2017 IEEE International Symposium on Mixed and Augmented Reality Adjunct Proceedings

Daily Living Skills Training in Virtual Reality to Help Children with Autism Spectrum Disorder in a Real Shopping Scenario

Ali Adjorlu*
Aalborg University Copenhagen

Emil Rosenlund Høeg*
Aalborg University Copenhagen

Luca Mangano†
Aalborg University Copenhagen

Stefania Serafin*
Aalborg University Copenhagen

Abstract

In this paper, we present a study conducted to investigate the feasibility and effectiveness of Virtual Reality (VR) applied to daily living skills (DLS) training of individuals diagnosed with Autism Spectrum Disorder (ASD). In collaboration with a teacher at a school for children and adolescents with mental disorders, a head-mounted display based VR simulation of a supermarket was built and evaluated with the purpose of developing the shopping skills of students diagnosed with ASD. A comparative between-group experiment was conducted on 9 participants, with initiated VR training following a baseline assessment in a real supermarket. After running seven VR training sessions over 10 days for the treatment group, participants were assessed again in the real supermarket. Results show some benefit of training DLS using VR as discussed in the paper.

Index Terms: H5.1 [Multimedia Information Systems]: Artificial, augmented, and virtual realities—; H 5.2 [User Interfaces]: User-centered design —

1 Introduction

Autism spectrum disorder (ASD) is a pervasive and lifelong neurodevelopmental disorder, characterized by severe impairment in behavioral, social and communicative skills of individuals diagnosed with it [19, 29]. The fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5) characterizes ASD by dividing it into three levels of severity. The third level is the most severe, requiring very substantial support in daily living skills (DLS) due to serious deficits in verbal and nonverbal communications skills [2]. The higher functioning individuals placed in the first and least severe level of the DSM-5 are capable of verbal communication but still require support due to difficulties processing sensory input and understanding social cues. The deficits described in DSM-5 are associated with lower DLS among individuals diagnosed with ASD compared to their IQ-matched peers [16, 26]. This causes the diagnosed individuals to feel overwhelmed and anxious in unfamiliar or crowded environments such as a public transportation, educational institutions and, of course, supermarkets. Consequently, a vast number of adults diagnosed with ASD rely on support from their parents or social services. According to [11], the societal cost during the lifespan of each person diagnosed with ASD in the USA is 3.2 million dollars, which consists mainly of lost productivity and costs of adult care services. As adults, only two-thirds of the ASD diagnosed individuals can provide themselves with basic personal care [34], while many experience unemployment or underemployment.

A study conducted by Hus Bal et al. [4] examined the independent variables for DLS attainment in a sample group diagnosed with ASD ranging from 2 to 21 years of age. Via interviews and questionnaires of both the parents and the diagnosed individuals, the study suggested a decline ability to develop DLS during adolescence, particularly for domestic skills such as food preparation, shopping and household safety. The study concludes that DLS should be a focus of educational plans for young individuals diagnosed with ASD. Another study [32] reported that young people diagnosed with ASD and Down syndrome developed their DLS across adolescence and early 20's equally. However, while the group diagnosed with Down syndrome continued their development, the group diagnosed with ASD attainment of DLS appeared to solidify in their late 20s. In another long term study conducted by Szatmari et al. [35] DLS improved during childhood and then flattened in late adolescence for an ASD diagnosed sample. This clearly highlights the need for early education programs to specifically target DLS in order to promote functional independence during adulthood. Approximately 1 in 100 individuals are diagnosed with ASD [3, 6] while more recent estimates in the United States suggest even higher numbers [9]. This showcases both the importance and urgency of DLS learning interventions.

One of the first publications on DLS training was a procedure on how to teach a young boy diagnosed with ASD how to wear glasses without help from his parents in 1964 [39]. Since then, behavioral training interventions have been frequently researched to teach a broad variety of skills to autistic children. These approaches include behavioral skills training (BST) mediated by a teacher trying to role play and discuss right behavior in different social scenarios [25, 36], and video modeling (VM) which is a form of observational learning that involves recording a child performing and behavior successfully. The footage is then edited into 1-2 minutes video clips and is shown to the child multiple times per week, for approximately 2-8 weeks [1]. These interventions have shown to effectively teach new behavioral skills in specific cases. However, due to the ASD populations' decreased ability to generalize knowledge from one context to other, DLS interventions often fall short when the individual is presented with a deviation of the rehearsed task [18].

One way to present deviation in DLS training is simulation. Simulation is a technique in which the learner is actively placed in a modulated situation or context in which practice and learning can be achieved in a safe environment. Training in a simulated graphical representation of reality, also known as virtual reality (VR), invites the two interlocking questions: How readily can the acquired training be transferred to a real environment and social situation? And furthermore, what tools can be applied to facilitate optimal learning? For a VR simulation training to be successful, the transfer process should allow the corresponding real task to be completed effectively in the virtual environment according to predetermined criteria [5].

