TARGETING THERAPEUTIC GAMES TO ADULTS WITH AUTISM SPECTRUM DISORDER

Joseph Bills and Yiu-Kai Ng

Computer Science Department
Brigham Young University
Provo, Utah 84604, USA
roboiguana@gmail.com, ng@compsci.byu.edu

ABSTRACT

Video games could have potential therapeutic value for individuals on the autism spectrum, but little research has been done on targeting games to the diverse individual needs of adults with autism, and the problem is complicated by the inaccessibility of patient profiles. The problem of making personalized recommendations from limited information could be solved by using the patient's taste in games as a proxy for their clinical profile, based on a hypothetical model and updated in response to feedback. This model both enables personalized game recommendation from a cold start and allows the learned information to be generalized to other patients.

KEYWORDS

Game recommendation, autism spectrum disorder (ASD), adults

1. INTRODUCTION

Autism is a disorder that is defined by impairment in social communication and stereotyped behavior (Bartolome, 2013), which is known to be a spectrum disorder, and those having this disorder have a diverse range of strengths and weakness in these areas. For example, deficits in cognitive empathy were once considered to be a universal characteristic of autism, but later research showed this was actually modulated by alexithymia (Bird et al., 2010), which is present in around half of the people with autism (Hill et al., 2004). Games would be most effective if targeted to the needs of the individual, allowing development of the pivotal skills in which they have a relative deficiency such as social initiation. In practice, therapeutic games target specific areas, e.g., Mindlight targets anxiety (Wijnhoven et al., 2015), and these areas vary in patients with autism (White et al., 2009). In practice, autism therapies work best if tailored to the needs of the individual.

Researchers (Ng & Pera, 2018) have hypothesized that video games could be used as therapeutic tools for people on the autism spectrum. In particular, video games could be integrated into Pivotal Response Treatment (Hiniker, 2013), where essential skills are taught in a naturally motivated manner that results in increased functioning in a wide range of areas (Simpson, 2005). Since many people on the autism spectrum demonstrate strong interest in games (Mazurek et al., 2015), and games by design require mastery of certain skills in order to complete the game process, they represent a natural area to investigate for improving a wide range of skills of people on the spectrum. Most research on the subject has been done on children (Hiniker et al., 2013; Wijnhoven et al., 2015), but there is potential for similar therapy to be applied to adults as autism is a lifelong condition, with Cognitive Enhancement Therapy (CET) proving to have satisfactory effects (Eack et al., 2013). Numerous sorts of skills can potentially be developed, ranging from cognitive to emotional and motor to social, and all of these are important. Development in any of these areas can improve quality of life and productivity.

One problem with implementing a targeted approach is that the medical profiles of autistic patients are confidential, so they cannot be easily accessible. Instead, indirect measures need to be used to construct the patient's profile of strengths and weaknesses. We also want to ensure that games developed for autistic adults are fun for the individual so that they remain engaging. Autism is not rare and there is a high demand for effective autism therapies, so creating new therapies that are both effective and desired by the patients is of utmost importance.

Our solution to this problem is to use a profile of games the patient is interested in as a proxy for the clinical profile by assuming that some sort of underlying correlation exists between the game profile and some game feature that we can measure. An initial hypothetical model about what correlations may exist between games a patient likes and areas of a patient's weaknesses is used to make recommendations until empirical data has been gathered about which sorts of games are most effective for people with certain profiles. When empirical data is gathered, predictions

can be made by matching new user profiles with similar profiles and the games that proved effective for them. It is the first study to make personalized game recommendations for autistic adults that target both fun and effectiveness using empirical data gathered over the course of the study, and doing so are more effective than other approaches.

2. RELATED WORK

Granic et al. (2014), who study positive effects of video games, have shown that commercial games can have positive effects on social skills of their players, both in the short and long term if the games contain cooperative elements. Their research, however, do not cover what effects may be specific to people with autism.

Eack et al. (2013) have demonstrated that CET can significantly increase cognitive performance in certain areas for adults with autism. The therapy was originally created for adults with schizophrenia, who suffer from similar social skills deficits as adults with autism. The therapy involves computer-based brain-training exercises, demonstrating the potential for using digital interfaces in improving of cognitive abilities. The authors, however, have not investigated if games specifically may be effective.

