Automatic assessment of cognitive and emotional states in virtual reality-based flexibility training for four adolescents with autism

Jewoong Moon, Fengfeng Ke and Zlatko Sokolikj

Abstract
Tracking students’ learning states to provide tailored learner support is a critical element of an adaptive learning system. This study explores how an automatic assessment is capable of tracking learners’ cognitive and emotional states during virtual reality (VR)-based representational-flexibility training. This VR-based training program aims to promote the flexibility of adolescents with autism spectrum disorder (ASD) in interpreting, selecting and creating multimodal representations during STEM-related design problem solving. For the automatic assessment, we used both natural language processing (NLP) and machine-learning techniques to develop a multi-label classification model. We then trained the model with the data from a total of audio- and video-recorded 66 training sessions of four adolescents with ASD. To validate the model, we implemented both k-fold cross-validations and the manual evaluations by expert reviewers. The study finding suggests the feasibility of implementing the NLP and machine-learning driven automatic assessment to track and assess the cognitive and emotional states of individuals with ASD during VR-based flexibility training. The study finding also denotes the importance and viability of providing adaptive supports to maintain learners’ cognitive and affective engagement in a highly interactive digital learning environment.

Introduction
Autism spectrum disorder (ASD) is the pervasive mental disorder that accompanies low representational flexibility (Totsika, Hastings, Emerson, Lancaster, & Berridge, 2011). Representational flexibility indicates individuals’ mental ability to choose and use accurate representations to articulate and resolve tasks. Several studies stated that representational flexibility is an intellectual skill associated with individuals’ generic problem-solving skills (Deliyianni, Gagatsis, Elia, & Panaoura, 2016). Adolescents with ASD are underrepresented in science, technology, engineering and math (STEM) education (Hawley, Cardoso, & McMahon, 2013). Low representational flexibility of adolescents...
with ASD mainly explains their challenge in academic performance on STEM education. In particular, STEM subjects require learners to select, apply and perform multiple representations for design problem solving. Hence, using appropriate representations to design, develop and evaluate products are vital. However, low representational flexibility of adolescents with ASD limits their uses of multiple representations to rehearse and simulate their design ideas. Also, adolescents with ASD are likely to have difficulty in identifying and adapting design features in given varying circumstances.

In autism research, researchers have introduced virtual reality (VR)-based training to enhance students’ intellectual skills (Parsons & Mitchell, 2002). Using VR can benefit individuals with ASD to experience 3D simulations and manipulate their 3D-morphed design artifacts (Ke & Lee, 2016). Their hands-on manipulations of 3D products in an open-ended virtual environment can foster their design problem solving due to enriched multimodal representation experiences.

Despite emerging implementations of VR-based training for learners with ASD, how to assess cognitive and emotional states during such training remains an open question. Tracking the learning states to provide tailored learner support is a critical element of a VR-based highly interactive learning environment, especially for learners with ASD who have difficulty adjusting to different representations or varying circumstances. Promptly tracking individual learners’ cognitive and emotional states is a prerequisite for the provision of adaptive support. However, empirical explorations are still lacking to illustrate how the data-driven, automatic assessment can be contextually designed and implemented for learners with special needs or in the VR-based learning setting.

**Literature review**

**Representational flexibility in autism research**

Hayne (2006) defined representational flexibility as the ability to readily switch different modes of expression to create and reproduce knowledge. Representational flexibility also indicates
individuals’ capabilities to choose appropriate representations to account for a task and its features (Nistal, Van Dooren, Clarebout, Elen, & Verschaffel, 2009). Representational flexibility is associated with human memory representations between encoding and retrieving information (Jones, Pascalis, Eacott, & Herbert, 2011). Jones and Herbert (2008) explained that individuals’ practices in articulating conceptual knowledge can show their representational flexibility. They reported that simulating mental images with verbal articulations fosters learners’ practices of using multiple representations.

Prior research demonstrated the deficit of representational flexibility of individuals with ASD (Geurts, Corbett, & Solomon, 2009). Research on executive-functioning (EF) indicated that individuals with ASD have low EF skills, including a set of cognitive attributes—such as flexibility, inhibition, working memory (Chevalier et al., 2012; Sinzig, Morsch, Bruning, Schmidt, & Lehmkuhl, 2008). Researchers suggested that low flexibility is associated with the challenges of performing certain cognitive tasks. Specifically, given the cognitive tasks that require swift perspective-takings, individuals with ASD have low flexibility and they are likely to fail to switch attention from a prior task to a new one (Reed & McCarthy, 2012). In addition, research reported that individuals with ASD are likely to experience difficulty in performing task reconfiguration and adaptations in response to various environmental surroundings. Green et al. (2006) noted that the low flexibility of individuals with ASD features over-arousal and an inability to deal with unpredictable events. Ravet (2013) described low flexibility portrayed by certain individuals with ASD as: (1) rigid and repetitive behaviors, (2) tendency to focus on details than wholes, (3) difficulties in transfer and generalization, (4) problems with understanding abstract concepts and (5) lack of curiosity and motivation.

