Computers & Education: X Reality 2 (2023) 100013

Contents lists available at ScienceDirect

Computers & Education: X Reality

journal homepage: www.journals.elsevier.com/computers-and-education-x-reality

Programming for generalization: Confronting known challenges in the design of virtual reality interventions for autistic users

Matthew Schmidt [a,*], Noah Glaser [b], Carla Schmidt [e], Rachel Kaplan [a], Heath Palmer [c], Sue Cobb [d]

[a] University of Florida, United States
[b] University of Missouri, United States
[c] University of Cincinnati, United States
[d] University of Nottingham, United States
[e] Bayada Home Health Care, United States

ARTICLE INFO

Keywords:
Virtual reality
Autism
Generalization
Daily-living skills
Public transportation

ABSTRACT

This study explored how to program for generalization using a fully immersive virtual reality (VR) intervention for teaching transportation skills to autistic adults related to using a university shuttle. Using multi-methods, this research sought to identify similarities and differences in behavior between the VR environment and the real-world, participants' perceptions of telepresence in the VR environment, and how participants characterized similarities and differences between the VR environment and the real-world. Male adult participants (N = 6) completed training and then engaged in two VR-based training sessions of increasing complexity, after which they enacted what was learned in the real-world. Fidelity of implementation was high across VR and real-world sessions and no significant differences were found in behaviors between the VR and real-world sessions, providing evidence for skills generalization from the contrived VR setting to the naturalistic real-world setting. Participants reported high perceptions of telepresence (e.g., being there) and social presence (e.g., being there with others), and qualitative evidence suggests they made connections between the virtual world and real-world. Implications and directions for future research are discussed.

1. Introduction

This research report details the results of a longitudinal design-based study that aimed to design, develop, and evaluate a virtual reality (VR) learning intervention called Virtuoso for autistic adults. The VR training was designed to help autistic adults learn how to use public transportation services in a safe, effective, and controllable manner (Glaser and Schmidt, 2018). The study highlights the challenges associated with designing VR for autistic users, specifically the programming for generalization of skills learned in the VR environment to real-world public transportation. The report discusses the potential benefits of VR technology for autism interventions, including the potential for skills learned in VR to generalize to real-world situations. The report provides a comprehensive overview of the research and its findings, offering valuable insights into the potential of VR technology to support the learning needs of autistic individuals (Glaser & Schmidt, 2021; Schmidt et al., 2019; Schmidt & Glaser, 2021a, 2021b).

In the following sections, we present autism, a neurodevelopmental condition affecting 2% of people worldwide, and associated challenges that can impact education, employment, and quality of life. We then describe the potential of public transportation training to increase independence, employment opportunities, and social engagement for autistic people, as well as the barriers to accessing such training, including safety risks and a lack of available training options. We present the use of virtual reality technology for autism as particularly promising given its potential for generalization of skills to the real-world. Our intervention, entitled Virtuoso, aims to address the challenges of public transportation training by offering a safe and controlled training environment for autistic adults. The Virtuoso public transportation training intervention is detailed, with a discussion of the various design elements that were incorporated to enhance the learning experience and promote generalization.

* Corresponding author.
E-mail address: matthew.schmidt@ufl.edu (M. Schmidt).

https://doi.org/10.1016/j.cexr.2023.100013
Received 10 November 2022; Received in revised form 23 February 2023; Accepted 25 February 2023

2. Literature review

2.1. Autism and public transportation

Autism is defined as a neurological spectrum condition with individuals on the autism spectrum often experiencing challenges related to social communication and interaction, and restricted or repetitive behaviors or interests (American Psychiatric Association, 2013). Those who are autistic tend to have a broad range of connected conditions that vary tremendously in terms of types and severity, as autism frequently manifests alongside cognitive disabilities and conditions such as epilepsy, anxiety, attention deficit hyperactivity disorder, sensory integration challenges, cognitive impairments, and obsessive-compulsive disorder (Brondino et al., 2019; Müller et al., 2008; Sharma et al., 2018). Although a range of services and intervention modalities exist to support autistic people, autism is a lifelong neurodevelopmental condition. Approximately 2% of people worldwide are affected by autism (Baron-Cohen, 2017), with some noting that this figure may be higher, as improvements in diagnostics and the celebration of the neurodiversity movement has helped to bring attention to individuals that may have been left out of past diagnoses (Brown et al., 2021).

Challenges associated with autism often have deleterious effects on an individuals' educational performance, vocational success, and quality of life (Taylor et al., 2014). Long-term social inclusion and employment outcomes for autistic people are poor (Coleman & Adams, 2018), with some estimates suggesting that between 50% and 75% of autistic adults are unemployed or underemployed (Eaves & Ho, 2008). Vocational and educational activities are critical factors that influence independence and quality of life (Lorenz et al., 2016); however, it can be difficult to access these opportunities without transportation (Scott et al., 2015). Autistic people are unlikely to use public transportation (Lubin & Brooks, 2021) and tend to rely on their parents for travel (Deka et al., 2016). Training on how to use public transportation has the potential to increase employment opportunities, encourage community and social engagement, and promote greater levels of independence (Taylor et al., 2014), and is desired by the autistic community, particularly for transition-aged youth (Deka et al., 2016). However, a number of barriers exist, including a lack of time and resources for teaching safety skills needed for independent travel and a lack of training options for transition programs (Feeley et al., 2015).

Despite potential benefits of public transportation training for autistic people, training opportunities are virtually non-existent, and few studies address this issue (Rezae et al., 2020). This could be due, in part, to the inherent risks associated with public transportation training (Feeley et al., 2015). In light of these risks, training providers must decide whether to train in naturalistic or contrived settings (Dixon et al., 2010). From a research perspective, contrived settings allow for more experimental control of the various risks that are present in naturalistic settings. However, skills learned in contrived settings must be programmed to generalize to naturalistic settings because "Training does not serve its purpose if the targeted skill fails to generalize" (Dixon et al., 2010, p. 632).

2.2. Autism and virtual reality technology

Technology supports have received great attention for autism interventions, as they are believed to align with needs and strengths of autistic people (Glaser & Schmidt, 2021; Grynszpan et al., 2014). Interest in developing VR interventions for autistic people has been steadily growing for over 20 years (Aresti-Bartolome & Garcia-Zapirain, 2014), beginning with preliminary research that examined the acceptability of VR equipment and potential learning effects with autistic children (Strickland, 1996, 1997). VR is a computer-generated model of reality in which a user can interact with and get information from the model through the use of ordinary human senses (Hale & Stanney, 2014). It is typically associated with three-dimensional environments that are capable of inducing psychological sensations of feeling 'present' commonly referred to as telepresence (Slater et al., 2009; Steuer, 1992). Research suggests potential learning benefits such as predictability, structure, customizable task complexity, control, realism, immersion, automation of feedback, assessment, and reinforcement (Bozgeyikli et al., 2018).

Perhaps the most compelling benefit of VR is the potential for skills learned in VR to generalize to the real-world (Bozgeyikli et al., 2018; Parsons et al., 2006). Generalization is believed to occur when experiences within VR contexts are highly realistic, leading users to behave in VR similarly to how they would in the real world. Realism plays an integral role in helping to promote perceptions of telepresence (e.g. Jung & Lindeman, 2021). The combination of realism and telepresence is seen as vital for exerting social influence in a VR environment and promoting behavioral change (Blascovich, 2002). Therefore, it is often assumed that if a VR environment is made to be realistic, then what is learned in VR will more readily generalize to the real-world (Dalgarno & Lee, 2010; Parsons, 2016). However, although the potential for generalization is seen as promising, evidence supporting generalization from VR to the real-world has just begun to emerge (Bozgeyikli et al., 2018; Dixon et al., 2019).

Given the affordances of the technology, VR can be used to simulate real-life transportation experiences, such as riding, in a controlled and safe environment. This can help individuals with autism develop the skills and confidence needed to successfully navigate transportation systems in the real world. In this study we present how a team of researchers created and evaluated Virtuoso, a novel VR intervention that was designed to support autistic individuals in their transportation needs. The questions that guided this research are provided in Table 1.