Mazurek et al. (2015) investigate what autistic adults opinions on video games are, in terms of their positive and negative effects. One interesting finding is that contrary to popular perception, people have reported more positive social effects than negative social effects. They have also noted the different qualitative elements that influence if someone likes a game. While the study does look into both factors for enjoyment and therapeutic value, they do not come up with a personalized recommendation system like this study aims to do.

Ng and Pera (2018) have developed a system for recommending therapeutic games to adults on the autism spectrum based on personal preference. One of the significant differences of our works and Ng's is that ours accounts for individual differences in what games may be most effective as well as enjoyable, but not the latter.

3. OUR PROPOSED MODEL

A game possesses attributes that can be categorized as either being *qualitative* or *therapeutic*, which determine if a game fits a user's taste and will help with his clinical needs, respectively. For example, genre can act as qualitive attribute because some players prefer strategy games while others prefer action games, while the inclusion of brain-training tasks can be a therapeutic trait. These categories are not strictly separate because elements that affect how someone enjoy a game may also relate to potential therapeutic areas. These attributes can be accessed by learning from *labels* that VideoGameGeek, a social video game website, and other websites provide in structured form on the pages of individual games, or by extracting *phrases* from the descriptions of games on sites such as Wikipedia.

Qualitative traits are elements of a game that appeal to different people, while *therapeutic traits* could help strengthen areas of weakness a person may process. People's enjoyment of a game can be estimated based on the game's attributes and the person's individual taste, while therapeutic value is based on a hypothesis that certain gameplay elements could challenge areas of weakness and strengthen them. For example, difficult action games may challenge fine motor control. Table 1 gives our hypothetical model. The first column enumerates areas of weakness based on those included in "Bridges We Build: The Art of Making Friends" (Scenicview Academy, 2016) which act as psychological constructs. This set of areas of weaknesses was then restricted to those for which existing games could potentially aid with in isolation after categories with too much overlap to be distinguishable were merged. The next two columns are *labels* and *key phrases* that are hypothesized to be correlated with games that possess qualities that could challenge that area. The last column lists examples of games which have those labels on VideoGameGeek and have the intended qualities.

Table 1. The Hypothetical Model

Weakness	VideoGameGeek Label	Key Phrases in Gameplay Sections of Wikipedia Articles	Example Games
Communication	Cooperative (Mode)	Team, cooperative	Secret of Mana, Portal 2, Diablo II
Maintaining eye contact	First Person Shooter (Genre)	FPS, 1st person, player's view	Halo, Call of Duty, Golden Eye
Responding to others	Hotseat (Mode)	Turn-based, social,	Civilization, Advance Wars, Worms

Awareness about sensitive subjects/ Social etiquettes	Simulation (Genre)	Context, controversial, etiquette, manners	The Social Express
Introducing self/ Making friends	MMO (Genre), Massively Multiplayer (Mode)	Online multiplayer, MMORPG, friends	World of Warcraft, Runescape
Handling feedback	Sandbox (Genre), Multiplayer (Mode)	Share, comment, team, creative	Minecraft, Teraria, LEGO Worlds
Resolving conflict	RPG (Genre), Simulation (Genre), Moral Choices (Theme)	Diplomacy, conflict resolution, non-violent	Undertale, Fallout 3
Paying attention	Educational (genre), puzzle (genre)	Focus, details	Brain Age Concentration Training
Difficulty in motor skills	Action (Genre), Wii (Platform), Kinect (Franchise)	Motion controlled, typing, high difficulty, precise	Mario Teaches Typing (Fine), Wii Sports Resort (Gross)
Sensory difficulty (Listening, seeing)	Rhythm(Genre), Music (Theme)	Sound, queues, graphics	Electroplankton

Since we do not have any information about a patient's clinical profile, we must make inferences about it from the gaming profile that he provides, which consists of a set of games that he enjoys, and the set of labels and phrases found to be associated with those games extracted from VideoGameGeek and Wikipedia, respectively. The hypothesis we are operating from is that if an adult with autism is already playing games that challenge a participate areas of his weakness, then that area is not a personal weakness for him, and thus it would not be fruitful to recommend games that train only that area. Games will be filtered from a candidate pool of games specifically designed to target defined areas of weakness, so that only games that target at least one of the areas that the patient is assumed to have a weakness in are included. It's assumed that a user has a weakness in all the areas to begin with unless their profile matches one of the areas in the model, in which case the user profile is said to have *hit* that area.