Correspondingly, prior research suggests that adolescents with ASD may experience the difficulty of performing analytical thinking in STEM problem-solving tasks. STEM fields generally require learners to conceptualize their ideas, implement iterative evaluations and revise them to produce their artifacts (Wei, Jennifer, Shattuck, McCracken, & Blackorby, 2013). Also, learners in the STEM tasks are asked to keep a holistic view of how multiple variables are interconnected in their problem solving. Learners, therefore, need to enhance their representational-flexibility skills to swiftly identify, compare, analyze the data and consequences during STEM problem solving. Rigid or inflexible thinking and poor representation skills will create challenges for performers during STEM problem solving.

Recent research has demonstrated the potential of using virtual reality (VR) in STEM education to support inquiry and problem-based learning. For example, researchers have studied VR-based teaching of anatomy with medical students (Birt, Stromberga, Cowling, & Moro, 2018; Moro, Stromberga, & Stirling, 2017). They found that the VR lessons yielded a high level of enjoyment and perceived usefulness. Pirker Holly Lesjak Kopf and Gütl (2019) implemented a VR-based virtual laboratory of physics. This VR environment enabled students to implement multiple physics experiments (eg, falling coil experiment and Van de Graaff Generator). Furthermore, Davis, Phillips and Kulm (2018) introduced a VR-based music-mathematics learning program that enables learners to play and create songs to acquire an understanding of the underlying mathematical concepts (eg, fractions, ratios and patterns).

**Training interventions to enhance representational flexibility**

Recent research raised questions of how to embrace and support learners with ASD for their future college enrollment and future career development in STEM fields (Batten, 2005). Research has investigated how to enhance autistic learners’ flexibility-related skills by using purposeful training. In this sense, previous studies introduced training interventions to help individuals
with ASD to promote representational flexibility (Cotugno, 2009; de Vries, Prins, Schmand, & Geurts, 2015). Kenworthy et al. (2014) implemented Unstuck and On Target (UOT), which was an executive-functioning intervention for children with ASD. This school-based intervention delivered instructional modules to specifically promote learners’ flexibility skills—such as big-picture thinking and planning. This intervention provided learners with open-ended activities, such as game design, roleplaying and gameplay.

In a recent decade, autism researchers have used virtual reality (VR) to enhance representational flexibility for individuals with ASD. Researchers suggested that open-ended and 3D visual stimuli enable learners to enhance representation skills by incorporating new information with different modalities (Ritter et al., 2012). Parsons and Mitchell (2002) stated that computer-based representations of tasks in VR can enhance the flexibility of learners with ASD. VR offers hands-on design practices that enable learners with ASD to plan, create and evaluate their artifacts in the virtual world. By participating in design activities, individuals with ASD can become skillful in multitasking and planning. Also, the ill-structured and multifaceted nature of design problems in VR-based training allows learners with ASD to perform and master attention-switching behaviors. For example, Ke and Lee (2016) implemented Opensimulator-supported VR social-skills training for children with high-functioning autism (HFA). HFA children joined 3D design problem-solving activities in a VR environment. The result of the study suggested that HFA children improved their flexibility by designing and creating 3D artifacts.

Although the previous autism research proposed VR-based training as a potential avenue to promote representational flexibility, empirical explorations are still lacking. Despite emerging VR-based interventions for learners with ASD, existing research appeared limited to inform on how to facilitate their problem solving in VR-based training systematically. Despite the enriched and immersive visual stimuli in VR, the complexity of 3D design tasks in VR may invoke cognitive challenges and hence additional learning supports are necessary. To offer timely learning supports for learners with ASD, identifying their cognitive and emotional states during the training is essential. As such, developing an automatic learning-state assessment mechanism is vital to a VR-based representational flexibility training.

Automatic assessment

Autism research has implemented various measures to assess flexibility improvements of individuals with ASD. Prior studies implemented external EF tasks, such as the Wisconsin card sorting task (WCST) (Kongs, Thompson, Iverson, & Heaton, 2000) and the Stroop test (Stroop, 1935). These assessment approaches showed several limitations in assessing the representational flexibility of individuals with ASD. First, prior research mostly used external measures but failed to track students’ learning state or behavioral data (Marcu et al., 2013). Those measures could not track the situational actions of individuals with ASD—intertwined with their given contexts. Second, existing assessments could not deliver timely and in situ feedback because assessing representational flexibility is complicated and time-consuming—implying that it may be too late to offer effective learning support during the training.

Prior research on ubiquitous computing has indicated a positive role of automatic assessments in promptly detecting how, why and when individuals with special and diverse needs will experience intellectual challenges during an educational intervention (Plötz et al., 2012). Consistently, recent research on cognitive-affective computing also highlights the significance of a data-driven automated assessment that captures students’ cognitive and emotional states in computerized learning environments (Dessi et al, 2019; Grawemeyer et al, 2017). Such an automated real-time assessment helps to identify students’ perceived level of challenges to inform when and how
formative feedback should be presented in computer-based learning. Concerning the diverse and complicated behavior states of individuals with ASD, using an automatic, data-driven assessment is especially important. For example, Plötz et al. (2012) developed an automatic assessment system that tracks the problem behaviors of individuals with ASD (e.g., aggression, disruption and self-injury). Using an attached on-body behavior-tracker, the system estimated the problem behaviors of individuals with ASD. Via machine-learning techniques, this assessment system segregated each behavioral episode from the participants’ sensory data to compute the frequency of target behaviors across intervention periods.