2.3. Project description

Virtuoso is a suite of VR technologies that was developed in collaboration with autistic adults enrolled in a day program and their service providers at a large Midwestern university. Virtuoso was developed to provide the day program with a formalized immersive learning intervention to teach skills associated with using a publicly-available university shuttle in and around the university campus (Glaser et al., 2022; Schmidt et al., 2019; Schmidt & Glaser, 2021). The ability to access and use public transportation is a critical aspect of independent living (Felce, 1997), as it provides much-needed access to medical, vocational, and community opportunities (Shier et al., 2009).

Prior to the development of Virtuoso, the day program lacked a formal and systematized public transportation routine. Given that associates of the day program needed to use public transportation on a regular basis to travel to vocational training sites and engage in day-to-day programming around campus, a number of issues arose, the most poignant of which had to do with the inherent risks associated with transportation training. Therefore, the use of immersive technologies was seen as a promising solution to this real-world problem. However, we were unable to locate any off-the-shelf public transportation VR tools. Therefore, we opted to develop Virtuoso. This decision was predicated on the following justifications.

Table 1
Research questions that guided the study.

RQ Number	Research Question
RQ1	How similar or different was behavior in the Virtuoso-VR environment relative to behavior in the real-world?
RQ2	To what extent did participants perceive that they were present in the Virtuoso-VR environment?
RQ3	How did participants characterize similarities and differences between the Virtuoso-VR environment and the real-world?

1. Lack of formal training: Prior to the development of Virtuoso, there was no formal or systematic shuttle training available, leading to challenges and risks for participants.
2. Real-world risks: Real-world public transportation training exposes participants to a variety of risks that are not feasible to manage in a live environment.
3. Safe and controlled scenarios: VR allows training to be experienced safely and repeatedly in controlled scenarios, reducing the risks associated with real-world training.
4. Synthesis of important skills: VR also allows for the synthesis of important subordinate skills, such as interpreting a map or schedule, which can be useful across a range of vocational scenarios.
5. Access to vocational training: Day program participants need to use public transportation to travel to vocational training sites, and a systematic approach to shuttle training using VR would provide them with the skills and confidence necessary to do so.

Virtuoso's learning design included a multi-tiered, four step approach which progressed from simple to more complex skills. These were learned using both low- and high-tech VR hardware and software. Our pedagogical model progressed thus: (1) introduction to the skill using a digital social narrative, (2) modeling of the skill using a spherical, 360-degree videos, (3) rehearsal of the skill using fully-immersive head-mounted display-based VR, and (4) real-world enactment of the skill (Schmidt & Glaser, 2021). Substantial effort was made to maintain parity between all four aspects of our pedagogical model so as to create digital twins (Fig. 1). During the introduction to the skill stage, a trained staff member introduced Virtuoso by sharing a comic strip style social narrative on a tablet. In the second stage learners were presented with four 360-degree videos that modeled the four discrete skills of catching the shuttle and was designed using best practices for video modeling (Schmidt et al., 2019).These videos were experienced by users in both the lower-immersion Google Cardboard head-mounted display (HMD) and the fully-immersive HTC Vive or Oculus Rift HMDs (see Fig. 1). In the third stage, learners were led through the Virtuoso-VR environment by an online guide (OG), a trained clinician in the form of an avatar who presented instructional content, provided opportunities to rehearse the previously reviewed skills, and prompted participants.

Virtuoso-VR consisted of two virtual environments, Virtuoso VR Level (VRL1) and Virtuoso VR Level 2 (VRL2). Between the first and second "levels" of the Virtuoso-VR training, the VR environment was made increasingly more realistic and additional diversity was added to the training. The additional instructional teaching prompts used in VRL1 were later faded, allowing VRL2 to be experienced as a more naturalistic environment. For example, in VRL1, red dotted lines were used to direct the participant on the correct path to walk, areas of the map were boxed and highlighted to draw attention to important information, and additional directional signs were used to indicate walking or shuttle boarding prompts. These were removed in VRL2, allowing participants to independently practice navigating important routes and encounter necessary information with reduced environmental prompts. Further additional environmental details such as bushes, trees, fire hydrants, and more notably non-playable characters were added to make the campus more realistic. The fourth and final stage took place in the real-world, providing participants an opportunity to perform the skills they had previously learned and practiced virtually. Ultimately, this pedagogical model sought to promote synthesis of important subordinate skills (e.g., interpreting a map, interpreting a schedule) using a range of evidence-based practices, such as system of least prompts (Doyle et al., 1988), prompt fading (Cengher et al., 2018), video modeling (McCoy & Hermansen, 2007), and incorporation of established generalization heuristics (Stokes & Osnes, 2016a).

2.4. Programming for generalization in virtuoso

Generalization of skills refers to behavioral change that takes place across time, persons, and settings (Stokes & Baer, 1977). A challenge with autism interventions is establishing conditions such that the skills acquired in contrived settings will generalize outside of these controlled environments (Dixon et al., 2019). Generalization occurs when a learner displays a behavior in a context that differs from that of the instruction, without direct training in the different setting or with a varied response. Systematic planning for generalization includes a selection of target behaviors that can elicit natural reinforcement in the varying situations in which the behaviors are desired. A nine-part classification system related to the assessment and programming of generalization was developed to describe findings from an analysis of 120 studies (Stokes & Baer, 1977). This system was further refined into three categories recommended to increase the probability of an intervention's generalization and maintenance: (1) exploit current functional contingencies, (2) train diversely (Fig. 2), and (3) incorporate functional mediators (Stokes & Osnes, 2016b). How these heuristics were incorporated into the design of Virtuoso is presented in Table 2 below.

3. Methodology

The purpose of this multi-methods research was to investigate generalization between the contrived VR setting and the naturalistic real-world setting with autistic adults enrolled in a day program at a large Midwestern university.

4. Participants

Participants were recruited using purposive sampling (Etikan et al., 2016). Inclusion criteria were: (1) confirmed autism diagnosis, (2) ability to verbally communicate, (3) level of cognition, (4) ability to engage in a

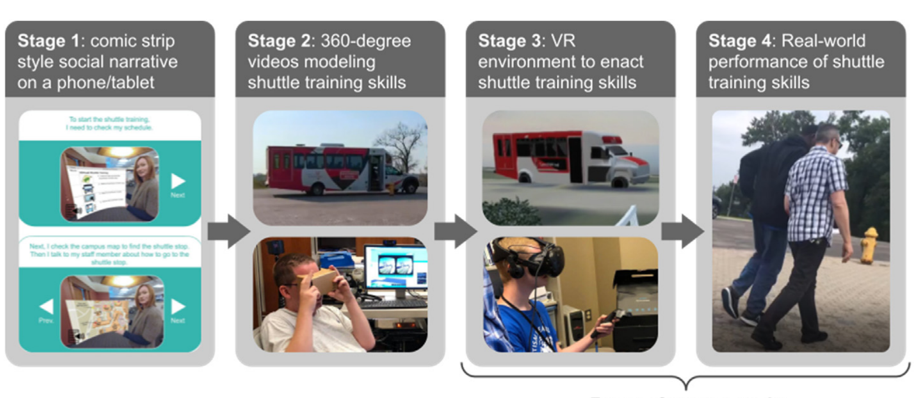

Fig. 1. An overview of Virtuoso's multi-tiered, four step pedagogical approach and technologies used in each step.

Fig. 2. Training diversely in the “Walk to Shuttle” step of Virtuoso training.