In order to diversify results for recommended games across patients, areas of weakness will be weighted by how *frequently* they are filtered across all patients, so that an area of weakness will be ranked *higher* if the area occurs less frequently among other areas of weakness. To calculate exactly how high to rank an area of weakness based on the total number of users who had a hit, the total number of users who had hits must be calculated across all the users. After the games are filtered and categorically ranked by area of weakness, they can be recommended to the patients in sorted order so that games with the *minimal* rank score are recommended first which is equivalent to first sorting by *area of weakness*, and then by *fun*.

$RankScore(Profile, Game) = \sum_{User}(Hit_{User, Area\text{-}of\text{-}Weakness(Game)}) - |Labels(Profile) \cap Labels(Game)| / |Labels(Profile) \cup Labels(Game)|$

where $Hit_{User, Area\text{-}of\text{-}Weakness(Game)} = 1$, if the user had a hit in that area of weakness targeted by that game, and 0 otherwise, and Area-of-Weakness(Game) is the area of weakness a particular therapeutic game targets.

Algorithm. Get_Preliminary_Sorted_List_For_All_Users
Input. A list of users, a list of areas of weakness, and a list of candidate games
Output. A sorted list of games
  1. For each user
    a. Request a list of games, *L*, that the user likes for his profile
    b. Initialize a list of *Hits* for each area of weakness, *W*, as **False**
      For each area of weakness, *W*
        For each game in *L*
          i. Look up its page, *P*, on VideoGameGeek
            For each label, *A*, in the hypothetical model for that area of weakness *W*
              If *A* is found on *P*, then
                Record the *Hit* for that area of weakness *W* as **True**
          ii. Look up its article, *T*, on Wikipedia
            For each key phrase, *H*, in the hypothetical model for that area of weakness *W*
              If *H* is found in the gameplay section of *T*, then

Record the *Hit* for that area of weakness *W* as **True**
2. For each area of weakness, *W*
    Initialize Count(*W*) := 0
    For each user, *U*
      If there is a Hit for *U* in *W*, then
        Count(*W*) := Count(*W*) + 1
3. Sort the list of games by *decreasing order* based on their counts for the area of weakness it targets

Example 1. Let Candidate-List-of-Games = {Attention-Trainer, Communication-Trainer, Conflict-Resolution-Trainer, Eye-Contract-Trainer, Feedback-Trainer, Sensory-Trainer}, and Users = {Alice, Bob, Carol}. Also, let Alice-Profile = {Portal 2}, Bob-Profile = {Halo 2, Golden Eye 007}, and Carol-Profile = {Undertale}.
    Assume that the relevant labels of each game are *RL*(Portal 2) = {Cooperative, Puzzle}, *RL*(Halo 2) = {Cooperative, First Person Shooter, Multiplayer}, *RL*(GoldenEye007) = {First Person Shooter, Multiplayer}, and *RL*(Undertale) = {RPG, Moral Choices}. The computed total hits for each area of weakness are Attention (1), Communication (2), Conflict-Resolution (1), Eye-Contact (2), Feedback (2), and Sensory (0), and Algorithm Get_Preliminary_Sorted_List_For_All_Users generates as output the sorted list of games {Communication-Trainer, Eye-Contact-Trainer, Feedback-Trainer, Conflict-Resolution-Trainer, Attention-Trainer, Sensory-Trainer}. □

3.1 Empirical Data

As it stands, there is no known research evaluating the relationship between video game categories and these areas of weakness, so the proposed relationship is strictly hypothetical. However, this hypothesis can be tested by gathering empirical data. Empirical data is collected by measuring how users show improvement in these areas based on responses to questions in a questionnaire. An initial baseline for a patient can be established by having them answer an ad-hoc questionnaire that uses questions gleaned from psychometrically evaluated tools, such as the Autism Quotient (Stevenson & Hart, 2017) or similar custom questions that have been labeled according to the areas of weakness we defined. Examples of such questions are found in Table 2. For each question, the patient is asked to rate themselves as "*Strongly Agree*", "*Agree*", "*Slightly Agree*", "*Neither Agree Nor Disagree*", "*Slightly Disagree*", "*Disagree*", and "*Strongly Disagree*". For each response, a value between -3 and 3 will be assigned depending on whether the question positively or negatively correlates to that area of weakness. The total score for an area of weakness is the sum of the score for each individual question relating to the area. Note that because these tools were evaluated using our own constructs this can only be used as a baseline, not as a substitute for a clinical profile. Without external validity only relative improvement can be measured, not absolute scores, but relative improvement is enough to validate our model.