Recent research specifically sought ways to investigate students’ speeches to uncover their underlying mental states (Beggrow, Ha, Nehm, Pearl, & Boone, 2014) as part of an automatic assessment. Research believes that language and spoken discourses externalize individuals’ thoughts and represents the dynamics of their mental states (McNamara, Allen, Crossley, Dascalu, & Perret, 2017). Several empirical studies used natural-language-processing (NLP) techniques to examine students’ speech data to interpret their mental states. For instance, Maenner et al. (2016) implemented a machine-learning algorithm to diagnose ASD surveillance of children. Using children’s words and phrases as model-training data and the classification approach of random forest (optimal model accuracy = .409) compared the automatic ASD diagnosis results with those of clinicians’ evaluation. Nakai et al. (2017) designed and implemented an automatic detector that captures the abnormal word utterances from children with autism via machine-learning techniques. This study sampled a total of 1,026 single-word utterances from the study participants and implemented support vector machine (SVM) to yield automatic classification results. As another example, Cho et al. (2019) designed a vocal-features-driven classification model to diagnose ASD from individuals’ informal conversations. They distilled various text features—such as frequencies of total words and pronoun usages—and implemented human coding to prepare the training data.

Although these recent studies on automatic assessment have used different machine-learning techniques, they did not aim to examine students’ skill improvement in a training program but mostly focused on diagnosing the mental illness and abnormal behaviors of individuals with ASD. Moreover, although VR-based training should assess learning dynamically to be adaptive with various learners with ASD, how to design and implement automatic assessments to swiftly capture learners’ cognitive and emotional states remains an open question.

Research purpose
This study explores the design and implementation of the automatic assessment to track both cognitive (representational flexibility) and emotional (six basic emotions) states of individuals with ASD via the speech data during VR-based flexibility training. We employed both NLP and machine-learning techniques for speech data mining. The primary research question of this study is: How does speech data mining enable the tracking of cognitive and emotional states of adolescents with ASD during their VR-based flexibility training?

Method
Participants
We sampled four male adolescents with ASD living in Southeastern US. The ages of participants 1–4 were 19, 13, 16 and 16 respectively. All participants had self-reported and documented (with Gilliam Autism Rating Scale, GARS-3) diagnosis of autism, can speak, read and write, and received grade level or above academic instruction. The study received institutional review board (IRB) approval and the written consent from all study participants and their guardians.
To develop the current assessment mechanism, we used a random sample of the recorded study session data (including baseline and intervention sessions) as the testing data set. The number of sampled study sessions by each participant ranged from 8 to 23. All adolescents with ASD joined in the program have prior 3D gameplay experiences.

**VR-based flexibility training**

We created a VR-based flexibility training program using *Opensimulator*, an open-source VR platform that enables learners with ASD to perform multiple information representations and multimodal (text, voice and nonverbal) interactions. Each study participant joined multiple rounds of intervention sessions over 8 to 16 weeks. During the intervention program, participants were asked to perform scientific design problem solving in a set of authentic scenarios. These 3D-simulation design problems aim to foster the development of domain-specific representational competence—flexibility in representing and solving the science- and math-related problems of forces and Newton’s laws of motion. For example, they were requested to design and construct a 3D elevation bridge in a virtual village. They needed to identify the villagers’ needs and particular design criteria for the bridge construction. The participants came up with their designs using 3D blocks in the virtual world or drawing sketches on a virtual white board. Also, to simulate the elevating motions of the bridge, they needed to code and apply scripts to the bridge. The program embedded a set of entry-level scripting modules that trained the participants (who were novice in scripting) on both pseudo coding and coding using a 3D block-based visual programming plugin. To correctly edit the magnitude of the forces underlying the elevation-bridge motions, they had to apply the knowledge of Newtonian physics. We aligned the domain competencies framed by the design tasks with middle-school physics and math content and practice standards (ie, Next Generation Science Standard, Common Core Math Standard). Figure 1 demonstrates how the participants of this study performed their design problem-solving tasks in VR.

To scaffold the participants’ design problem solving, two facilitators joined to provide contextualized cuing and prompting. They also puppeteered multiple task-related social characters, such as the village council board members and villagers, to present background task information and design clues. Both facilitators were graduate students of education major and followed a structured character-puppeteering protocol during the intervention sessions. Asides from character- or agent-based scaffolding, we provided direct instruction on the math, science and programming knowledge involved in the tasks with interactive tutorials presented on the virtual media boards.
Automatic assessment development

As shown in Figure 2, this study implemented three assessment-development phases: (1) model training, (2) model testing and (3) model validation. First, in model training, we prepared a training dataset that represents the verbal examples of representational flexibility. The training data were prepared by synthesizing the results of both a systematic literature review and the speech data of the other children with ASD who joined the pilot session. We then extracted the lexical features of the training data that allows candidate classifiers to detect representational flexibility growth. Second, in model testing, we collected four participants’ written and oral speech data from their VR-based training sessions and converted all data to scripts. We segregated the participants’ verbal utterances with multiple data instances. We ran a multi-label classification approach with candidate classifiers to test these classifiers’ performance. Third, for model validation, with k-fold cross-validation, we used two predictive performance indicators to evaluate the candidate classifiers.