30-min task, (5) expressed interest in taking part in the project, and (6) at least 18 years old. Participants were excluded if they had a history of significant behavioral challenges such as physical aggression or were unable to verbally communicate. Based on these criteria and recommendations of the program's director, a total of seven participants were identified for inclusion, from which six agreed to participate. All six ASD participants were male, with an average age of 26.6 years old. Studies involving autistic participants tend to have small sample sizes (Palmen et al., 2004) because it is a low-incidence disability (Baron-Cohen, 2017). Therefore, the sample size of this work is consistent with other autism research in the field. Further, that all participants were male is expected, given that ASD is far more prevalent in males than females (Lord et al., 1982), with a ratio of around 3:1 (Loomes et al., 2017).

A full breakdown of ASD participants is provided in Table 3, in which we provide details from the Peabody Picture Vocabulary Test (PPVT), the Social Responsiveness Scale (SRS), and the Behavior Rating Inventory of Executive Function (BRIEF). We are including these measures as is common in ASD research (Kwok et al., 2015; Chan et al., 2017). The PPVT measures one's comprehension of single-word vocabulary words and is often used as an early screening measure for the identification of developmental conditions. The scores of the PPVT are standardized and an age equivalent is calculated. The SRS is used to identify the presence and severity of social challenges. The overall T-Score is a quantified value of social challenge and severity. The BRIEF is used to assess executive function and is a commonly used measure to assess the abilities of a broad range of people who have learning disabilities and other neurological conditions.

4.1. Data collection and analysis

Study procedures took place in the office space of the principal investigator on the university campus. Informed consent was obtained by a trained researcher prior to the beginning of the study. In cases where a participant was under the legal guardianship of another, consent was obtained by the guardian and assent was obtained from the participant. Participants engaged in three 1-h research sessions (see Fig. 3) one week apart. After each session, participants were asked if they would like to return to continue the study.

Quantitative and qualitative multi-methods were used to collect data, as summarized in Table 4. Given that there is no established protocol for testing the generalization of learning skills, researchers must determine their own approach to generalization testing, which may involve modifying the content, context, setting, or a combination of these elements (Khowaja et al., 2020). Therefore, methods appropriate for evaluating generalization were identified by the research team from the autism and VR literature (Dixon et al., 2019; Massey & Wheeler, 2000; Palmen & Didden, 2012). These included (1) percentage of task completion (Smith et al., 2016), (2) total time on task (Palmen & Didden, 2012), and (3) operationalized behaviors (Osgood, 2022) for prompts and responses (see Table 5).

4.2. Data sources

Multi-methods utilizing both quantitative and qualitative data sources were used, as summarized in Table 4 and detailed in the following sections.

Temple Presence Inventory (TPI). The Temple Presence Inventory (TPI) is a no-cost, validated inventory for measuring dimensions of telepresence within many types of media content (Lombard et al., 2009). The TPI was developed to assess the degree to which people experience a sense of presence in virtual reality environments and measures several aspects of telepresence, such as spatial presence (e.g., How much did it seem as if the objects and people you saw/heard had come to the place you were?), social presence (e.g., To what extent did you feel you could interact with the person or people you saw/heard?), mental immersion (e.g., To what extent did you experience a sensation of reality?), social realism (e.g., The events I saw/heard could occur in the real world.), among others. Upon the completion of each VR scenario, participants completed the TPI. Data from their responses were entered into a spreadsheet for later analysis.

Virtuoso-VR Session Recordings. During each VR-training session, Open Broadcaster Studio (https://obsproject.com/) was used to capture screen, webcam, and audio recordings of participants and the OG. Videos from participants and the OG's perspectives were merged into a single video (see Fig. 4), allowing researchers to observe and analyze session activities simultaneously. A behavioral coding system was created, and codes were applied to these videos using the V-Note computer-assisted qualitative data analysis software (Bremig, LLC of PA).

Field Notes. A trained research assistant took notes of observed activities that were scanned and digitized, stored in a secure cloud sharing platform, transcribed, and then organized in a text document for further analysis. Notes were broad in nature with a focus on documenting details pertinent to our research questions. In particular, notes were taken when the research assistant observed behaviors and communication that conveyed details about the realism of digital assets.

Semi-Structured Interviews. After each session, a semi-structured interview protocol was used to ask participants about their experience with using the VR head-mounted displays, how they perceived the VR

Table 2
Generalization heuristics and examples of how these heuristics were enacted in Virtuoso.

Stokes and Osnes' (2016) Generalization Heuristics that were Enacted in Virtuoso	Description of How Heuristics Enacted	Examples
Exploit current functional contingencies. • Use modified consequences to teach new skills. • Utilize consequences found in the natural environment.	• Plan for participants to contact modified and naturally-rewarding consequences throughout training similar to those contacted outside of VR training • Contingencies used in natural state or modified • Occurrences of generalization reinforced throughout training • As target behaviors learned, verbal and gestural praise provided • Praise faded over time	• Participants earn a badge of achievement by entering the shuttle • Waiting for shuttle contacts reinforcement of getting on the shuttle for a ride • Completion of tasks on the provided schedule provides naturally occurring consequence
Train diversely. • Repeat training activities • Use scenarios that evolve in fidelity and complexity over time • Provide training with a variety of real-world exemplars • Fade instructional prompts • Explicitly teach natural antecedents • Teach activities using a variety of methods	• Repeated training activities with increasing demands • Training with variety of real-world exemplars • VR environment that evolves in complexity • Prompt fading	• Multi-stage process of (1) observing a digital social narrative, (2) watching 360-degree videos of discrete skills needed, (3) rehearsing skills in a VR environment of increasing complexity and difficulty (Fig. 2), (4) demonstrating learned skills in naturalistic environment • Heavily scaffolded instructional teaching prompts used in VRL1 (e.g., dotted lines, signage, & cues) removed in VRL2
Incorporate functional mediators. • Incorporate real-world visual stimuli • Incorporate common, salient physical stimuli • Incorporate self-mediated stimuli	• Practicing identifying physical stimuli by looking at schedules, finding stops and buildings on a map, walking to shuttle stops, standing at stop waiting for shuttle • Represent real-world stimuli in VR environment i.e., buildings on map, street signs, shuttle stops, etc. • Learning to use self-mediated stimuli like maps, schedules, and signs.	• OG pointed out visual stimuli (e.g., signs or a fountain) that could help participants recall the path to the shuttle stop • OG asked questions like "where are we now" to help the participant connect the VR environment to the real-world • OG and learner avatars boarded the shuttle together, with the OG reminding participants to "look for an empty seat and stay together" on the virtual shuttle.

environment, and what their general thoughts and impressions were. Interviews were transcribed for later analysis.

Generalization Session Recordings. Generalization activities were video recorded by a trained researcher. These videos were stored in a secure cloud sharing platform for later analysis.

4.3. Analysis for RQ1: how similar or different was behavior in the VR environment relative to behavior in the real-world?

Behavioral coding using coding scheme. V-Note Pro was used to code the video mashups from screen, webcam, and audio recordings (Fig. 4). Researchers created a behavioral coding scheme, then coded prompt and response behaviors directly onto a timeline of the video. Researchers compared OG and participant behaviors across sessions**.** Durations were first coded, and then frequency of the OG's prompts and participants' responses were coded using a behavioral coding system that was created and refined by the research team (see Table 5). These coded data were exported for further analysis.

Coder calibration and interobserver agreement analysis were performed. Coders were trained until at least 80% simple agreement was achieved (Ledford & Gast, 2010). For inter-observer agreement, (IOA), 30.4% of all videos were coded by two coders, and Kappa coefficients were calculated, averaging 0.83 (min = 0.71; max = 1.0; SD = 0.096). Given that 0.6–0.8 is a reliable Kappa metric for IOA (Kraemer et al., 2012), IOA was determined to be acceptable.

Video analysis using V-Note Pro. Precise measures of total time on task and percentage of task completion were also established using V-Note Pro. To determine the total time on task, researchers applied duration-based codes using V-Note. Two independent coders coded tasks from start to end. Results between coders were nearly identical (off by milliseconds). Values were input into a spreadsheet and later used to calculate descriptive statistics.