*e-mail: {adj,erh,sts}@create.aau.dk
†e-mail: lu.mangano@gmail.com

978-0-7695-6327-5/17 $31.00 © 2017 IEEE
DOI 10.1109/ISMAR-Adjunct.2017.93

According to several sources [12, 23, 24], multimodal interfaces can generally create better learning conditions through inter alia cognitive integration and limited cognitive load compared to single modality learning interfaces [21]. Furthermore, technology-based learning has been suggested to be especially effective learning tool for children affected with ASD [38]. In essence, a VR simulation can immerse the learner in a situation where tasks can be performed and repeated in safe surroundings i.e., safer learning environments. In [8] it is argued that DLS teachings that are less verbally mediated, and involve more spatial and visual cues are more preserved in young people diagnosed with ASD.
There are a number of studies that apply visual cues in virtual environments to teach DLS to children and adolescents diagnosed with ASD. Didehbani, et al measured the effect of interaction with a desktop virtual environment to train social skills on thirty children diagnosed with ASD between the age 7-16 [15]. The participants completed a variety of true-to-life social scenarios. Didehbani, et al. included four stages in the experiment: introduction, modeling, rehearsal, and feedback. By logging in and controlling an avatar in the virtual environment, a coach provided the user with both introduction and feedback while the child was modelling and rehearsing the right behavior. A pre- and post-measurements of emotion recognition, social attribution, attention, and executive function were conducted. The results show an improvement in emotion recognition, social attribution and executive function of the user diagnosed with ASD. Most research on virtual environments and ASD focuses on social training similar to Didehbani. One of the studies examining the potentials of DLS training in virtual environments was conducted by Sara Parsons, et al. [27]. Parsons, et al. examined the effects of a desktop-based virtual café and bus environment to rehearse adaptive behaviors with children diagnosed with ASD. Using keyboard and mouse, the user could move around the virtual café, order food and choose a table. The bus experience was designed to rehearse waiting in the queue and choosing a place to sit. It was evaluated via a descriptive, qualitative study with two participants. Results of the experiment indicated that the users enjoyed the virtual environment experience, and provided specific examples on how they could generalize their experiences to real life scenarios. There are several desktop-based VR studies that focus on children with ASD such as Self, et al. [31] solution to teach fire and tornado safety skills, or Josman, et al [14] studying the effects of virtual environments in teaching safe street crossing skills to children and adolescent diagnosed with ASD. Maskey, et al. developed and evaluated CAVE-based virtual reality environments to perform DLS training for children diagnosed with ASD [17]. Nine scenarios, all of which were reported to be uncomfortable situations for the test participants were designed and evaluated through single subject studies. These scenarios included a bus ride, shopping, and a school classroom experience. The results show a positive effect of the training on four out of the nine participants. However, CAVE for virtual reality experiments is expensive to build and require dedicated space, which is not always available in public schools. In this paper, we present a display based VR experience designed to teach shopping skills to a young target group diagnosed with ASD. Due to the novelty of affordable consumer VR equipment, there has been limited research on interventions in VR to teach DLS to young people diagnosed with ASD [27]. A study conducted by Newbutt et al. [22] investigated the ASD populations acceptance, enjoyment and sense of presence during head-mounted display based virtual reality experiences. Results of the study show a positive attitude towards head-mounted display based virtual reality experiences from individuals diagnosed with ASD.

In this paper we present an experiment aimed to evaluate head-mounted display VR based supermarket shopping training systems at a Valhøj School in Copenhagen. Amongst the hundreds of students, the school welcomes children and young adults diagnosed with ASD. Part of their training for the students diagnosed with ASD consists of home economics and DLS training, which is conducted as field trips to the local supermarket. This specific training was designed after several discussions with the teachers of the school. The aim of the study is to help students to develop the necessary skills to conduct a safe shopping experience independently. According to the teachers, many of the students have been observed to have difficulties in particular with managing money, locating and selecting products, and experience general discomfort.

2 Methods

A total of 9 participants diagnosed with ASD ranging in age from 12 to 15 were scheduled to take part in the study. The group included 8 males and 1 female, which is roughly consistent with the overall gender ratio of 5:1 (male: female) diagnosed with ASD [10]. A between-group study was planned with the students divided into a treatment and control group by the teacher due to the authors limited access to the participants (4 in experimental group: 5 in control group).

The experiment was conducted in a period of ten days starting with a baseline and ending with a post treatment assessment in a real supermarket (between April 18th and April 28th, 2017). In between the two assessments in the real shopping center, the treatment group was required to train with the VR supermarket system once a day for a total of 7 sessions, while the control group would receive no treatment. The VR training sessions were mediated by a teacher while the authors had no direct contact with the students during the whole process. Furthermore, a parent or legal guardian of each participant were provided with a written informed consent of the research study and they all assent to participate.

2.1 Baseline and post-treatment assessment

The baseline and post-treatment assessments were established in a real shopping scenario. Each participant was assigned one out of nine unique shopping lists consisting of four products each (two from the general shelves, and minimum one from the fruits and vegetable and diary products section). In order to evaluate efficiency and effectiveness of the participants shopping, start and end times of the shopping sessions were recorded for each individual. Furthermore, the teachers were asked to photograph the baskets with products and shopping lists visible after the assessment in order to gather data on the number of right products picked up by each individual during the assessment.

Additionally, the participants were presented with a questionnaire after both the baseline and post-treatment assessments in the real supermarket. The purpose of the questionnaire was to measure the participants' self-reported confidence level, and a prospected hypothetical confidence level if they were to shop alone the next day. Furthermore, the questionnaire was used to evaluate how easy it was to shop in the real supermarket as well as an open-ended question on which assistive element they used to locate the products. The questionnaire was inspired in part by pictorial confidence rating scale used by Maskey et al. [17] to subjectively evaluate the level of confidence and comfort. Lastly, key events were observed and noted via a purposely built web application that was developed for smartphone use. Three of the authors and two research assistants conducted the observation in the field, by covering different areas of the supermarket while avoiding direct contact with the sample group. All of the test participants were assigned specific shopping lists on distinguishable colored papers to ease the data collection for the observers. The teachers told the participants to keep the shopping list papers visible at all time, while the observers did not know which color belonged to the experiment group until after the assessments in order to avoid any bias.

2.2 VR Training

The VR training was conducted by a teacher at Valhoj school in Copenhagen. The treatment group trained with the VR intervention every day between the baseline and final evaluation. The training required the participants to enter the virtual supermarket and locate and place 4 items from their virtual shopping list in to their virtual shopping basket. The shopping list would change every day. During the VR sessions, the system logged the following data:

- Product touched
- Hint asked
- Product put in the basket
- Product grabbed
- Product released
- Total time spent in the experience

After each session, the participants were asked to fill in questionnaires aimed to assess their comfort during the experience and their comfort if they were to go to a supermarket by themselves the following day as well as their satisfaction and ease-of-use with the VR experience.