Model training

Operationalization of classification labels

We operationalized cognitive and emotional states of the study participants when designing our automatic assessment. We considered the participants’ enactments of representational flexibility as the cognitive states of this study. The target competencies of representational flexibility in this study are (1) attention-switching, (2) multimodal representation, (3) pattern development and (4) pattern contextualization. Attention-switching indicates learners’ ability to switch attention or action based on changed rules and contextual demands (Reed & McCarthy, 2012). With regard to representational flexibility, this competency is associated with learners’ ability to identify how well they understand contextual demands to choose appropriate representations. Multimodal representation indicates learners’ capability to identify and apply multiple representations during design problem analysis and solving (Linebarger & Norton-Meier, 2016). This competency is associated with learners’ ability to create different representation formats in terms of representational flexibility. Pattern development refers to the ability to identify or delineate a pattern (or rule) during design problem analysis and solving (MacDonald, Westenskow, Moyer-Packenham, & Child, 2018). Relevant to representational flexibility, this competency specifically describes learners’ ability to synthesize and explicate hidden patterns of existing representations. Lastly, pattern...
contextualization indicates learners’ ability to identify an implementation context of the endorsed pattern or customize the pattern based on a novel implementation context (Yee & Bostic, 2014). Associated with representational flexibility, this competency explains learners’ capability to come up with a novel representation based on existing representation contexts.

In terms of the emotional states during the training, we labeled six primary emotions based on Ekman’s (1970) emotion categories: anger, disgust, fear, joy, sadness and surprise. Table 1 shows the details of the cognitive (target competencies) and emotional (six primary emotions) states.

Training data preparation
To track the states of adolescents with ASD during the training, a classification model of both cognitive (enactment of representational flexibility) and emotional state (six emotions) is essential. To build a classification model for each state type, we needed to prepare a training dataset. Since there is no validated training dataset—particularly indicating the target cognitive state of this study, the initial development of the training data is needed. On the contrary, for tracking the participants’ emotional states, we did not prepare the training data. We instead ran the existing emotion-classification model designed by Colneri et al. (2018). This model already contains the enough amount of the training dataset (1,175,847 dataset) and showed good performance to detect the target emotions of this study (accuracy training data = above 90%; accuracy validation data = above 70%). To achieve more precise results on the emotional states, we chose the existing emotion-classification model with good data-validation results.

To prepare the training dataset for the target cognitive state, we collected a selection of verbal examples that demonstrated the occurrences of representational flexibility. There were two steps of preparing verbal examples for the training dataset. First, we conducted a systematic literature review of qualitative research that particularly contained individuals’ exemplary discourses related to representational flexibility. Second, we ran a total of 20 one-hour pilot training sessions for three children with ASD, who were not participants of this study. The activities of the pilot training sessions were similar to the design tasks in the training program of the current study. We video- and screen-captured all pilot training sessions and then, sampled the participants’ speech data. We used the text-to-speech converting software Otter.ai that transcribed the participant’s speeches to texts (Otter.ai, 2019). We then synthesized sampled verbal records from both the literature review and pilot training sessions. We finally prepared a total of 106 text data instances as a training dataset of the classification model.