One way within subjects ANOVA. Within-subjects 1-way ANOVA was used. Participants were measured multiple times on the dependent variable (total time to complete each task and each session) under different levels of the independent variable (different stages of the shuttle training program). The goal of the ANOVA was to determine if there was a significant difference in the mean total time to complete each task and each session across the different stages of the shuttle training program. The Shapiro-Wilk test was used to check the normality of the data for each task. If the data was normally distributed, the 1-way within-subjects ANOVA was performed. If the data was not normally distributed, the Friedman rank sum test was used instead.

4.4. Analysis for RQ2: to what extent did participants perceive that they were present in the VR environment?

Temple Presence Inventory. The Temple Presence Inventory (TPI) is a validated inventory for measuring dimensions of telepresence (Lombard et al., 2009). The TPI contains individual questions organized by dimensions of telepresence. Values are calculated by taking the aggregate of the relevant questions on a scale from 1 to 7, ranging from "Not at all" to "Very Much." Data from the TPI were input into a spreadsheet for analysis using methods detailed by the instrument's authors.

Analysis for RQ3: How did participants characterize similarities and differences between the VR environment and the real-world?

Virtuoso VR Session Recordings. Researchers applied a coding scheme within V-Note Pro to identify representative examples and video excerpts in which participants commented on the similarities and differences between the VR environment and the real-world. These codes were classified as "Ecological Validity" statements, and were separated into prompted and unprompted statements (see Table 6). These statements were first identified, then exported for further analysis and synthesis.

Semi-structured Interviews. Upon completion of each VR training session, participants took part in a semi-structured interview, each of which was annotated using thematic analysis. To thematically analyze our data, we started by familiarizing ourselves with the data by reading through the interview notes several times. Next, we began to identify and highlight segments of our data that seemed relevant. This process was completed by two members of the research team to generate preliminary codes. We then met and formalized the themes into a codebook and began the process of charting them to seek the relationships between

Table 3
Participant demographics and scores from the Peabody Picture Vocabulary Test (PPVT), Social Responsiveness Scale (SRS), and Behavior Rating Inventory of Executive Function (BRIEF).

	PPVT			SRS	Behavior Regulation	Meta-cognition	Global Executive
Participant Name and Description	Raw Score	Standard Score	Age Equivalent (y:m)	T-Score & Range	T-Score & Percentile	T-Score & Percentile	T-Score & Percentile
Travis: 29-year-old male diagnosed with ASD, Smith Lemli Opitz Syndrome, attention deficit hyperactivity disorder (ADHD), and auditory processing disorder	196	93	18:11	69; Moderate	88; 98	90; 99	94; 99
Andy: 24-year-old male diagnosed with ASD, anxiety disorder, and attention deficit disorder	161	72	10:11	71; Moderate	93; 99	72; 97	84; 99
Evan: 35-year-old male diagnosed with ASD and an intellectual disability	162	72	11:1	66; Moderate	48; 70	52; 68	50; 66
Kevin: 25-year-old male diagnosed with ASD and anxiety disorder	165	73	11:6	65; Mild	68; 90	63; 82	67; 91
Jonah: 22-year-old male diagnosed with ASD,	65	20	4:1	75; Mild	86; 98	82; 99	88; 99
Keith: 25-year-old male diagnosed with ASD and down syndrome	57	20	3:9	82; Mild	62; 86	98; 99	88; 99

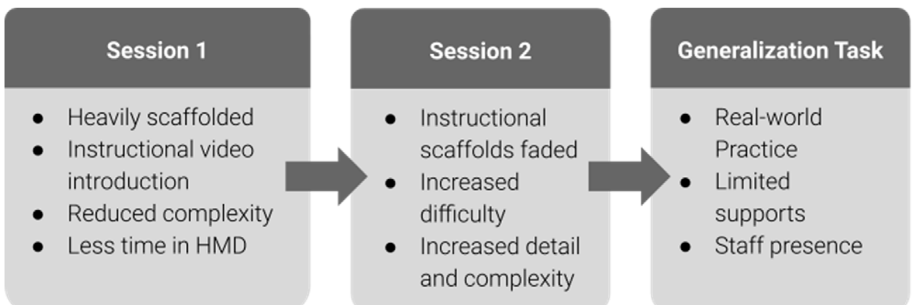

Fig. 3. Activities within each research session.

Table 4
Data analytic approach.

RQ #	Focus of Analysis	Data Sources	Method
RQ1	• Percentage of Task Completion (Smith et al., 2016) • Total Time on Task (Palmen & Didden, 2012) • Comparison of OG prompting and participant response behaviors across sessions (Osgood, 2022)	• Virtuoso VR Session Recordings	• Behavioral coding using coding scheme • Video analysis using V-Note Pro • One way within subjects ANOVA
RQ2	• Participants' perceptions of presence	• Temple Presence Inventory (Lombard et al., 2009)	• Aggregate scores • Descriptives
RQ3	• Participants' characterizations of VR relative to real-world counterparts.	• Virtuoso VR Session Recordings • Field Notes • Semi-structured Interviews	• Thematic analysis • Open and axial coding

Table 5
Prompt and response behavioral codes and operational definitions.

Code	Operational Definition	Examples
Prompt	Any verbal instruction, demand, or feedback from the OG to the participant used to elicit a response	"Where are we right now?" "Stand on the pod."
	OG avatar points to, looks at, or touches an item or area to elicit a response	OG points to pod to indicate which pod participant should step on.
	OG moves the avatar in a way to elicit a response.	OG stands on the pod to show the participant which pod to stand on.
Response	Participant responds to the OG prompt.	OG says, "Read the first step on our schedule." Participant then says, "Look at map and identify the southwest shuttle stop" OG says, "Come over here and get on one of the green pods." Participant then moves the avatar to the green pod.
	Participant provides a verbal response to a prompt.	OG asks, "Where are we?" Participant responds, "On the shuttle."
	Participant's avatar points to, looks at, or touches an item or area in response to a prompt.	OG asks, "Which way is the shuttle stop?" The participant points in the direction of the shuttle stop.
	Participant moves avatar in response to a prompt.	OG says, "Please step on the pod". Participant moves the avatar onto pod.

them. From there, we selected the most important and relevant themes and interpreted and elaborated on them, considering their relevance to our research questions. Themes were organized by comparing and contrasting participant perceptions of VR environment attributes with the real-world.

Field Notes. Field notes were used to triangulate findings from Virtuoso VR session recordings and semi-structured interviews. Field notes were reviewed and annotated using the same coding scheme as for the Virtuoso VR session recordings (Table 6). Field notes were also analyzed using similar thematic analysis techniques as those used for analysis of the semi-structured interviews Themes were organized by comparing and contrasting the participants' perceptions of the VR environment with attributes with the real-world. The triangulation of data from Virtuoso

Fig. 4. Mashup video of (1) participant view of VR space, (2) webcam video of participant in real-world, (3) OG view of VR space, and (4) webcam video of OG in real-world.

Table 6
Prompted and unprompted codes for ecological validity.

Code	Code Description	Sample Quotes
Prompted Ecological Validity (EV) Statements	EV is the extent to which the virtual world is perceived as being similar to the real-world. Coded when EV statements are made in response to an OG question. Participant responds to OG question about EV of VR environment.	OG: "Where are we right now?" Participant: "We're in Teachers' College." OG: "Does this remind you of the program offices?" Participant: "Not really."
Unprompted EV Statements	EV statement in response to environmental stimulus. Participants make comments or remarks about the EV of VR environment without prompting.	Participant: "That's my desk." Participant: "That's [the OG's] office." Participant: "These trees look fake."

VR Session Recordings, semi-structured interviews, and field notes allowed for a more complete and robust understanding of participants' perceptions of ecological validity and provided evidence of multiple perspectives regarding the same phenomenon.