3 VR Intervention

The VR intervention was developed using Autodesk Maya and Unity. HTC Vive was chosen to run the application due to its effective room scale tracking. A custom computer was build for the purpose of the research, consisting of an Intel i7 7700k processor, GTX 1080 GPU, and 16GB of DDR4 ram. According to [5], a low road towards the transfer of knowledge requires rehearsal of skills in a context very similar to to the context in which the skills are going to be performed. Therefore, the virtual environment was designed and implemented to look like a local supermarket next to the school in which the baseline- and post assessments were conducted. The user can move around in a 1:1 virtual version of the supermarket via walking within the Vive lighthouse area and VR teleportation. The VIVE lighthouse area works via two lighthouse stations that send out infrared light signals. These signals are than captured via the infrared sensors placed on the VIVE head mounted display and controllers in order to achieve location tracking. Teleportation in virtual reality works by using the controller to point at the desired location and pressing a button. The user is then teleported to the specified location. The teleportation interaction was already familiar to the sample group due to their previous experience playing VR games on the HTC Vive with the teacher. Using the HTC Vive controllers, the user can look at a shopping list, grab items from the shelves and place them in the virtual shopping basket (figure 1).

The layout of the store was reconstructed by creating 3D models of the supermarket's shelves, signs, and flags and placing them in the virtual environment. All sections were reconstructed with custom made models specific to the original supermarket, i.e., shelf-systems and signs. A comparison between pictures taken in the real supermarket and their virtual version can be seen in figures 2, 3 and 4.

A total of 32 specific products was modeled, texturized and placed on the shelves. The textures for the virtual products were created using photographs of the actual products. Teleportation was added due to the limited physical space available for walking due to HTC Vive's tracking coverage limit of 5*5 meter. A Vive sensor was placed on a real shopping basket to be used in the simulation in the participants of-hand. A 3D model of the basket was implemented in the VR experience to further increase immersion through valid reinforced sensorimotor actions (figure 5).

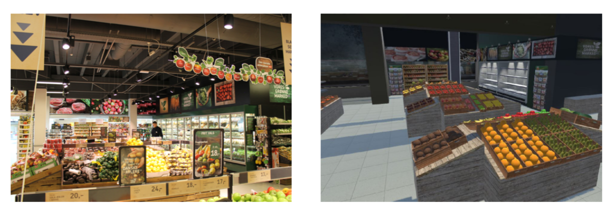
Figure 1: Virtual shopping list and basket

(a) Physical (b) Virtual

Figure 2: Comparison between the real shelfs and their reconstructed equivalent in the VR supermarket

(a) Physical (b) Virtual

Figure 3: Comparison between the vegetable and fruit department and their reconstructed equivalent in the VR supermarket



(a) Physical (b) Virtual

Figure 4: Comparison between the real freezers in the dairy department and their reconstructed equivalent in the VR supermarket

An arrow pointing towards the direction of the next product on the shopping list was implemented as a hint. The hint could be activated by the teacher when the teacher felt that the student needed guidance.

In order to teach the basic interactions to the sample group before the evaluation, a tutorial scene was designed and developed. The tutorial used voice instructions to introduce interaction schemes such as closing and opening of the visible shopping list and grabbing

(a) Physical (b) Virtual

Figure 5: Physical basket with the attached vive sensor and its reconstructed equivalent in the VR supermarket

items from the shelve and placing them in the basket.

4 Results

The goal of this study was to investigate whether skills developed in a VR environment could be transferred to a similar real world situation. Triangulation of 3 distinct methods were used to evaluate pre- and post-experience performance of the user:

- Task completion time and effectiveness
- Questionnaires
- Observations

Furthermore, data logging performed by the VR-system was used for data cross verification, as well as questionnaires filled out by the sample group and the teacher.

4.1 Task completion time

The task completion time describes the participant's exact amount of time spent shopping in the supermarket. The time of entering the supermarket was noted by the teachers while the completion time was recorded by photographing each students shopping basket as soon as they returned to the teacher upon finishing their given shopping assignments. The Exchangeable Image File Format encoding of each image contains information on the exact time the image was taken and was used as reference for measuring task completion time. Results are reported in Figure 6. Both groups spent more time in the supermarket during the post-treatment assessment compared to the baseline assessment. This could be due to one of the independent variables: the number of customers in the supermarket which was much higher during the post-treatment compared to the baseline assessment. The treatment group was 11 seconds slower than the control group in the baseline evaluation, while during the post-treatment evaluation the treatment group was on average 52 seconds faster than the control group.

4.2 Effectiveness

Effectiveness is frequently used in the field of Human-Computer Interaction (HCI) to conduct usability evaluations and is often referred to as Completion Rate (CR) [30]. Effectiveness is calculated as:

$$Effectiveness = \frac{ST}{TAT} * 100\%$$

where ST represent the successful tasks and TAT represent the total amount of tasks. Task success or failure is coded as 0 for failure: incorrect product/no product bought, 1 for success: correct product bought while 0.5 is used to describe partial completion criteria: the wrong quality or wrong type of product bought. Effectiveness results for the baseline- and post-treatment evaluation can be seen in 7.