The classification labels of the cognitive state in this training dataset were neutral, positive and negative with every of the four representational-flexibility competencies. The positive cases indicate a good example of representational flexibility whereas negative cases refer to poor behaviors on representational flexibility. If a record does not include any clear evidence of either positive or negative evidence, we counted it as neutral. For example, for the competency of Attention-switching, a positive case is the act of stating multiple design clues associated with a design quest. A positive case of the competency Multimodal representation is the portrayal of explaining or prototyping one’s design idea via different representation formats, such as 3D blocks and 2D visuals. A positive case associated with the competency pattern development is the expression related to the analysis of the relationship among various features of a design artifact. Lastly, for the competency pattern contextualization, a positive case shows that an individual either implements or customizes a design pattern or solution across multiple given design contexts. In comparison, the negative instances of the aforementioned competencies were failed or unsuccessful attempts of the aforementioned expressions or enactments. Multiple human coders with experts iteratively reviewed and revised the initial classification labels of the training data. Examples of the training dataset are provided in the Appendix.
<table>
<thead>
<tr>
<th>Type</th>
<th>Target competency</th>
<th>Definition</th>
<th>Literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cognitive state (representational flexibility)</td>
<td>Attention switching</td>
<td>Ability to switch attention or action based on changed rules and contextual demands</td>
<td>Reed and McCarthy (2012)</td>
</tr>
<tr>
<td></td>
<td>Multimodal representation</td>
<td>Capability to identify and apply multiple representations</td>
<td>Linebarger and Norton-Meier (2016)</td>
</tr>
<tr>
<td></td>
<td>Pattern development</td>
<td>Ability to identify or delineate a pattern (or rule) during design problem analysis and solving</td>
<td>MacDonald, Westenskow, Moyer-Packenham, and Child (2018)</td>
</tr>
<tr>
<td></td>
<td>Pattern contextualization</td>
<td>Ability to identify an implementation context of the endorsed pattern or customize the pattern based on a novel implementation context</td>
<td>Yee and Bostic (2014)</td>
</tr>
<tr>
<td>Emotional state</td>
<td>Angry</td>
<td>Affective states when individuals tend to escape from things that are bothering</td>
<td>Ekman (1970)</td>
</tr>
<tr>
<td></td>
<td>Fear</td>
<td>Emotional responses to an anticipated threat or thoughts about potential risks</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Disgust</td>
<td>Unpleasant feelings from surroundings</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Sadness</td>
<td>Feeling of disappointment, grief and disinterest</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Surprise</td>
<td>Verbal reactions to show startle responses following something unexpected</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Joy</td>
<td>Pleasant emotional states that are featured by contentment and joy</td>
<td></td>
</tr>
</tbody>
</table>
Feature extraction
We extracted defining feature of the positive, neutral and negative cases from the training dataset (106 text data instances). The goal of the feature extraction was to distill key lexical characteristics of the training dataset that can teach the initial classification model to detect meaningful verbal utterances of target competencies. We used the software LightSIDE, a natural-language-processing (NLP)-driven workbench that implemented both text pre-processing and feature extraction (Mayfield & Rosé, 2013). The software allows researchers to select and conduct machine-learning algorithms to choose the best classification model to fit given linguistic contexts. We chose the bag-of-word model that decomposed a structure of the raw text and then encoded it as binary codes so that a computer can read and understand a collection of the text features from the training data. We selected the following feature extraction conditions—unigram, bigram, trigram, line length, POS bigram, POS Trigram, Line length, word/POS pair, normalizing N-Gram counts, stem N-Gram and skip stop-words in N-Gram. In the end, we extracted a total of 994 textual features from the training dataset.

Model testing
Data collection and segregation of the testing data
To test the classification model emerging from the training data, we collected both video- and screen-recorded intervention sessions of the four participants. Using a total of 66 one-hour recorded intervention sessions, we sampled a total of 585 text instances from the participants’ speech data, served as the testing data set of this study. To make the data instances contain meaningful discourses from the raw text data, the segregations of speech data were important (McNamara et al., 2017). In this study, we segregated the participants’ speech data ranged from 5–10 minutes considering natural periods of discourses. The average word count of a single data instance is 71.79.

Model implementation
To track the cognitive and emotional states of the study participants, we implemented a supervised multi-label classification approach (Katakis, Tsoumakas, & Vlahavas, 2008). Multi-label classification designates labels of data instances into either one or more than two classes. This approach aimed to resolve the text categorization problem, in which each data instance belongs to multiple types of classification classes. The multi-label classification approach considers the data as not reciprocally exclusive. In comparison to other classification techniques, multi-label classification is useful to detect the data instance that may be associated with multiple competencies.

Under the multi-label classification, we adopted multiple approaches when conducting classifications. To track the participants’ cognitive states, we performed dichotomous classifications to count the frequency of each label (neutral, positive and negative) across the four representational-flexibility competencies. To track the participants’ emotional states, we ran the emotion analyzer Tweet Profiler (Colnerič & Demsar, 2018), supported by the machine-learning software Orange (Demšar et al., 2013). This emotion analyzer was trained by a deep-learning technique (latent semantic indexing model). To estimate the emotional states, we used probabilistic classifications because individuals’ emotional states are hardly classified by only using speech data. Therefore, in this study we conducted the probabilistic estimations to precisely detect the likelihood of emotion occurrences. Although this study focuses on detecting the enactment of representational flexibility in terms of the cognitive state, we also aimed to estimate participants’ general emotional states when they interacted with the training program. We adopted a relatively reliable classification model that contains multimodal training data for emotion detection and enhances generalizability through deep-learning driven techniques.
We used the LightSIDE to select multiple candidate classifiers with the best performance. We used several techniques or approaches of classifiers: (1) Support vector machine, (2) Naïve Bayes, (3) Decision tree and (4) Logistic regression. A support vector machine is a classifier to divide or map the data into the vector space via a separating hyperlane. Given the training dataset with labels, the support vector machine estimates the output with an optimal hyperlane. For example, if there are two label classes on the graph as a vector space, a support vector machine draws a hyperplane that classifies the labels depending on the location of the datasets. Naïve Bayes is another classifier that uses probabilistic computations by using Bayes theorem. The Bayes theorem estimates the probability of an event occurrence given the likelihood of another event that has already happened. A decision tree is a classification algorithm that shows a hierarchically connected visual model of decisions with possible consequences. Hierarchical connections among nodes and branches demonstrate an entire classification rule. Last, logistic regression refers to the probabilistic model to determine dichotomous classes or events. Logistic regression uses the marginal probabilities of the training data and then, those probabilities are better calibrated compared to other classifiers.