Table 2. The questionnaire used by the proposed model

Weakness	Questions
Communication	I frequently find that I don't know how to keep a conversation going.
Maintaining eye contact	I find it easy to work out what someone is thinking or feeling just by looking at their face.
Responding to others	I know how to tell if someone listening to me is getting bored.
Introducing self/Making friends	I find it hard to make new friends.
Awareness about sensitive subjects/Social etiquettes	Other people frequently tell me that what I've said is impolite, even though I think it is polite.
Handling feedback	I get upset when my work is criticized.
Resolving conflict	When someone disagrees with me, I can maintain composure.
Paying attention	If there is an interruption, I can switch back to what I was doing very quickly.
Difficulty in motor skills	I enjoy sports.
Sensory difficulty (Listening, seeing)	I find it hard to focus in a noisy environment.

Patients can be reassessed using the same questionnaire after each recommended game they have played in order to monitor how their responses have changed. Based on their responses, a score can be recalculated for each area of weakness, and performance is defined as the new score minus the original score. This information can be used to validate the hypothesis or update the model. If improvement was found in a specific area of weakness, the games the patient played can be labeled as having *affinity* with the profile, and the targeted skill should be *filtered out*. This **affinity score** for a user is represented as a vector with dimension for each candidate game. If no improvement was found, the games will be labeled as having *negative affinity* with the profile and *filtered* from future recommendations. As the process is iterated, the filtering will narrow, ensuring that all areas of weakness will eventually be considered, and they would be exposed to an effective game if any exist. To incorporate fun as well as therapeutic value into the data, the patient will also be asked how much he has enjoyed a game, where they either say they liked the game (1), disliked the game (-1), or were indifferent (0). Based on this another value will be added to the affinity score, but one with a smaller absolute value than that is assigned based on whether it was effective, so half the rating is added. This is done in order to ensure that therapeutic value has more weight.

$$Affinity\_Score(User, Game) = Sgn(Updated\_Score_{Area\ of\ Weakness(Game),User} - Baseline\_Score_{Area\ of\ Weakness(Game),User})$$
$$+ Rating(User) / 2$$

Once there are either no more games to recommend to a player, or a set period of time (e.g., 2 weeks) has passed since this user had his first game recommended to him, the player's profile is considered to be validated. After enough profiles have been validated to make predications, the hypothetical model can be abandoned for certain users as we will now have empirical data for training a machine learning model that can be used to make more accurate predictions than the hypothetical model by predicting affinity scores from a user's profile. Whether or not there is *enough* data to make a prediction for a given user is measured by whether the net similarity between the user's profile and all validated profiles exceeds a certain threshold value, which is defined below. Between any two profiles a *similarity score* is defined in order to gauge if a prediction could be made from existing data. This similarity score is based on the entire set of labels gleaned from VideoGameGeek for all the games that were played, not just those that were included in the hypothetical model, as well as significant monograms and bigrams extracted from the Wikipedia descriptions of those games. It can be calculated using *Jaccard similarity* as

$$Similarity\_Score(A, B) = |\ (Labels(A) \cup Phrases(A))\ \cap\ (Labels(B) \cup Phrases(B))\ |\ /$$
$$|\ Labels(A) \cup Phrases(A)\ \cup\ Labels(B) \cup Phrases(B)\ |$$

where A and B are any two user's profiles.