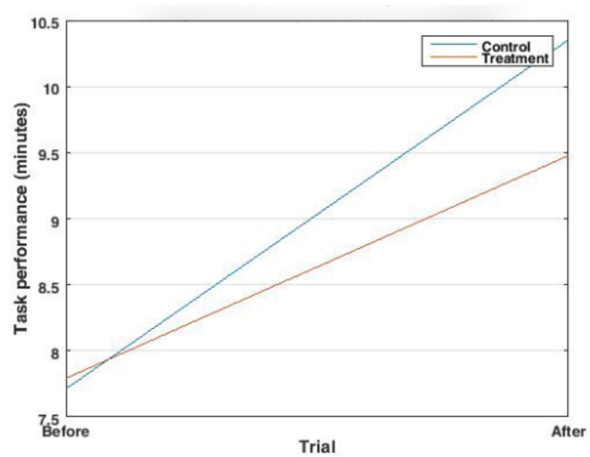


Figure 6: Line plot showing average Task completion time(TCT) of the treatment- and control groups between baseline and final shopping assessment

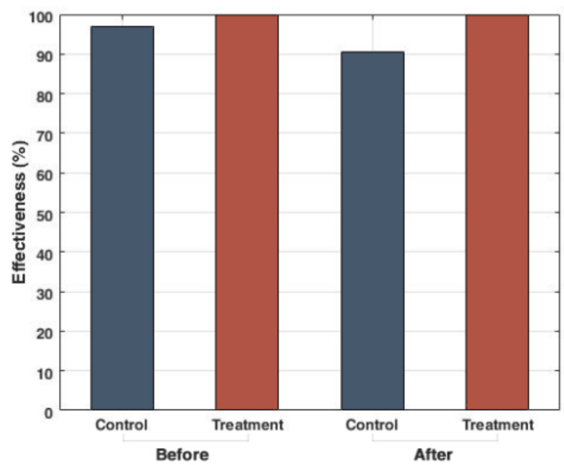


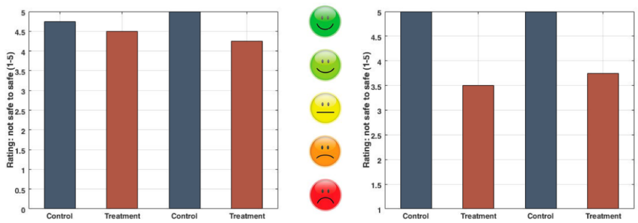
Figure 7: Average shopping effectiveness (%) of tasks completed correctly between control group (blue) and treatment group (red)

Control groups effectiveness score decreased from 97% in the baseline evaluation to 91% in the post-treatment evaluation. Participant **H** scored 3.5/4 during the baseline assessment (partial completion score due to only buying one pack of crispbread instead of two). During the post-treatment session participant **E** scored 3/4 (did not buy any crispbread) while participant **I** scored 3.5/4 (bought mild cheese instead of mature cheese). The treatment group efficiency score was 100% during both the baseline- and post-treatment assessments.

4.3 Pre- and post-treatment questionnaires

Results of the pre- and post-treatment questionnaires are reported in Figures 8 and 9.

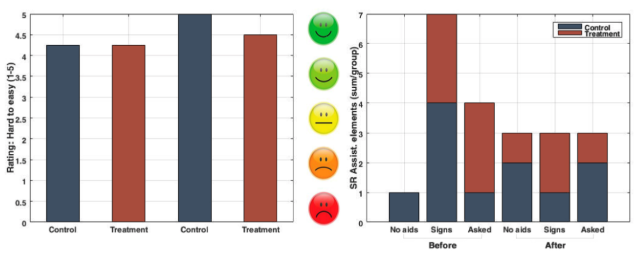
The control group's average confidence rating increased from 4.75 in baseline assessment to 5 in post-treatment assessment while their prospected confidence did not change. The treatment group 's average confidence decreased from 4.5 at the baseline to 4.25 in the post-treatment assessment, while their average prospected confidence increased from 3.5 to 3.75. The control group 's average perception of how easy it was to shop in the real supermarket increased from 4.25 to 5 between the assessments while the treatment



Confidence-level

Prospected Confidence-level

Figure 8: Self-evaluation of the average confidence per group om the shopping situation, and of the prospected confidence in shopping alone in the future



Ease of shopping

Assistive elements used

Figure 9: Self-evaluation of how easy it was for the participant to shop today, and the summed self-reported assistive elements used in the baseline and post-treatment evaluation in the supermarket

group saw an increase from 4.25 to 4.5. Finally, for the question regarding assistive elements, 4 participants in the control group and 3 in the treatment group reported having used signs for help in the baseline evaluation while the number decreased to 1 for the control group and 2 in the treatment group during the post-experiment evaluation. 4 participants reported that they asked for help in the baseline (1 control group, 3 treatment group) while 2 from the control group asked for help in the post-treatment assessment and only 1 asked for help in the treatment group.

4.4 Observations

The most compelling data gained trough observation recorded for baseline- and post-treatment evaluation is the number of times the participants asked for help, represented by figure 10.

The treatment group was never observed to ask teachers for help but asked several times (3) other customers for help. They have also been found to ask once the supermarket clerks for help. On the other hand, during the baseline evaluation, control group has been observed to use all three entities help (1-clerks, 2-teachers, 2-customers). During the final evaluation, the control group was observed to not ask teachers or clerks anymore, and only twice other customers were inquired. Furthermore, it was observed that participant **D** was accompanied by one the schools teachers into the supermarket during the baseline assessment. However, during the post-treatment assessment he was observed entering the supermarket alone.

4.5 Difference in difference

There was a significant increase in the number of other customers in the supermarket from the baseline to the post-treatment evaluation. Therefore a difference-in-difference (DiD) estimate was applied which is an useful method when both the control and treatment group are mutually influenced by external factors [7]. DiD calculates the difference in the difference, where the change is computed

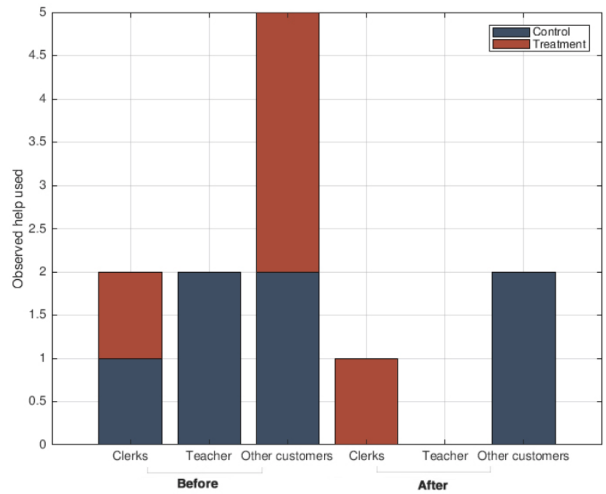


Figure 10: Helped asked by control group and treatment group

as the value after treatment minus value before treatment. The DiD regression-estimates was tested for significance with significance levels of 1%, 5%, and 10% (= 0.01, = 0.05 & = 0.10) due to small sample size. the results are presented in Figure 11, where the 3-coefficient is the standardized coefficient resulting from the DiD-regression analysis. The predictors are independent variables