Model validation

We evaluated the predictive performance of multiple candidate classifiers using two metrics: accuracy and Cohen’s kappa. Accuracy is the percentage of correctly classified data instances from the entire dataset. For example, if a classification model shows the accuracy score 0.6, this result indicates that the model performs well at 60%. Cohen’s kappa is another indicator to evaluate the classification model performance—specifically examining whether the result is computed by chances. This statistic compares the accuracy of the classification model with the accuracy of random computations. A kappa value ranges from 0 to 1. According to Fleiss’s kappa, the range of the kappa coefficient indicates excellent (0.75 < X ≤ 1), fair to good (0.40 < X ≤ 0.75) and poor (0 < X ≤ 0.40) performance respectively (Hartling et al., 2012).

To evaluate the selected candidate classifiers, we used the fivefold cross-validation. K-fold cross-validation is a resampling technique that evaluates the performance of a classifier on new data (Kim, 2009). In this technique, a parameter K refers to the number of groups that a given data sample should be split. In other words, if using fivefold cross-validation, it indicates that the algorithm will run five times—each time fourfold used for training and onefold is used for testing the model.

Results

Model validation results

We implemented the model validation through two phases: K-fold cross-validation and manual evaluations by experts. Implementing the candidate classifiers is to detect the representational flexibility and emotional states of adolescents with ASD and the model validation is to evaluate the performance of these candidate classifiers. First, using the fivefold cross-validation, we evaluated how well the candidate classifiers tracked participants’ enactment of representational flexibility. Table 2 shows the overall result of the performance metrics from the four classifier types (ie, Support vector machine, Naïve Bayes, Decision tree and Logistic regression). According to the results of performance metrics (accuracy and Cohen’s kappa), the classifiers by both support vector machine (accuracy = .674avg) and logistic regression (accuracy = .660avg) appeared best when tracking the competencies of representational flexibility. The kappa scores of both approaches of classifiers were also satisfactory (support vector machine: .448avg, logistic regression: .413avg), which indicated that the classifiers’ performance was reasonable (fair to good: 0.40 < X ≤ 0.75).
Second, we also conducted a two-round manual evaluation of the automatic assessment result through expert reviews to confirm the accuracy of the candidate classifiers. At the first-round evaluation, a trained coder conducted the manual evaluation of the classification results. The trained coder attended all training sessions and was trained based on a VR-based cognitive training intervention protocol with the descriptions of the target competencies on representational flexibility. In the second-round evaluation, two expert reviewers joined to mutually evaluate the results of the first-round evaluation. They reviewed the evaluation results and discussed their discrepancies until they meet 100% agreement with each other.

**Multi-label classification results**

The time-series graph in Figure 3 demonstrates the results of the multi-label classifications—showing both cognitive (bar graph) and emotional (line graph) states of the study participants. The left axis of the graph indicates the frequency level of the cognitive states, whereas the right axis of the graph shows the probability level of the emotional states. The multi-label classification results in Figure 3 suggested that the four study participants demonstrated various patterns of cognitive and emotional states during VR-based flexibility training. Specifically, participants 1 and 2 showed similar patterns of competency growth on representational flexibility and

---

**Table 2: Performance metrics results of candidate classifiers**

<table>
<thead>
<tr>
<th>Classifiers</th>
<th>Support vector machine</th>
<th>Naïve Bayes</th>
<th>Decision tree</th>
<th>Logistic regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attention-switching</td>
<td>.679 (.497)</td>
<td>.518 (.285)</td>
<td>.528 (.253)</td>
<td>.679 (.495)</td>
</tr>
<tr>
<td>Multimodal representation</td>
<td>.679 (.457)</td>
<td>.641 (.451)</td>
<td>.547 (.233)</td>
<td>.632 (.363)</td>
</tr>
<tr>
<td>Pattern development</td>
<td>.660 (.423)</td>
<td>.585 (.361)</td>
<td>.660 (.444)</td>
<td>.650 (.398)</td>
</tr>
<tr>
<td>Pattern contextualization</td>
<td>.679 (.413)</td>
<td>.547 (.242)</td>
<td>.490 (.098)</td>
<td>.679 (.397)</td>
</tr>
</tbody>
</table>

Note: Accuracy (kappa).
emotional changes, while participant 3 showed an obviously different pattern in both cognitive and emotional states. Participants 1 and 2 showed increases in positive cases of cognitive and emotional states during their training, whereas participant 3 mostly showed negative instances of the cognitive and emotional states. Compared to other participants, participant 3 had difficulty in identifying given resources of a virtual world and then, developing his 3D bridge model. Participant 3 also chose to use text messages rather than voice communication during the training. In comparison to other participants, he showed more difficulty in clearly articulating his design ideas and questions. Compared to participants 1 and 2, participant 4 showed fewer negative cases of the target competencies. The assessment results indicated that the participants showed different patterns of learning states when experiencing identical design problem-solving tasks.