If a user profile's net similarity across all validated profiles is greater than a threshold value, then his game recommendations will be filtered based on calculated affinity to that user rather than by the hypothetical model. The *threshold value* should be dependent on the machine learning model and the similarity measure. As the maximum value for similarity between two profiles is 1 and thus the maximum possible sum similarity is just the number of elements, a simple way to estimate the threshold is just use the minimum number of instances to train the model.

$$Threshold\_Value = Maximum\ Possible\ Similarity\_Score \times Minimum\ Number\ of\ Instances\ Needed\ To\ Use\ Model$$

Algorithm. Profile_Validation
Input. A user and a list of sorted, recommended games
Output. A vector of Affinity Scores
1. Establish the user's Baseline Score in each Area of Weakness using the questionnaire answered by the user
2. Initialize the user's *Affinity Scores* with each game to *zero*
3. While there are still games in the recommendation list and time remaining for the user to be evaluated
    a. Remove the first game, *G*, from the recommendation list and recommend *G* to the User
    b. Wait for the user to complete *G*
    c. Prompt the user to rate *G*
    d. If the user likes *G*, then
        Set *Affinity_score(G)* := 0.5
      Else
        Set *Affinity_Score(G)* := -0.5
    e. Have the user take the questionnaire again

f. Calculate the Relative Improvement in the Area of Weakness *G* targeted

g. If improvement was found, i.e., *New Score* > **Baseline Score**, then

    i. *Affinity_Score*(*G*) := *Affinity_Score*(*G*) + 1

    ii. Filter out all the games that targeted the particular Area of Weakness

  Else

    iii. *Affinity_Score*(*G*) := *Affinity_Score*(*G*) - 1

3.2 Data Clustering

If enough data, i.e., at least 20 instances as there are ten areas of weakness and two extremes for affinity scores, is collected, a decision tree could be trained to predict whether each individual game would be useful to a user based on the user's profile. However, with limited data due to the slow empirical process another method may be more efficient. (There would have to be at least as many data instances as there are labels before entropy can be meaningfully interpreted.) Hence, a semi-supervised model based on first clustering validated profiles with similar affinity scores and then associating an unvalidated profile using K-Nearest Neighbors (KNN) may work better with limited information as it assumes users fall into a smaller number of classes. Here KNN used unweighted voting, and in the event of a tie the furthest neighbor is excluded. KNN with an arbitrary small *k*-value, such as four, on *k*-1 clusters could probably start making predictions sooner as it assumes simpler boundaries between qualitatively distinct categories of user profile. For this model, the threshold is *k*+1 to ensure that at least one element will not be included in the set of nearest neighbors. KNN predicts which cluster a new profile would be assigned to, and returns the vector of affinity scores associated with the cluster. Clusters are determined using a variation on K-Means for which distance is defined as cosine similarity between the affinity scores of a validated game and an estimated mean. Once profiles are assigned to clusters, that cluster can be added as a label which can then be predicted for unvalidated profiles using KNN, and from the cluster a set of affinity scores is predicted. The predicted affinity scores returned for a cluster will be the same as the *mean* that defines the cluster. Games will be ranked by affinity score and the highest-ranking game will be recommended.

Algorithm. Recommend_Games
Input. A list of users, and a list of candidate games
Output. A sorted recommendation list for each user
1. Initialize the set of Validated Profiles as an empty set
2. Preliminarily sort the list of games for all users using the Get_Preliminary_Sorted_List_For_All_Users Algorithm
3. For each user, *U*

  a. Initialize *Net Similarity* to *zero*

  b. For Each Validated Profile

    i. Calculate the Similarity Score between the Validated Profile and *U*'s Profile

    ii. Add this *Similarity Score* to the *Net Similarity Score*

  c. If Net Similarity Score > Threshold, i.e., *Net Similarity* > *K* + 1, then

    i. Use K-Nearest Neighbors to assign the *U*'s Profile to a cluster of Validated Profiles

    ii. Set the Predicted Score for each game in the recommendation list as the average Affinity Score in the cluster

    iii. Sort the Recommendation List by Predicated Scores in decreasing order

  Else

    iv. Initialize *U*'s Recommendation List to the Preliminary Sorted List

    v. For Each Area of Weakness

      (a) If *U* has a Hit in that area, then

        Filter All Games Targeting that Area out of the User's Recommendation List

    vi. Further sort the List by comparing games that target the same Area of Weakness and rating the one with a Higher Similarity Score to the User's Profile Higher

  d. After the Recommendation List is filtered & sorted, make recommendations using the Profile_Validation algorithm