Predictor(s)	Coeff. (β_3)	Lower CI	Upper CI	*p*-value
1) Raw DiD-estimate	-1	4.53	2.44	0.53
2) Effectiveness	-1.58	-5.13	1.97	0.36
3) Effectiveness Required help	-3.11	-6.27	0.05	0.054*
4) Effectiveness Required help Est. Traverse	-3.26	-7.06	0.55	0.087 *
5) Effectiveness Required help Green section	-3.92	-7.56	-0.29	0.036**

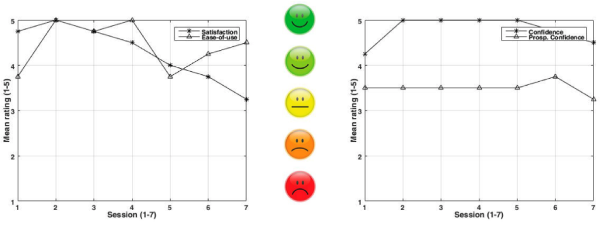
Inference: *** *p<0.01; ** p<0.05; * p<0.1*

Figure 11: Difference in difference (DiD) results

added in order to explain casual inference i.e., elements that might explain the variation in the dependent variable. Five different scenarios have been defined: raw DiD-estimate of the treatment effect with no independent variable as predictors, effectiveness as predictor, required help as predictor (Assistive element from the questionnaire ”asked for help”), traverse as predictor (measure as the optimal route to get all products for the shopping lists and finally a binary traverse-estimate of whether the shopping list contains items from the green section of the supermarket as a predictor. The last predictor was based on the green section of the supermarket, which had significantly more products in a smaller area and in a more unstructured and disorderly setup with less space to navigate. The results of the DiD can be seen in figure 11. The raw DiD-regression without independent variables resulted in a DiD- estimate 3 = 1. The result is not significant at a 10% (P > 0.10) level of significance. Nor was there any significant difference when adding Effectiveness as a predictor (3 = -1.58). However, the DiD-estimate treatment group had significant (P < 0.10) change in performance with the Required help-predictor, and significantly ($P < 0.10$) better performance with added estimated traverse-predictor. The most significant difference ($P < 0.05$) in performance was achieved with the Effectiveness-, Required help- and Green section-predictor.

4.6 VR-session evaluation

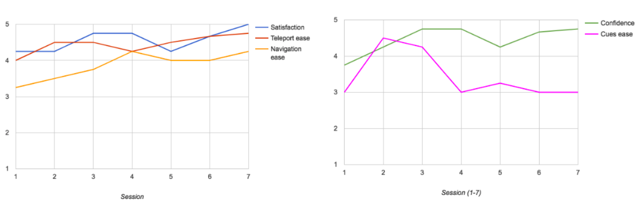
The treatment group went through 7 sessions with the VR supermarket. The results from the self-assessment questionnaires conducted after the participants finished each VR supermarket experience are presented in figure 12.



Ease-of-use/satisfaction   Confidence-rating

Figure 12: Ease-of-use (triangle) and satisfaction (asterisk) of the system. The pictorial scale used, and confidence in the VR experience (asterisk) and prospected confidence in a real shopping scenario (triangle)

Ease of use in the first session has an average of slightly below 4, but increases to 5 and remain relatively steady trough sessions 2-4, where it decreases during the 5th session and steadily increases again in session 6-7. Overall satisfaction appears to decrease steadily from around 5 in the two last sessions and ending a little over 3 in the final session. Confidence ratings presented in Figure 12 are stable from the second to the 5th session, but decreases slightly in the last two sessions, a similar trend can be noticed in prospected confidence, which remains unchanged from session 1-5, but makes a slight fluctuation in the two last sessions. The teacher explained the decline in the last three sessions, describing the he tried to push the students out of the comfort zone by negating the students access to the hint (3D arrow pointing at the right product), telling them to try to find the product by looking at the shelfs in the VR supermarket. The teacher was also asked to submit a questionnaire after each participant finished their VR experience. Figure 13 shows the evolution of the results means over time.



(a) Satisfaction, teleport ease, navigation ease   (b) Confidence, cues ease

Figure 13: (a) Satisfaction = blue, teleport ease = red, navigation ease = orange. (b) Confidence = green, cues ease = pink

The graph 13(a) reports satisfaction (blue), confidence (red), and teleportation technique ease (orange), while graph 13(b) presents cues usage ease (green), and navigation ease (pink). Each category has been represented based on means of the rating each participant was evaluated by the teacher. All of the categories have seen a tendency to increase, with the exception of cues ease. The cues usage ease dropped on the 4th session, i.e. the beginning of the second week of training. On the same day, it can be observed that teleportation ease slightly dropped, but returned and surpassed to previous values the following sessions. Confidence has seen an increase during the sessions (3.75 in session 1 and 4.75 in session 7), with the exception of session number 5, in which it dropped by half a point. Satisfaction had a similar behavior, but started from 4.25 and ended at 5.0. Navigation ease steadily increased since session 1 (3.25) and reached its peak in session 4 (4.25), afterward it dropped half a point, but it was recovered on the last session.

Data were logged each day and subsequently extracted from the testing computer. Task completion time was logged for each participant in the treatment session 1-7 (figure 14) (Note that participant A was not present the last day, and could therefore not participate in the session). Fig. 15(a) shows the number of products touched (i.e.