5. Results

5.1. *Results for RQ1: how similar or different was behavior in the VR environment relative to behavior in the real-world?*

5.1.1. *Percentage of task completion*

Analysis of percentage of task completion revealed that in both the VR sessions and the real-world session, all four tasks were completed by all participants (i.e., check daily schedule, check map, walk to shuttle, and wait for shuttle). A single exception to this was for participant Keith in VRL1, when a technology error precluded his ability to complete the wait for shuttle task. Percentage of task completion for all participants in VRL1 was 95.83%, in VRL2 was 100.00%, and in the real-world was 100%.

5.1.2. *Total time on task*

Total time to complete each task and each session was calculated for each participant and then compared using within subjects 1-way ANOVA. The Shapiro-Wilk normality test indicated that data for the following tasks were normally distributed: check schedule ($W = 0.95$, $p = 0.41$), check map ($W = 0.99, p = 0.99$), and walk to shuttle $W = 0.97, p = 0.88$). Hence, a 1-way within subjects analysis was performed on these data. Data for the wait for shuttle task were not normally distributed ($W = 0.74, p = 0.0002$) as were data for total session time ($W = 0.86, p = 0.01$). Therefore, the Friedman rank sum test was used for this data. Findings suggest no significant differences among sessions for time to completion for the following stages: check schedule ($p = 0.06$), check map ($p = 0.15$), walk to shuttle ($p = 0.19$), and wait for shuttle ($p = 0.51$). Comparison of total session time between participants also suggested no significant difference ($p = 0.57$). Given this lack of significant differences, results suggest that total time to complete each individual task was similar for all participants. Further, total time to complete sessions was also similar for all participants.

Descriptive statistics illustrating time to complete sessions and tasks were aggregated across participants (see Table 7). For the check daily schedule task, a decreasing trend in average time to completion was observed, with a difference of 33 s evident between VRL1 and real-world sessions, suggesting a potential carryover effect. A similar trend was evident for the check map task, also with a 33 s difference between VRL1 and real-world sessions. For the walk to shuttle task, VRL1 and VRL2 were identical, although the average to completion was 43 s longer for performing the task in the real-world. For the wait for shuttle task, average times differed from 1 s to 32 s between sessions. On average, total time to completion was fairly similar between sessions, ranging from 7:01 to 7:32.

5.1.3. *Comparison of OG and participant behaviors across sessions*

OG prompting behaviors and Virtuoso participant response behaviors were analyzed across sessions and tasks. Descriptive statistics were calculated and are presented in Table 8.

To compare behaviors between the following sessions, analysis of variance (ANOVA) was performed on: (1) VRL1 and VRL2, (2) VRL1 and real-world, and (3) VRL2 and real-world. Homogeneity of variance was calculated for OG and Virtuoso participants (VP) across conditions, with findings suggesting normal distribution of the data. Generally speaking, behavior across sessions and tasks was similar. However, some differences were observed. For VRL1 and VRL2, significant differences were found in OG behavior for the wait for shuttle task and for VP behavior in the walk to shuttle task. For VRL1 and real-world, significant differences were found in OG and VP behavior for the walk to shuttle task. For VRL2 and real-world, significant differences were found in OG behavior for the walk to shuttle task and for VP behavior in the wait for shuttle task.

Table 7
Descriptive statistics for participant time on tasks and total task time in VRL1, VRL2, and Real-world training.

Task	Sessions											
	VRL1				VRL2				Real-World			
	Mean	Min	Max	SD	Mean	Min	Max	SD	Mean	Min	Max	SD
Check Daily Schedule	1:24	1:13	1:42	0:11	1:04	0:36	1:32	00:19	0:51	0:32	1:09	0:17
Check Map	1:32	0:50	2:20	0:33	1:25	1:02	1:43	0:15	0:59	0:34	1:28	0:19
Walk to Shuttle	2:54	2:05	4:08	0:46	2:54	2:04	4:35	1:01	3:37	2:46	5:28	0:56
Wait for Shuttle	1:35	1:03	3:32	1:05	2:07	0:29	6:52	2:23	1:34	0:50	1:41	0:49
All Tasks	7:11	6:01	7:19	0:52	7:32	4:49	12:58	3:00	7:01	4:58	8:16	1:15

Table 8
Descriptive statistics for OG and Virtuoso participant (VP) prompt and response behaviors.

		Tasks															
		Check Daily Schedule				Check Map				Walk to Shuttle				Wait for Shuttle			
Session	User	Mean	Min	Max	SD	Mean	Min	Max	SD	Mean	Min	Max	SD	Mean	Min	Max	SD
VRL1	OG	5.2	4.0	9.0	1.9	7.8	4.0	10.0	2.4	9.5	2.0	15.0	4.2	0.7	0.0	2.0	0.8
	VP	4.8	4.0	8.0	1.6	8.	5.0	10.0	1.8	8.7	3.0	15.0	4.9	0.2	0.0	1.0	0.5
VRL2	OG	3.5	0.0	6.0	2.4	7.0	3.0	9.0	2.2	13.0	10.0	20.0	3.6	8.2	1.0	17.0	5.7
	VP	4.0	0.0	7.0	2.6	6.7	3.0	8.0	2.0	12.5	9.0	19.0	3.4	7.3	1.0	15.0	5.0
Real-world	OG	5.5	3.0	9.0	2.4	8.2	5.0	12.0	2.5	6.3	5.0	11.0	2.4	8.3	5.0	13.0	2.8
	VP	5.3	3.0	8.0	1.9	7.8	5.0	12.0	2.5	6.8	4.0	11.0	2.5	8.0	5.0	13.0	3.1

An ANOVA comparison of OG prompting behaviors and Virtuoso participant (VP) response behaviors in different sessions was performed. The four tasks compared were (1) checking the daily schedule, (2) checking the map, (3) walking to the shuttle, and (4) waiting for the shuttle. The comparisons were made between VRL1 and VRL2, between VRL1 and the real world, and between VRL2 and the real world. For the comparison of VRL1 and VRL2, there was no significant difference in the OG prompting behaviors for checking the daily schedule ($F(1,10) = 1.72$, $p = 0.22$) or checking the map ($F(1,10) = 0.39$, $p = 0.54$). However, there was a significant difference in the OG prompting behaviors for walking to the shuttle ($F(1,10) = 2.28$, $p = 0.16$) and waiting for the shuttle ($F(1,10) = 10.28$, $p = 0.01$). There was no significant difference in VP response behaviors for checking the daily schedule ($F(1,10) = 1.04$, $p = 0.33$), checking the map ($F(1,10) = 0.81$, $p = 0.39$), or waiting for the shuttle ($F(1,10) = 0.08$, $p = 0.79$), but there was a significant difference in their response behavior for walking to the shuttle ($F(1,10) = 10.91$, $p = 0.01$). For the comparison of VRL1 and the real world, there was no significant difference in the OG prompting behaviors for checking the daily schedule ($F(1,10) = 0.07$, $p = 0.80$), checking the map ($F(1,10) = 0.06$, $p = 0.82$), or walking to the shuttle ($F(1,10) = 2.37$, $p = 0.16$). However, there was a significant difference in the OG prompting behaviors for waiting for the shuttle ($F(1,10) = 41.33$, $p = 0$). There was no significant difference in the VP response behaviors for checking the daily schedule ($F(1,10) = 0.45$, $p = 0.52$), checking the map ($F(1,10) = 1.87$, $p = 0.20$), or walking to the shuttle ($F(1,10) = 2.47$, $p = 0.15$), but there was a significant difference in their response behavior for waiting for the shuttle ($F(1,10) = 10.06$, $p = 0.01$). For the comparison of VRL1 with the real world, ANOVA results showed that both the OG and VP behavior was significantly different in the "Wait for Shuttle" task ($p < 0.01$). Results also suggest a significant difference in VP behavior in the "Wait for Shuttle" task ($p < 0.01$). The "Check Daily Schedule," "Check Map," and "Walk to Shuttle" tasks, however, showed no significant differences between the OG and VP behavior ($p > 0.05$). Full results of the ANOVA are presented in Table 9.