  e. Record empirical data by adding *U*'s Profile and Affinity Scores to the list of Validated Profiles

  f. Use K-Means to Divide the Validated Profiles into *k*-1 Clusters

4. VALIDATION TESTS

In order to test the validity of the method, an empirical study must be conducted. However, we can assess some other aspects of the model without gathering empirical data. One way to do this is look at the internal correlation of the hypothetical model, seeing if the games picked up from the labels on VideoGameGeek correspond with games containing key phrases on Wikipedia. In general, if any two measures are highly correlated, they likely measure the same construct, though they should not be perfectly correlated as in that case using multiple measures is redundant. For each area of weakness that an individual game may help strengthen, a hit can be separately calculated using the labels in VideoGameGeek and key phrases from Wikipedia. Correlation between hits, VideoGameGeek hits, and Wikipedia hits in each area of weakness over a collection of games can be calculated using the Pearson Φ *coefficient*, where each count is based on the number of games that were *hit* or *missed* on either VideoGameGeek or Wikipedia.

$$B_a = \Sigma_g\,(V_{g,a} \times W_{g,a}), \quad V_a = \Sigma_g\,(V_{g,a} \times (1 - W_{g,a})), \quad W_a = \Sigma_g\,((1 - V_{g,a}) \times W_{g,a}), \quad N_a = \Sigma_g\,((1 - V_{g,a}) \times (1 - W_{g,a}))$$
$$\Phi_a = (B_a \times N_a - V_a \times W_a) / \mathrm{Sqrt}(B_a \times N_a \times V_a \times W_a)$$

where $V_{g,a}$ is the game g matched that area a on VideoGameGeek and $W_{g,a}$ is the game g matched that area a on Wikipedia, and B, V, W, and N stand for total instances of cases where games matched both VideoGameGeek and Wikipedia (B), only video game geek (V), only Wikipedia (W), or neither (N), respectively.

4.1 Game Evaluation

Another measure that can be used is estimating the likelihood that the hypothesized terms for an area of weakness would get any hits in practice. If a hit is unlikely to occur, then the measure would be a useless discriminator between profiles in practice. Likewise, a measure that always produces a hit would be useless, though this situation is less likely to occur. To estimate this, most popular games on VideoGameGeek could be sampled. Without having sample profiles, there is no way to estimate how popular a game must be for it to be represented, so arbitrarily the top 1,000 most popular games are chosen for the sample. From these games, the total number of hits in each area could be calculated, where a hit is registered for a game either if it contains the label on its VideoGameGeek page or the corresponding Wikipedia article contains the key phrase in its gameplay section. Values other than 1,000 can also be used to estimate frequency of hits at different levels of popularity, but without sample game profiles variance cannot be calculated to know the actual distribution of played games among patients and thus the real probability that any hits occur. Values of 1,000 can be tested to see if the distribution changes at different orders of magnitude.

Finally, while empirical data is needed to determine if the recommended games are effective, returned games that register a hit can manually be looked at to see if they contain the feature that is hypothesized to challenge areas of weakness. The initial hypothetical model was based on identifying specific games where those features were identified, so the hypothesis entails that not only do these features help with areas of weakness in practice, but also that these labels and key phrases are good predictors for these features. As those features are identified, their presence can be tested without further empirical validation; it would just require looking through a sample of games that is large enough to be statistically significant.

4.2 Control Group Evaluation

The effectiveness of the entire procedure can be evaluated by clinically testing each patient with autism spectrum disorder who volunteer for the project who is not assigned to testing just the hypothetical model. Each patient here will be evaluated at the start of the experiment, and again at the end of a trial period. This trial period should be around a month to give patients time to work through all of the games. These patients should be randomly assigned into three groups: a *control group* that is recommended no games, a *baseline group* that play games that are recommended solely based on personal enjoyment, and a *test group* that have games recommended for them using the system described here. Ideally each group would have at least nine people in it so that clustering can be tested. To ensure that clustering is tested, the patients in the 3rd group should be staggered so one patient starts every three days, ensuring that a validated set is built up before later patients start receiving recommendations. Even though the patients are staggered, the total number of hits is calculated for the entire group at the beginning. Each group can be reassessed to determine what percentage of the people in them show significant improvement, with the clinicians reporting the degree of improvement for each patient according to whatever measure they choose. The resulting

percentages can be reported without revealing any personal information by using a double-blind approach where clinicians are ignorant to what group patients are assigned to, but the computer system knows, and the percentages of users who show improvement in each group are generated automatically when all the reports are made.