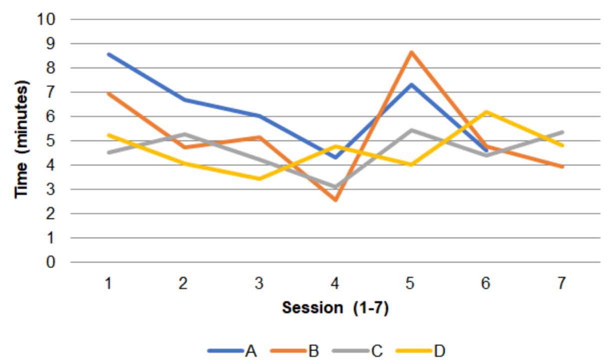
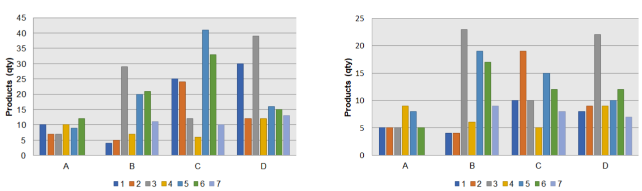


Figure 14: The time (minutes) spent in the VR supermarket each session by each participant

controller touches the products collider) by each participant, and fig. 15(b) shows a number of products grabbed.



(a) Products touched   (b) Products grabbed

Figure 15: (a) Number of times the user touched any a product during the 7 VR training sessions, (b) number of times each user grabbed a product during the 7 VR training session

On average across all sessions participant A touched 9 products and grabbed 6, B touched 14 and grabbed 12, C touched 22 and grabbed 11, and D touched 19 and grabbed 11. Least products touched and grabbed were B in session 1 (i.e. not more than 4 products both touched and grabbed) and most objects touched were C in session 5 (41 products).

5 Discussion

Looking at the results, even though the treatment group did outperform the control group in the final assessment, it was not with a significant difference ($P > 0.10$) when considering the raw difference in difference estimate with no independent variables as predictors. However, many factors could hypothetically have contributed to the results. Several variations in conditions, ranging from the broad diversity of the ASD spectrum to the possible unavailability or limited accessibility of products, could be accounted for variations in performance. Additionally, the given search tasks (shopping lists) might have varied in the degree of difficulty, although sharing rudimentary qualities, requiring participants to walk at different speeds and distances. All these factors lead to the realization that task conditions, let alone the groups themselves, are not mutually comparable unless the same task-based evaluation is followed for all participants, or repeatedly measured until all combinations of participants and shopping lists are carried out.

To account for these factors, we assumed that variations in task completion time can be explained by other measured independent variables such as effectiveness, approximated travel distance (traverse), and increased difficulty (i.e., the occurrence of search task elements in areas estimated to increased completion time such as the vegetable section). Specifically, we weighed these independent variables as predictors in the Difference in Difference estimate as four explanatory assumptions that affected performance:

1. Individual task effectiveness
2. The addition of self-reported required help to achieve task completion
3. The addition of an approximated crude distance-to-travel estimate (traverse-estimate)
4. Binary traverse-estimate of difficulty based on products found in the *Green Section* of the supermarket.

Significant results were found in 3 of the 4 predictions: in scenario 2 ($p < 0.1$), scenario 3 ($p < 0.1$), and scenario 4 ($p < 0.05$). If we accept the assumption of the best resulting scenario, that these factors contributed to TCT, we can accept the alternative hypothesis that a significant difference was found between the treatment group and the control group at a 5% level of significance ($p = 0.05$). However, although the factors discussed could perhaps explain fluctuations in the results, the evidence is far from valid enough be conclusive. Overall, the small sample size results in a wide confidence interval, that prevents us from making strong assumptions based on the Difference in Difference alone. Therefore all the other factors measured, including individual performances and self-reported subjective opinions and that of the teachers must be considered as well.

Considering effectiveness, the results indicate that the treatment group had constant perfect score during both the baseline and post-treatment assessments. On the contrary, the control group effectiveness decreased from 97% in the baseline assessment to 91% in the final assessment. Based on this result, it could be hypothesized that having trained regularly with VR simulation could have had helped the participants in the treatment group to retain their ability to find products' locations more accurately and confidently. Finally, the close to perfect score on effectiveness during the pre-assessment evaluation invites the question: is effectiveness in shopping scenarios a suitable independent variable for this study?
The treatment group had a negative development of their the self-reported sense of confidence during the post-treatment assessment. This could be due to the environmental discrepancies between the real supermarket and its VR version. The VR supermarket simulation did not include any other human shoppers, since the implementation of several realistic virtual shoppers would have been too demanding in the context of this project. Furthermore, implementing virtual shoppers has the potential of inducing uncanny characters that might have intimidated the users.

As reported by the teacher, participant **D** has always had low confidence when visiting public environments. The participant **D** was observed to need help when going to the gym, or to the supermarket; where he never wanted to go alone. This was confirmed during the baseline evaluation, when **D** was observed to explore the supermarket together with a teacher who confirmed that the participant explicitly asked them to be accompanied otherwise he would not enter. Additionally, **D** was observed to show discomfort throughout the baseline evaluation. This pattern in the participant behavior was broken during the post-treatment assessment. Observations and teachers notes reported that the participant decided to approach the supermarket exploration alone during the final assessment. This result can be considered extremely positive, also considering that **D**'s effectiveness was 100% in the real supermarket. The participant has been observed to be able to overcome obstacles without the help of teachers; he has only been noted to ask another participant for help in finding one product. Deficits in self-initiated question asking have been described as one of the main characteristics of autism spectrum disorder [37] [33] [13], contributing to their reduced abilities with everyday living skills. By asking questions about the location of the products, participant **D** illustrated a positive development in his ability to approach others for help. Participant **D**'s positive development can also be observed during the VR training period; satisfaction and confidence levels raised to very positive values by the last training session.

