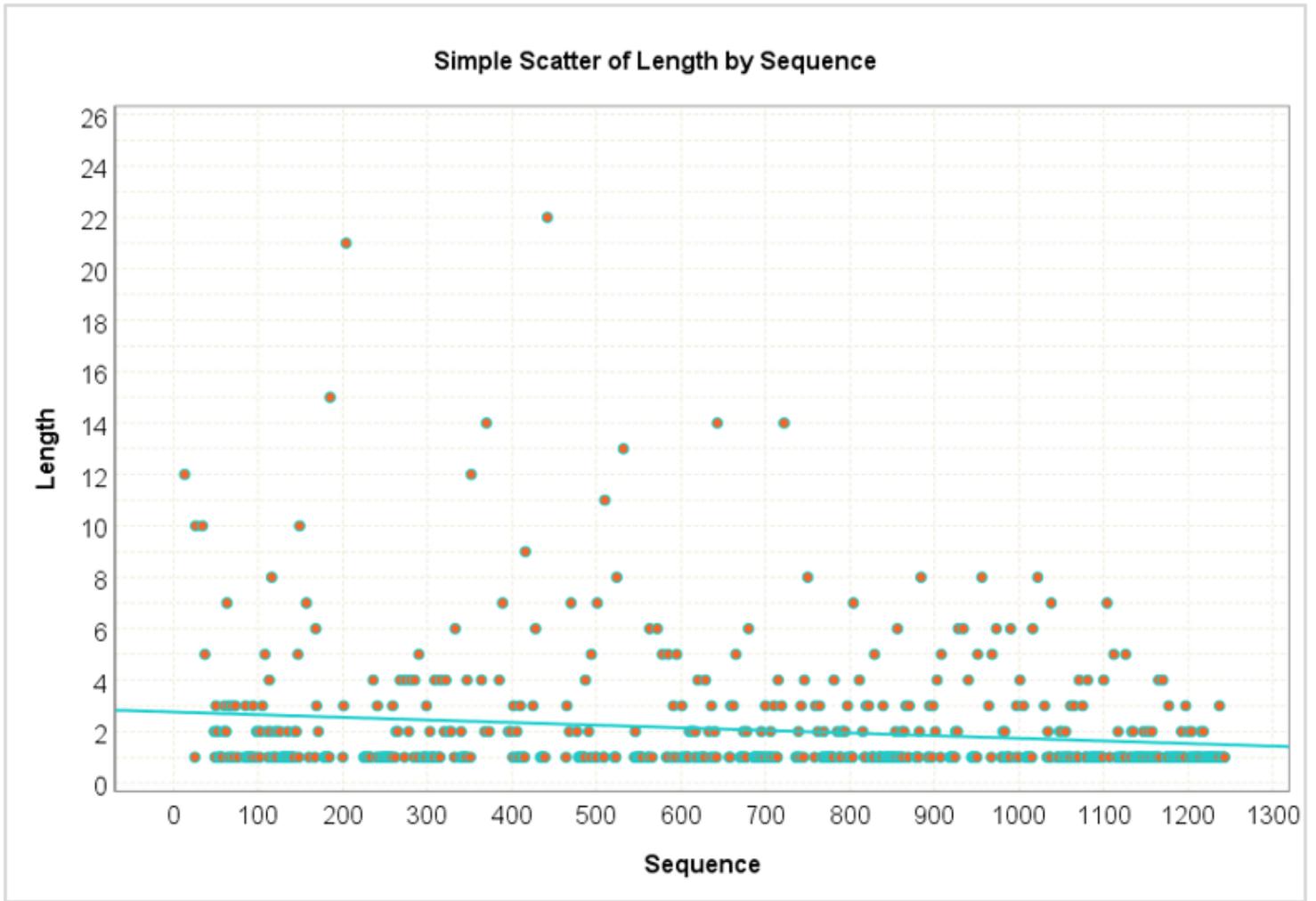


Figure 8

Scatterplot of Episode Length against Time

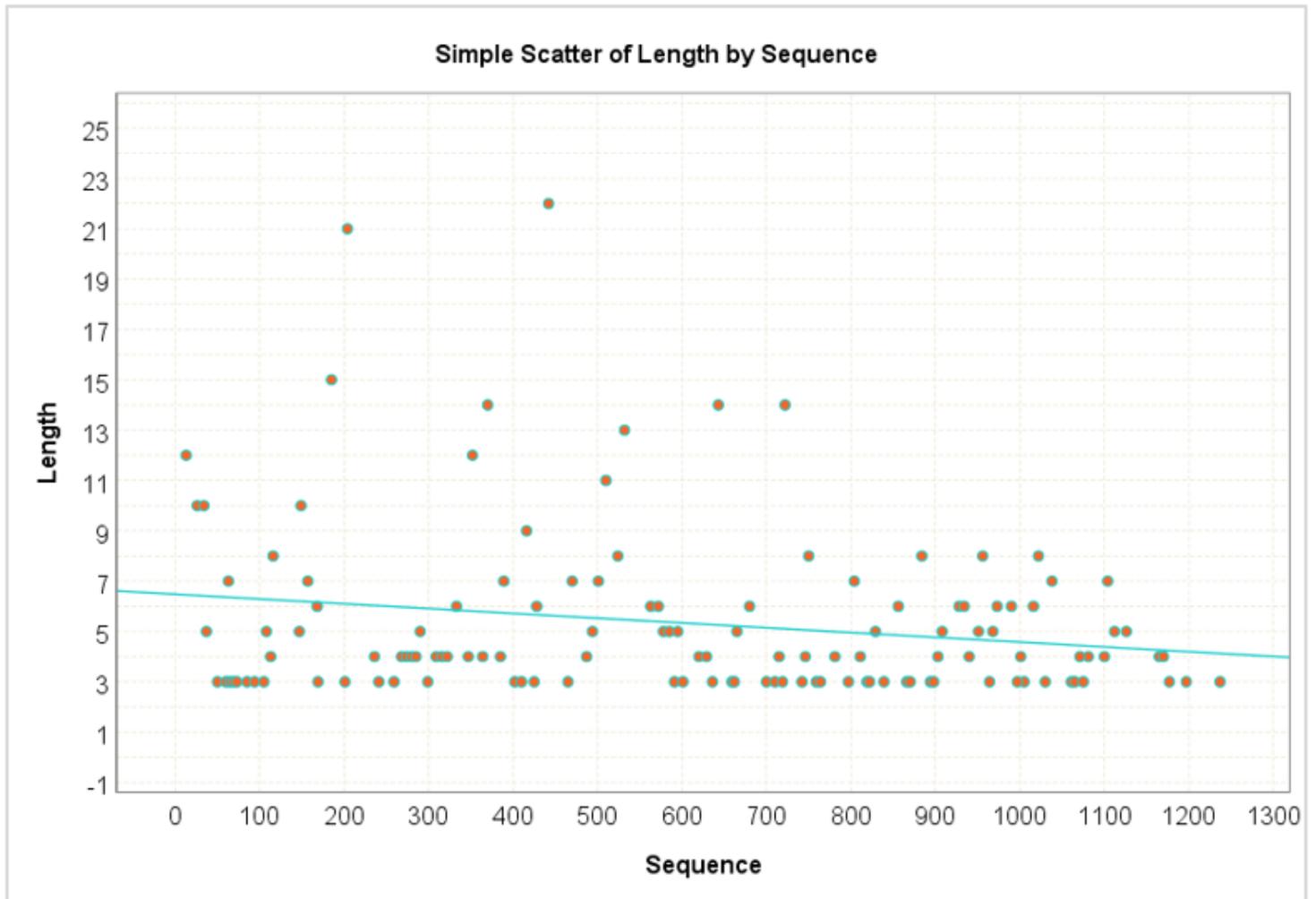


Figure 9

Scatterplot of Episode Length (> 2) against Time

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AppendixA.docx](#)