We conducted a post hoc Pearson correlation analysis to explore how the participants’ cognitive (representational-flexibility competencies) and emotional states (six emotions) were associated with each other. Table 3 shows the correlation analysis results. The results indicated multiple statistically significant correlations between the participants’ negative cases of representational-flexibility competencies and their emotional states. Notably, negative occurrences of representational-flexibility competencies were positively correlated with multiple negative emotions, whereas the positive emotional state (Joy) negatively correlated with the negative occurrence of multimodal representation ($r = -.34, p < .01$) and pattern development ($r = -.26, p < .05$). These results indicated a general positive correlation between the detected cognitive and emotion states.

**Discussion and Conclusions**

*Design and implementation of an automatic assessment*

Overall, using both NLP and machine learning techniques, this study demonstrated the feasibility of using an automatic, data-driven assessment to track both cognitive and emotional states of adolescents with ASD during VR-based training. To implement the computerized, data-driven assessment during the VR-based cognitive training, we have captured participants’ verbal inputs and utterances in real time and promptly converted the study participants’ speech data to text data. After every training session, the speech data-driven assessment would estimate the current cognitive and emotional states of adolescents with ASD. In this study, we evaluated the accuracy of the candidate classifiers to confirm how well they detect learners’ representational flexibility changes. According to the performance metric results, we found that two approaches of classifiers (support vector machine and logistic regression) as promising methods to be implemented in the future assessments of representational flexibility.

The study finding suggests that the proposed speech data-driven assessment along with machine-learning classifiers can detect the cognitive and emotional states of learners with ASD during their VR-based, interactive learning activities. The study also helped to inform us about when and how individuals with ASD would experience different challenges during design-based scientific problem solving. More broadly, the study results will contribute to the scholarly discussions on learning assessment design in the cognitive-affective computing research (Dessi et al., 2019; Grawemeyer et al., 2017). Specifically, the study demonstrates the viability of using speech data mining to track both targeted competencies and emotional states of adolescents in a VR-based learning environment.

*Alerting learners’ needs of adaptive learning support in VR-based flexibility training*

Using a data-driven assessment, as this study illustrated, can provide in situ information governing when and how learning supports should be presented to individual learners with ASD. As illustrated by Figure 3, both the cognitive and emotional states of adolescents with ASD
### Table 3: Correlation analysis results

<table>
<thead>
<tr>
<th></th>
<th>AS_P</th>
<th>AS_N</th>
<th>MM_P</th>
<th>MM_N</th>
<th>PD_P</th>
<th>PD_N</th>
<th>PC_P</th>
<th>PC_N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Anger</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>−0.074</td>
<td>0.115</td>
<td>0.002</td>
<td>.316**</td>
<td>0.024</td>
<td>0.147</td>
<td>−0.099</td>
<td>0.060</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.557</td>
<td>0.356</td>
<td>0.987</td>
<td>0.010</td>
<td>0.849</td>
<td>0.240</td>
<td>0.429</td>
<td>0.634</td>
</tr>
<tr>
<td>N</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td><strong>Disgust</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>−0.191</td>
<td>.281*</td>
<td>0.005</td>
<td>.252*</td>
<td>−0.092</td>
<td>.295*</td>
<td>0.000</td>
<td>0.055</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.124</td>
<td>0.022</td>
<td>0.966</td>
<td>0.041</td>
<td>0.461</td>
<td>0.016</td>
<td>1.000</td>
<td>0.661</td>
</tr>
<tr>
<td>N</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td><strong>Fear</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>−0.170</td>
<td>0.147</td>
<td>−0.076</td>
<td>.367**</td>
<td>−0.187</td>
<td>.303*</td>
<td>−0.119</td>
<td>0.073</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.171</td>
<td>0.238</td>
<td>0.547</td>
<td>0.002</td>
<td>0.132</td>
<td>0.013</td>
<td>0.341</td>
<td>0.562</td>
</tr>
<tr>
<td>N</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td><strong>Joy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>0.106</td>
<td>−0.202</td>
<td>0.065</td>
<td>−.344**</td>
<td>0.109</td>
<td>−.263*</td>
<td>−0.108</td>
<td>−0.088</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.399</td>
<td>0.103</td>
<td>0.603</td>
<td>0.005</td>
<td>0.386</td>
<td>0.033</td>
<td>0.389</td>
<td>0.484</td>
</tr>
<tr>
<td>N</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td><strong>Sadness</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>−0.220</td>
<td>.340**</td>
<td>−0.186</td>
<td>.540**</td>
<td>−0.191</td>
<td>.420**</td>
<td>−0.080</td>
<td>0.195</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.076</td>
<td>0.005</td>
<td>0.135</td>
<td>0.000</td>
<td>0.123</td>
<td>0.000</td>
<td>0.521</td>
<td>0.117</td>
</tr>
<tr>
<td>N</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
</tr>
<tr>
<td><strong>Surprise</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson correlation</td>
<td>−0.234</td>
<td>.430**</td>
<td>−0.209</td>
<td>.532**</td>
<td>−0.186</td>
<td>.534**</td>
<td>−0.134</td>
<td>.344**</td>
</tr>
<tr>
<td>Sig. (2-tailed)</td>
<td>0.059</td>
<td>0.000</td>
<td>0.091</td>
<td>0.000</td>
<td>0.136</td>
<td>0.000</td>
<td>0.283</td>
<td>0.005</td>
</tr>
<tr>
<td>N</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
<td>66</td>
</tr>
</tbody>
</table>