Table 9
ANOVA results comparing OG prompting behaviors and Virtuoso participant (VP) response behaviors between sessions.

		Task Name			
Session Comparison	User	Check Daily Schedule	Check Map	Walk to Shuttle	Wait for Shuttle
VRL1 & VRL2	OG	$F(1,10) = 1.72$, $p = 0.22$	$F(1,10) = 0.39$, $p = 0.54$	$F(1,10) = 2.28$, $p = 0.16$	$F(1,10) = 10.28$, $p = 0.01$*
	VP	$F(1,10) = 1.04$, $p = 0.33$	$F(1,10) = 0.81$, $p = 0.39$	$F(1,10) = 10.91$, $p = 0.01$*	$F(1,10) = 0.08$, $p = 0.79$
VRL1 & real-world	OG	$F(1,10) = 0.07$, $p = 0.80$	$F(1,10) = 0.06$, $p = 0.82$	$F(1,10) = 2.37$, $p = 0.16$	$F(1,10) = 41.33$, $p = 0$*
	VP	$F(1,10) = 0.45$, $p = 0.52$	$F(1,10) = 1.87$, $p = 0.20$	$F(1,10) = 2.47$, $p = 0.15$	$F(1,10) = 10.06$, $p = 0.01$*
VRL2 & real-world	OG	$F(1,10) = 2.03$, $p = 0.184$	$F(1,10) = 0.745$, $p = 0.408$	$F(1,10) = 14.29$, $p = 0.00$*	$F(1,10) = 0.00$, p = 0.95
	VP	$F(1,10) = 0.25$, $p = 0.63$	$F(1,10) = 0.07$, $p = 0.80$	$F(1,10) = 0.66$, $p = 0.44$	$F(1,10) = 30.6$, $p = 0$*

*$p > 0.05$.

5.2. Results for RQ2: to what extent did participants perceive that they were present in the VR environment?

Results from the TPI indicate that participants rated the realism of both VRL1 and VRL2 highly. Mean computed presence scores across participants for VRL1 were 92.2 ($SD = 7.23$). Mean computed presence scores across participants for VRL2 were 93.3 ($SD = 5.81$). Overall, the results from the Temple Presence Inventory for VRL1 and VRL2 show that participants generally had high scores in most categories. The mean scores for VRL1 were: Spatial Presence (91), Social Presence - Actor within medium (100), Social Presence - Passive Interpersonal (91), Engagement (Mental Immersion) (82), Social Realism (94), Perceptual Realism (95), and Overall Score (92). The mean scores for VRL2 were: Spatial Presence (88), Social Presence - Actor within medium (98), Social Presence - Passive Interpersonal (92), Engagement (Mental Immersion) (93), Social Realism (97), Perceptual Realism (98), and Overall Score (93.32). In both VRL1 and VRL2, participants generally had high scores for Spatial Presence, Social Presence - Actor within medium, Engagement (Mental Immersion), and Perceptual Realism. The mean scores for Social Presence - Passive Interpersonal and Social Realism were also high, but with a larger standard deviation. Computed values across TPI subscales for VRL1 and VRL2 were also highly rated, ranging from 83 to 100, with ratings for VRL2 being slightly higher, suggesting that this level may have been perceived as being more realistic (see Table 10).

5.3. Results for RQ3: how did participants characterize similarities and differences between the VR environment and the real-world?

Results for RQ3 indicate that, in general, participants found the 3D assets included in VRL1 and VRL2 to be comparable to real-world counterparts. However, this sentiment was not as strongly conveyed in regards to humanoid avatars used by both the OG's and non-playable characters (NPCs) seen in VRL2, as discussed below.

5.3.1. Perceptions of digital asset realism

Across VRL1 and VRL2 sessions, participants overwhelmingly stated that they felt like they were actually on campus and that the 3D models used to represent the buildings seemed realistic. For example, in VRL1, Evan said that he felt like he was “really there” when he exited the office onto the university campus. He readily identified campus buildings and was able to tell the OG the location of buildings that he frequently visits in his day-to-day life, for example, “that's McMicken Hall.” This ability to recognize 3D models of campus buildings was shared by other participants, such as Andy who pointed out that he could “see the Teachers' College” while walking to the shuttle stop. Participants also commented on other virtual assets and made remarks about visual and behavioral fidelity. For example, in VRL1, Andy pointed out that “the wind is blowing” after noticing that the branches on the trees were moving. Travis also commented on some of the environmental details such as saying the sky was “spot on” and that the graphics of the trees were “fairly good”. Further, he explained that he felt like he was “actually on campus.” The virtual shuttle was also perceived as being realistic. Transcript excerpts illustrate that participants had no problem recognizing shuttles and were able to identify their routes. For example, in VRL1, Kevin said “I felt like I was really getting on the shuttle.”

5.3.2. Perceptions of avatar realism

Humanoid avatars were the asset with the most discrepancy in participants' perceptions. The avatars used were perceived as being realistic and contributing to the 3D environment's realism by some participants, although they also were perceived as having some non-realistic characteristics. For example, Kevin told us that he thought the OG's avatar was realistic looking. In fact, all participants were able to tell the research team accurately who the real-world counterpart of the OG's avatar was. However, the avatar's mannerisms and movements were sometimes perceived as being unrealistic. Travis told us that he was confused as to why her mouth was not moving when the OG was talking, but also stated, “I get it. It's a video game,” a sentiment similar to that reported in Cobb et al. (2002) that “it's just a VE [virtual environment]” (p. 19). Later, in VRL2 he told us that he found the OG to be realistic because of the “behavioral realism”. Another example is when Andy stated that the OG was realistic but also felt the avatar was robotic, although he also stated he understood it was controlled by a human. Despite varied perceptions of avatar realism, it appears that participants perceived the OG avatar as being controlled by a human. In regards to NPCs, participants seemed unsure if the humanoids were pre-programmed or were controlled by other users. After completing VRL2, Kevin told us he was unsure if they were controlled by “robots”, but he was sure they were supposed to be other students on campus. Another example of this is when Andy explained that the NPCs were realistic but they “felt different” and they were not controlled by a real person.

6. Discussion

6.1. Discussion of results for research question 1

RQ1 investigated how similar or different behavior in the VR environment was relative to behavior in the real-world. To approach this question, percentage of task completion, total time on task, and OG and participant behaviors across sessions were analyzed. Differences in OG prompting behavior for the Wait for Shuttle task in VRL1 could be explained by variances in time for the shuttle to arrive. The virtual shuttle was programmed to arrive approximately every 2 min, meaning that participants would have to wait longer if they arrived at the shuttle stop directly after the shuttle had departed. In these cases, researchers observed the OG would sometimes go off-script to keep the participant engaged, asking questions such as “Do you remember which shuttle we are getting on?” or providing reminders such as “Don't forget, the shuttle will be here soon so we need to wait patiently.” This led to an increased number of prompts and respective participant responses. Data analysis suggests high fidelity of implementation, which contributes to intervention validity and reliability and is critical for promoting consistent, predictable, effective outcomes (Lemire et al., 2022; Rojas-Andrade & Bahamondes, 2019).

With evidence supporting fidelity of implementation, this research looked to compare participant behaviors between the VR sessions and the real-world session. Importantly, findings suggest no significant differences between participant response behaviors in VR sessions and the real-world session. That is, participant response frequencies were similar between VR sessions, suggesting that learning may have generalized proximally (i.e., from one VR training session to the next). In addition, response frequencies were similar between VR sessions and the real-

Table 10
Participant results from Temple Presence Inventory for VRL1 and VRL2 on a scale of 0–100.