If no significant improvement can be seen in either the *baseline* or the *test group* over the *control group*, then the preliminary hypothesis that the recommended games are effective therapy must be rejected. In this, no conclusion can be made about the relative advantage of a recommendation system that targets to individual weaknesses over one that does not. The fundamental construct of what games are used and how they are administered must be revised before it can be determined if this targeted system gives a relative advantage over an untargeted one. If the *test group* significantly outperforms the *baseline group*, then it can be concluded that this system works better for targeting individual weaknesses by *rejecting* the *null hypothesis*.

5. CONCLUSION

Previous research on recommendation systems for therapeutic games focused only on personal preference for the game (Ng & Pera, 2018). By recommending games based on individual therapeutic value as well as qualitative value, we can ensure that patients will not only have fun, but that they will also develop the skills they need. We hope to solve this problem by finding a correlation between game preference and areas of weakness. Then recommendations can be made based primarily on that correlation, and secondarily on games that are similar to those that are enjoyed. By targeting games to a patient's individual needs, they can be used more effectively, leading to a more efficient mastery of essential skills.

REFERENCES

Book

Scenicview Academy, 2016. *Bridges We Build: The art of making friends*.

Journal

Bird, G. et al, 2010. Empathic Brain Responses in Insula are Modulated by Levels of Alexithymia but not Autism. In *Brain: A Journal of Neurology*, Vol. 133, Issue 5, pp. 1515-1525.

Eack, S. et al, 2013. Cognitive Enhancement Therapy for Adults with Autism Spectrum Disorder: Results of an 18-Month Feasibility Study. In *Journal of Autism and Developmental Disorders*, Vol. 43, No. 12, pp. 2866-2877.

Granic, I. et al, 2014. The Benefits of Playing Video Games. In *American Psychologist*, Vol. 69, No. 1, pp. 66-78.

Hill, E. et al, 2004. Brief Report: Cognitive Processing of Own Emotions in Individuals with Autistic Spectrum Disorder and in their Relatives. In *Journal of Autism & Developmental Disorders*, Vol. 34, No. 2, pp. 229-235

Mazurek, M. et al, 2015. Video Games from the Perspective of Adults with Autism Spectrum Disorder. In *Computers in Human behavior*, Vol. 51, Part A, pp. 122-130.

Simpson, R., 2005. Evidence-based Practices and Students with Autism Spectrum Disorders. In *Focus on Autism and Other Developmental Disabilities*, Vol. 20, No. 3, pp. 140-149.

Stevenson, J. and Hart, K., 2017. Psychometric Properties of the Autism-Spectrum Quotient for Assessing Low and High Levels of Autistic Traits in College Students. In *Journal of Autism and Developmental Disorders*, Vol. 47, No. 6, pp. 1838-1853.

White, S. et al, 2009. Anxiety in Children and Adolescents with Autism Spectrum Disorders. In *Clinical Psychology Review*, Vol. 29, No. 3, pp. 216-229.

Wijnhoven, L. et al, 2015. The Effect of the Video Game Mindlight on Anxiety Symptoms in Children with an Autism Spectrum Disorder. In *BMC Psychiatry,* Vol. 15, No. 1, pp. 138.

Conference

Bartolome, N. et al, 2013. Autism Spectrum Disorder Children Interaction Skills Measurement Using Computer Games. *Proceedings of the 18th* IEEE *International Conference on Computer Games: AI, Animation, Mobile, Interactive Multimedia, Educational & Serious Games (CGAMES)*, pp. 207-211.

Hiniker, A. et al, 2013. Go Go Games: Therapeutic Video Games for Children with Autism Spectrum Disorders. *Proceedings of the 12th International Conference on Interaction Design and Children (IDC)*, pp. 463-466.

Ng, Y. and Pera, P., 2018. Recommending Social-Interactive Games for Adults with Autism Spectrum Disorders (ASD). *Proceedings of the 12th ACM Recommender Systems Conference*. Vancouver, Canada, pp. 209-213.