Considering the VR training specifically, it can be observed that participants constantly reported high levels of self-confidence. The decline during the last two VR training sessions can be explained by the fact that the teacher tried to push the participants towards the boundaries of their comfort zone by temporarily negating access to the compass. Mentioning one participant, the teacher said that he asked for compass, but he was told to look at the shelves. Afterwards, the participant found the product he was searching for himself. Such approach can justify the final increase in self-reported and teacher-reported satisfaction; it can be assumed that the achievement of finding a product without the use of hints could boost a participant's self-confidence. Similarly to confidence, satisfaction saw a clear decline when the teacher reduced the access to hint, resulting in an unsettling experience for the participants.

Can skills taught in a virtual supermarket be transferred to a real context?

The aim of the repeated usage of the system was to allow low road transfer [28] of adaptive shopping skills from VR to real shopping experience. This was achieved through the repetitions of product-search processes in the safety of the simulated environment [5], and stripping the environment from additional obstructive and intimidating elements, allowing the participants to focus on the improvement of their skills [21].

The Difference in Difference analysis showed some significant difference in the performance of treatment over control groups, while still being far from valid due to small sample size. A peculiar outcome of this study, however, is participant **D**'s behavior, which was outstandingly different between baseline and final evaluation (i.e., approached the shopping alone). Based on this analysis, we confidently confirm the possibility of near transfer from the virtual to the real experiences, although it might be covert to the participants. Other studies involving VR intervention for individuals diagnosed with ASD also report variability in their participant's ability to transform knowledge from the virtual- to the real world environment. Pearson et al. [20] study on virtual environment intervention for teaching social understanding to adolescents with ASD showed an increased understanding of social conventions in three out of their six participants, while Josman et al. [14] VE street-crossing intervention only showed a transfer of knowledge for three out of their six participants. This heterogeneity of responses to VR interventions in participants diagnosed with ASD calls for future research that should take the differences in characteristics in this population into account.

Lastly, with this study has not been possible to quantitatively assess participants' retention of the newly transferred skills, and it could be interesting to assess whether results improve with a longer rest period between training sessions, and follow-up on the maintenance of such skills.

The self-reported satisfaction declined 35% during the experiment, and would likely have continued to decline. Participants simply were not stimulated enough, and the lack of motivation could ultimately cause them to deny training with the VR simulation. During the final evaluation session, the teacher remarked this, and suggested that some elements to keep the children motivated should be implemented if they were to continue training in VR. The teacher suggested individual competitive scores as a tool to motivate them to perform better over time, trying to improve their previous scores.

6 Conclusion

In this paper, we proposed and evaluated a head-mounted display based VR training simulation designed to rehearse DLS of shopping. While the between group measurement of pre- and post-treatment evaluations in a real supermarket did provide some significant results, the results cannot be generalized due to limited number of participants. Despite this, one of the participants who usually is uncomfortable with shopping scenarios and who refused to enter the supermarket alone during the pre-assessment accomplished the post-treatment shopping assessment with perfect results. The study indicates some positive effects of a head-mounted display based VR simulation to train DLS of individuals diagnosed with ASD, however further research is needed to measure the long-term effects of such intervention. In future research, we will investigate other forms of VR-based DLS training for children diagnosed with ASD, including VR experiences captured using 360 cameras. We will also investigate how to deliver personalized training to reflect the different nuances of autism disorders.

Acknowledgments

The authors wish to thank the teacher Claes Kiilsgaard, students and parents from the Valhøj school for their participation, feedback and support during this study. Additionally, we would like to thank Rødovre municipality for financing the VR equipments used in this study. Finally, huge thanks to Anastassia Andreasen and Razvan Paisa for their help with observations during the baseline- and post treatment assessments.

References

[1] K. D. Allen, D. P. Wallace, D. Renes, S. L. Bowen, and R. V. Burke. Use of video modeling to teach vocational skills to adolescents and young adults with autism spectrum disorders. *Education and Treatment of Children*, 33(3):339–349, 2010.

[2] A. P. Association et al. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[3] G. Baird, E. Simonoff, A. Pickles, S. Chandler, T. Loucas, D. Meldrum, and T. Charman. Prevalence of disorders of the autism spectrum in a population cohort of children in south thames: the special needs and autism project (snap). *The lancet*, 368(9531):210–215, 2006.

[4] V. H. Bal, S.-H. Kim, D. Cheong, and C. Lord. Daily living skills in individuals with autism spectrum disorder from 2 to 21 years of age. *Autism*, 19(7):774–784, 2015.

[5] C. Bossard, G. Kermarrec, C. Buche, and J. Tisseau. Transfer of learning in virtual environments: a new challenge? *Virtual Reality*, 12(3):151–161, 2008.

[6] T. Brugha, S. Cooper, S. McManus, S. Purdon, J. Smith, F. Scott, N. Spiers, and F. Tyrer. Estimating the prevalence of autism spectrum conditions in adults. *London, England: The NHS Information Centre*, 2012.

[7] D. Card and O. Ashenfelter. Using the longitudinal structure of earnings to estimate the effect of training programs. 1984.

[8] N. R. Council et al. *Educating children with autism*. National Academies Press, 2001.

[9] D. M. N. S. Y. Developmental, . P. Investigators, et al. Prevalence of autism spectrum disorder among children aged 8 years-autism and developmental disabilities monitoring network, 11 sites, united states, 2010. *Morbidity and mortality weekly report. Surveillance summaries (Washington, DC: 2002)*, 63(2):1, 2014.

[10] E. Fombonne. Epidemiology of pervasive developmental disorders. *Pediatric research*, 65(6):591–598, 2009.

[11] M. L. Ganz. The lifetime distribution of the incremental societal costs of autism. *Archives of pediatrics & adolescent medicine*, 161(4):343–349, 2007.