VRL1								
	Travis	Kevin	Andy	Evan	Jonah	Keith	Mean	SD
Spatial Presence	86	79	96	100	100	83	91	9.00
Social presence - Actor within medium	100	100	100	100	100	100	100	0.00
Social presence - Passive interpersonal	43	100	100	100	100	100	91	23.00
Engagement (mental immersion)	71	100	86	100	100	33	82	26.00
Social realism	86	100	100	100	89	89	94	7.00
Perceptual realism	94	100	93	100	100	83	95	7.00
Overall score	83	93	96	100	97	83	92	7.23
VRL2								
	Travis	Kevin	Andy	Evan	Jonah	Keith	Mean	SD
Spatial Presence	89	79	86	100	100	75	88	11.00
Social presence - Actor within medium	86	100	100	100	100	100	98	6.00
Social presence - Passive interpersonal	86	100	100	100	100	67	92	14.0
Engagement (mental immersion)	71	100	86	100	100	100	93	12.0
Social realism	90	100	100	100	100	89	97	5.00
Perceptual realism	93	100	93	100	100	100	98	4.00
Overall score	88.10	92.86	92.86	100	100	86.11	93.32	5.81

world session, suggesting potential distal learning generalization (i.e., from VR training to the real-world).

One exception was participant behaviors during the Walk to Shuttle task in VRL2, which could be explained by a lack of fluency with the VR controls for some participants, which in one case (Kevin) was exacerbated by severe motor control challenges, as well as a technical glitch in which one of the participant's avatars got stuck in the virtual geometry. This finding is in line with other autism and VR research which indicates that autistic users sometimes struggle with using VR controls (Bozgeyikli et al., 2016). This required additional prompting by the OG to help the participant move his avatar back onto the virtual sidewalk. Conversely, the real-world lacked technical glitches or the need to navigate one's avatar using controllers, meaning that there was no need for additional prompting or responses. Future designers could consider adapting their control scheme for users requiring a higher degree of assistance. For example, there is also the potential for an on-rails control scheme that automatically navigates the character through the instructional environment which could make the system easier to use and remove the need for additional prompting. This recommendation is made with the caveat that on-rails approaches may be less effective if the goal of intervention is to promote the development of spatial navigation skills (Jang, Vitale, Jyung, & Black, 2017).

As discussed previously, generalization is unlikely to occur as a matter of course. Researchers must actively design and program for generalization (Stokes & Osnes, 2016). Virtuoso was designed intentionally to embody generalization heuristics while promoting useful skills for public transportation utilization for autistic adults (Schmidt et al., 2020). The primary aim was to create an intervention that was grounded in well-established, evidence-based practices and to improve the intervention based on iterative design, enactment, evaluation, reflection, and revision. As such, the findings related to Research Question 1 are congruent with others who have applied these heuristics for VR interventions (e.g., Dixon et al., 2019).

6.2. Discussion of results for research question 2

RQ2 investigated the extent to which participants perceived themselves to be present within the Virtuoso-VR learning environment. The psychology of telepresence has long been recognized as a fundamental factor in influencing behavior in VR. For example, seminal work by researchers at the Stanford Virtual Human Interaction lab suggests that telepresence and realism are vital for exerting social influence in VR environments (Blascovich, 2002). It is broadly understood that if users are able to feel psychologically present then the actions that they participate in within the virtual environment can take on a deeper meaning and users can then form a better mental model of their behaviors to that of the real world.

To examine this, researchers analyzed TPI data to obtain mean computed presence scores across multiple dimensions of telepresence (i.e., social realism, mental immersion, spatial presence, etc.). Findings suggest that participants perceived high degrees of telepresence in both VRL1 and VRL2. Of note, aggregate results indicate that participants perceived VRL2 as promoting a higher sense of telepresence than VRL1, although this was not uniform across all participants. In VRL2, participants completed the same tasks, but using a modified version of the VR environment (i.e. scaffolds were removed, more environmental details were added, etc.). This echoes findings of prior research, which suggest that more realistic VR environments can induce higher perceptions of telepresence (e.g. Bailenson et al., 2006; Jung & Lindeman, 2021).

Because of the heterogeneity of the participants and established challenges using standardized measures with autistic populations, the extent to which telepresence would actually manifest in the study was uncertain. Indeed, autism research suggests that individual differences such as IQ, anxiety, and resilience can influence participants' perceptions of telepresence in VR environments (Malihi et al., 2020). That participants' perceptions of telepresence were uniformly high is reassuring and contributes further evidence that, when VR environments are carefully designed and implemented, participants can feel virtually present, which is a critical prerequisite to establishing a sufficiently realistic experience to support generalization. However, this claim is tempered with Parsons' (2016) assertion that realism is but one of many factors that contribute to generalization, and that designers of VR interventions for autistic people must consider a range of factors, including technology type, user demographics, usage contexts, necessary supports, and intended learning objectives. There remains a need to better understand how autistic learners experience telepresence and the role of telepresence in addressing standing issues of VR generalization. These issues have some overlap with the interpretation of findings for Research Question 3, discussed in the following section.

6.3. Discussion of results for research question 3

Research Question 3 investigated how participants characterized the similarities and differences between the VR environment and the real-world. To approach this research question, researchers used a variety of qualitative data sources and analysis techniques to unveil participants' impressions of realism regarding 3D assets. Findings suggest that participants overwhelmingly found the campus models, terrain, and humanoid avatars to be recognizable and that, in most cases, these digital representations matched their expectations of how their real-world counterparts would look and behave. That is, the photographic and behavioral fidelity of the VR environments were sufficient for participants to recognize digital assets as being analogs of real-world counterparts. According to Blascovich et al. (2002), behavioral fidelity "refers to the degree to which virtual humans and other objects within IVEs behave as they would in the physical world" (p. 111). These scholars position behavioral fidelity as a contributor to social influence, which plays both an implicit and explicit role in promoting conceptual, attitudinal and behavioral change (Izuma, 2017). In social VR environments like Virtuoso, social influence is critical for the OG to socially facilitate the learning experience, i.e., using prompting to ensure that participants obey, conform to behavioral conventions, and perform the required tasks.

Social influence in virtual environments is not only impacted by the behavioral fidelity of the VR environment, but also social presence. Social presence is "the degree to which the user (e.g., the participant) believes that he or she is in the presence of and interacting with another veritable human being and that the behaviors of virtual humans within [immersive virtual environments] represent the actions of real individuals in the physical world in real time" (Blascovich et al., 2002, p. 111). In alignment with findings from the AS Interactive project (Cobb et al., 2002; Parsons et al., 2000; Lancaster et al., 2002), researchers found that some participants in Virtuoso perceived humanoid avatars to behave in non-realistic ways and to maintain characteristics that did not align with their expectations of the real-world. For example, some participants were disturbed by unnatural avatar animations (i.e., walking, gesturing, etc.), inconsistent mouth movements when the OG was speaking, and perceptions that the OG avatar seemed "robotic." The findings suggest that there was some confusion as to whether the avatars were controlled by a human, although the exact reasons for this are unclear. This could be explained by limitations in the avatar designs. It could also be explained by participants' prior experiences with immersive technologies (i.e., video games, VR, etc.) in that they may not have expected avatars to be human-controlled. Nonetheless, TPI findings suggest that most participants perceived high levels of social presence in the Virtuoso-VR environment. Therefore, the impact of these avatar designs on social presence remains unclear. Research suggests that human-controlled avatars promote superior behavioral responses than pre-programmed NPCs (e.g. Carter et al., 2014). To achieve superior behavioral responses, research suggests autistic participants must perceive the avatar as real and controlled by a human (Wallace et al., 2017). This could imply that researchers and designers should explicitly inform autistic users (i.e., participants) that the avatars with whom they are interacting are human-controlled. However, further research is needed to better understand the extent to which avatar design influences social presence (c.f. Gisbergen et al., 2019).