**Note:** AS_P = positive case of attention switching, AS_N = negative case of attention switching, MR_P = positive case of multimodal representation, MR_N = negative case of multimodal representation, PD_P = positive case of pattern development, PD_N = negative case of pattern development, PC_P = positive case of pattern contextualization and PC_N = negative case of pattern contextualization. ** = Correlation is significant at the 0.01 level (2-tailed), * = significant at the 0.05 level (2-tailed).
vary across VR-based intervention sessions. The graph shows that certain VR intervention sessions result in increasing negative cases of the target competencies, as well as the decreasing probability value of the positive emotional states. These findings suggest that it is necessary to provide personalized learning supports to keep learners engaged in VR-based training. Also, the correlation analysis results of this study corroborate that participants’ challenges in training are subsequently associated with negative emotional states. This study finding implies that it is warranted to provide learner-adaptive supports that will dynamically promote the study participants’ cognitive and emotional states during the training. Therefore, using the current automatic assessment enables researchers to gauge which types of design tasks or scenarios are likely to be challenging for the adolescents with ASD. Correspondingly, it can serve to alert the need for learner supports that enhance learners’ cognitive and emotional learning states.

Limitation and further research
Although this study has demonstrated the feasible implementation of the NLP-driven automatic assessment to track and assess representational flexibility of learners with ASD, the study has its limitations. First, the automatic assessment in this study used relatively a small number of training dataset that may increase the likelihood of biased results. Future research should collect more training datasets to enhance the classification accuracy as well as reduce the chance of biased results. It can also use machine-learning classifiers that adopt ensemble-learning techniques that can enhance the generalizability of classifications (Livieris, Drakopoulou, Mikropoulos, Tampakas, & Pintelas, 2018). Second, the assessment is not fully generalizable across different design tasks. The implementation of an NLP-driven classification model tends to be sensitive to contextualized linguistic features. Learners’ design problem solving relies on their actions to collect contextual demands in a given environment. Different design tasks with different backdrop stories and scenarios are highly likely to contain different contextualized information. For example, the exemplary task requests learners to design the elevation bridge based on the design clues provided by the virtual characters in the scenario. Their discourses include high occurrences of contextualized words and concepts related to a given scenario. Further studies should investigate how to configure feature extraction conditions as well as a training dataset to make the automatic assessment generalizable to other design task contexts.

Moreover, future research may use additional measures to corroborate the existing results of the automatic assessment. The current design of the automatic assessment draws on learners’ speech data. Such a data-mining approach will not detect learners’ nonverbal behaviors associated with representational flexibility. For example, learners with ASD could manipulate 3D artifacts without text or verbal utterances, thus, making the existing assessment limited to detect their competency growth. Future research should use online activity logs as an additional measure or source for VR-based learning-state data mining.

Using machine-learning techniques, we can design prediction models that estimate the future patterns governing the competency growth or emotional state change of adolescents with ASD. Subsequently, we can examine how and when various learning supports can be presented and tailored to dynamic behavior and reaction changes of adolescents with ASD.

Acknowledgement
This work was supported by the National Science Foundation [grant number 1837917].
**Statements on open data, ethics and conflict of interest**

This study was conducted with the university IRB (human subject protection) approval. Being constrained by the human subject protection policies and the qualitative nature of the data, the original study data are not open. Anonymous quantitative and qualitative analyses results are accessible upon request. There is no conflict of interest in the work that we are reporting here.

**References**


Appendix A

Examples of the training data (positive case)

<table>
<thead>
<tr>
<th>Target competency</th>
<th>Example descriptions</th>
</tr>
</thead>
</table>
| Attention switching                | “We may need to give us precaution that this bridge should be stable when many people cross it. We already took a look at two potential construction sites. One was located at the river stream with the narrow width. It seems better to build a bridge because it may reduce potential extra-costs.”
|                                    | “We examined what residents in this village strongly need. Although there is a small port alongside the downtown, it was too small. Even, the major port the city is managing is blocked by the river.” |
| Multimodal representation          | “What I am going to do is building each part of the model and putting them together to see how it looks.”                                                                 |
|                                    | “I think that I need to change the texture and color a little bit to fit what I am thinking.”                                                                 |
| Pattern development                | “Okay, now I know what the logic is.”                                                                                                               |
|                                    | “I understand how mass, weight, and force were inter-related.”                                                                                       |
|                                    | “We newly created a new pattern of the bridge mechanics, inspired by the given script that enable the crane to lift the containers in a timely manner. During the exploration, we noticed that there is a regular pattern/routine that lift certain building parts per several seconds. Also, we have realized that once we can systematically change the values and time period of the script, we can build more elaborated bridge.” |
| Pattern contextualization          | “So why don’t we just to go down and then come back to the square and then try modifying it as we discussed.”                                          |
|                                    | “Once we are considering the number of moving carts and people to the island, we should broaden the body of the bridge and then increase the force again.”     |
|                                    | “We could reuse this code to operate this bridge.”                                                                                                   |
|                                    | “I think that we can tweak this default code and then apply it to our bridge.”                                                                      |