[12] M. Gellevij, H. Van Der Meij, T. De Jong, and J. Pieters. Multimodal versus unimodal instruction in a complex learning context. *The Journal of Experimental Education*, 70(3):215–239, 2002.

[13] M. Hauck, D. Fein, L. Waterhouse, and C. Feinstein. Social initiations by autistic children to adults and other children. *Journal of autism and developmental disorders*, 25(6):579–595, 1995.

[14] N. Josman, H. M. Ben-Chaim, S. Friedrich, and P. L. Weiss. Effectiveness of virtual reality for teaching street-crossing skills to children and adolescents with autism. *International Journal on Disability and Human Development*, 7(1):49–56, 2008.

[15] M. R. Kandalaft, N. Didehbani, D. C. Krawczyk, T. T. Allen, and S. B. Chapman. Virtual reality social cognition training for young adults with high-functioning autism. *Journal of autism and developmental disorders*, pp. 1–11, 2013.

[16] M. J. Maenner, L. E. Smith, J. Hong, R. Makuch, J. S. Greenberg, and M. R. Mailick. Evaluation of an activities of daily living scale for adolescents and adults with developmental disabilities. *Disability and Health Journal*, 6(1):8–17, 2013.

[17] M. Maskey, J. Lowry, J. Rodgers, H. McConachie, and J. R. Parr. Reducing specific phobia/fear in young people with autism spectrum disorders (asds) through a virtual reality environment intervention. *PloS one*, 9(7):e100374, 2014.

[18] J. L. Matson. *Applied behavior analysis for children with autism spectrum disorders*. Springer, 2009.

[19] J. C. McPartland, K. Law, and G. Dawson. Autism spectrum disorder. *Encyclopedia of mental health*, pp. 124–130, 2016.

[20] P. Mitchell, S. Parsons, and A. Leonard. Using virtual environments for teaching social understanding to 6 adolescents with autistic spectrum disorders. *Journal of autism and developmental disorders*, 37(3):589–600, 2007.

[21] R. Moreno and R. E. Mayer. Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of educational psychology*, 91(2):358, 1999.

[22] N. Newbutt, C. Sung, H.-J. Kuo, M. J. Leahy, C.-C. Lin, and B. Tong. Brief report: A pilot study of the use of a virtual reality headset in autism populations. *Journal of autism and developmental disorders*, 46(9):3166–3176, 2016.

[23] S. Oviatt. User-centered modeling and evaluation of multimodal interfaces. *Proceedings of the IEEE*, 91(9):1457–1468, 2003.

[24] S. Oviatt, R. Coulston, and R. Lunsford. When do we interact multimodally?: cognitive load and multimodal communication patterns. In *Proceedings of the 6th international conference on Multimodal interfaces*, pp. 129–136. ACM, 2004.

[25] A. Palmen, R. Didden, and R. Lang. A systematic review of behavioral intervention research on adaptive skill building in high-functioning young adults with autism spectrum disorder. *Research in Autism Spectrum Disorders*, 6(2):602–617, 2012.

[26] C. J. Park, G. W. Yelland, J. R. Taffe, and K. M. Gray. Brief report: The relationship between language skills, adaptive behavior, and emotional and behavior problems in pre-schoolers with autism. *Journal of Autism and Developmental Disorders*, pp. 1–6, 2012.

[27] S. Parsons. Authenticity in virtual reality for assessment and intervention in autism: A conceptual review. *Educational Research Review*, 19:138–157, 2016.

[28] D. N. Perkins and G. Salomon. Teaching for transfer. *Educational leadership*, 46(1):22–32, 1988.

[29] A. J. Russell, C. M. Murphy, E. Wilson, N. Gillan, C. Brown, D. M. Robertson, M. C. Craig, Q. Deeley, J. Zinkstok, K. Johnston, et al. The mental health of individuals referred for assessment of autism spectrum disorder in adulthood: a clinic report. *Autism*, 20(5):623–627, 2016.

[30] J. Sauro and J. R. Lewis. *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann, 2016.

[31] T. Self, R. R. Scudder, G. Weheba, and D. Crumrine. A virtual approach to teaching safety skills to children with autism spectrum disorder. *Topics in Language Disorders*, 27(3):242–253, 2007.

[32] L. E. Smith, M. J. Maenner, and M. M. Seltzer. Developmental trajectories in adolescents and adults with autism: The case of daily living skills. *Journal of the American Academy of Child & Adolescent Psychiatry*, 51(6):622–631, 2012.

[33] W. L. Stone and L. M. Caro-Martinez. Naturalistic observations of spontaneous communication in autistic children. *Journal of Autism and Developmental Disorders*, 20(4):437–453, 1990.

[34] D. Strickland. Virtual reality for the treatment of autism. *Studies in health technology and informatics*, pp. 81–86, 1997.

[35] P. Szatmari, S. Bryson, E. Duku, L. Vaccarella, L. Zwaigenbaum, T. Bennett, and M. H. Boyle. Similar developmental trajectories in autism and asperger syndrome: from early childhood to adolescence. *Journal of Child Psychology and Psychiatry*, 50(12):1459–1467, 2009.

[36] J. Ward-Horner and P. Sturmey. Component analysis of behavior skills training in functional analysis. *Behavioral Interventions*, 27(2):75–92, 2012.

[37] A. M. Wetherby, C. A. Prutting, et al. Profiles of communicative and cognitive-social abilities in autistic children. *Journal of speech and hearing research*, 27(3):364–377, 1984.

[38] J. H. Williams, D. W. Massaro, N. J. Peel, A. Bosseler, and T. Suddendorf. Visual–auditory integration during speech imitation in autism. *Research in developmental disabilities*, 25(6):559–575, 2004.

[39] M. Wolf, T. Risley, M. Johnston, F. Harris, and E. Allen. Application of operant conditioning procedures to the behavior problems of an autistic child: A follow-up and extension. *Behaviour Research and Therapy*, 5(2):103–111, 1967.