7. Limitations

The findings presented in this paper should be interpreted in light of the limitations presented here. First, researchers did not explicitly test for mastery of the skill in the VR or real-world contexts. While data supports participants completing all tasks, mastery claims cannot be supported. Second, as with most research on VR with autistic populations, researchers did not evaluate long-term retention of the skill – an area for future research. Third, the small sample size used in this research limits the ability to generalize these findings to other research contexts. Fourth, the sample was composed completely of adult males – a recurring issue in autism research (Hull et al., 2020) – and within a specific urban university setting. This limitation is a recurring issue in autism research because diagnosticians encounter numerous challenges while assessing women for autism; mainly due to differences in presentation and the limitations of diagnostic tools that reflect these differences (Tsirgiotis et al., 2021). While the role of clinical judgment in diagnosing autistic people is crucial, there is limited understanding of the difficulties faced by diagnosticians when evaluating female clients (Duvekot et al., 2017). Further, autism is a low incidence disability which contributes to the small sample sizes in autism research. Since only approximately 1 in every 54 people in the US are autistic (Baron-Cohen, 2017), it is often challenging for researchers to find large sample sizes relative to specific inclusion criteria (e.g., Rubenstein & Furnier, 2021). Extending and diversifying study samples remains a direction for future research, both in the general autism research field and for research studies investigating the use of extended reality (Zener, 2019). Fifth, research has established that the use of surveys with autistic participants can lead to bias in results. As the research utilized the TPI, it is therefore susceptible to such bias. However, researchers did find substantial parity in responses between participants, though this could also be attributed to potential Hawthorne effect, in that researchers were present both for execution of study procedures and for proctoring of the TBI.

8. Implications for research and practice

Generalization from digital environments to the real-world has been widely cited as justification for the use of VR among autistic people. However, the research evidence supporting generalization is remarkably limited (Glaser & Schmidt, 2021; Mesa-Gresa et al., 2018), as is guidance on how to design VR environments to promote generalization (Schmidt & Glaser, 2021). Few studies explicitly investigate generalization effects. Those that do suffer from a range of methodological limitations, such as using self-report data (cf., Parsons et al., 2006) or reporting highly variable and inconclusive outcomes (cf., Self et al., 2007). In a field that has a paucity of empirical support for generalization, such studies are helpful in that they provide evidence of promise; however, research with more rigorous experimental control is needed to validly and reliably establish an empirical basis of support for generalization.

Few studies have attempted to approach the problem of generalization with such rigor. Notably, those that have tend to focus on relatively simple, discrete skills related to safety. For example, Strickland, McAllister, Coles, & Osborne (2007) developed a VR environment to promote fire safety skills, after which a proportion of study participants completed a generalization task without error. Josman et al. (2008) developed a VR training intervention to teach street-crossing skills to autistic children. Their findings indicated that about half of the participants were able to improve their pedestrian behavior following the VR intervention. Lastly, perhaps the most compelling and methodologically sound example of generalization can be found in Dixon et al. (2019), who used single-subject design research methods to evaluate a street-crossing intervention with three autistic children. Participants in this study were able to reach mastery in both the VR environment and in the real-world. Similar to the Virtuoso study reported here, Dixon and colleagues specifically programmed for generalization using heuristics proposed by Stokes and Osnes (2016). However, a key difference in this study was that, although the researchers used fully immersive head-mounted display-based VR, their VR intervention was based on 360-degree spherical videos, which provide only three degrees of freedom and different interaction possibilities (Glaser et al., 2022). Affordances of fully interactive, digitally-rendered VR systems that provide six degrees of freedom were absent, such as avatars, avatar control, a fully interactive environment, etc.

Meta-analyses and systematic reviews show the general lack of research reporting generalization outcomes in VR studies focusing on autism (Karami et al., 2021; Mesa-Gresa et al., 2018). This paucity of research extends to studies that report how to program for generalization in VR. Among these, the principal focus has centered around "a closer fit with the real world in order to assess cognition and support the generalization of learning" (Parsons, 2016, p. 154). However, this approach has been criticized as being conceptually narrow (Parsons & Cobb, 2011). Closer inspection suggests that this critique rests on how researchers have characterized the construct of realism, which has been primarily focused on visual aspects. The current study extends notions of realism beyond just visual aspects, considering issues of task realism (i.e., how well task analysis is operationalized in the VR environment), object and behavioral congruence (i.e., how objects and avatars behave relative to real-world counterparts), fidelity of implementation (i.e., providing the same instruction across all intervention modalities), perceptions of telepresence and social presence (i.e., feeling present in the VR environment and in the presence of others), and explicit programming for generalization (i.e., incorporation of generalization heuristics). This more holistic conceptualization of realism in VR could provide fertile conceptual ground for further research and development.

9. Conclusion

The work presented here contributes to the limited evidence base on generalization in VR for autistic users. Researchers have long maintained that designing VR interventions for autistic populations is a wicked problem (Schmidt, 2014) with no established or ideal solution, and solving one aspect of the problem might exacerbate other problems. Yet few researchers approach the problem of designing VR for autistic users from this perspective. Instead, the majority of research in this area is remarkably technocentric, attributing all questions and solutions to the technology itself. However, the technology underlying VR interventions remains remarkably immature, difficult to support in real-world intervention contexts, and particularly prohibitive for those who lack strong software development skills. As such, the technocentric stance could be problematically flawed when viewed in light of issues such as (1) challenges associated with realizing real-world tasks within VR technology systems (Glaser et al., 2021), (2) human-computer interaction issues related to cybersickness (Glaser et al., 2022), (3) technologies becoming quickly obsolete (sometimes in the midst of a study), and (4) the difficulty of sustaining and supporting VR interventions longitudinally. These issues highlight the importance of considering the complex range of factors beyond technology itself when designing VR interventions for autistic users.

In light of these issues, questions remain as to how researchers might more productively approach the design of VR interventions for autistic users in tremendously complex design contexts. We maintain that the issues we have presented will sustain until one of the following conditions is met: (1) the technology becomes easier to use and support by orders of magnitude and/or (2) intervention designers are able to partner with highly sophisticated software development studios to create products that are viable for mass consumption. Regarding the former, more useable and sustainable systems are predicted by computing laws, such as Moore's Law (Moore, 1965) and Kurzweil's Law (Kurzweil, 2004), which suggest that computing power will double every 1.5 years and that this advancement will lead to even more rapid technology acceleration. Extrapolating from this, it may be the case that VR is simply too immature in its current state for widespread implementation and use, but that this will inevitably change with time. Regarding intervention designers partnering with software development studios, precedent exists in the area of digital games based learning, where researchers have noted similar challenges (Hirumi et al., 2010). Given the similarities between video game development and VR development, designers of VR for autistic users should explore collaborating with more sophisticated development studios, as was the case with the Floreo VR system, which is perhaps the most successful commercial venture in this space (Ravindran et al., 2019).

Despite over two decades of research in the field of VR for autistic users, the surprising lack of support for generalization - one of the most fundamental justifications for the use of VR - is a major concern. To address this gap, we have taken action by explicitly programming for generalization and reconceptualizing notions of realism. Our hope is to inspire and encourage future research in this critical area.

Statements on open data and ethics

The datasets generated and/or analyzed during the current study are not publicly available due to the vulnerable nature of our participants.

Ethical approval

All research presented in this manuscript was reviewed by our institutional review board.

Declaration of competing interest

The authors declare no financial or other competing interests.

References

American Psychiatric Association, D. (2013). *Diagnostic and statistical manual of mental disorders: DSM-5* (Vol. 5). DC: American psychiatric association Washington.

Aresti-Bartolome, N., & Garcia-Zapirain, B. (2014). Technologies as support tools for persons with autistic spectrum disorder: A systematic review. *International Journal of Environmental Research and Public Health, 11*(8), 7767–7802.

Bailenson, J. N., Yee, N., Merget, D., & Schroeder, R. (2006). The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments, 15*(4), 359–372.

Baron-Cohen, S. (2017). Editorial Perspective: Neurodiversity – a revolutionary concept for autism and psychiatry. *Journal of Child Psychology and Psychiatry, 58*(6), 744–747. https://doi.org/10.1111/jcpp.12703

Blascovich, J. (2002). Social influence within immersive virtual environments. In R. Schroeder (Ed.), *The social life of avatars: Presence and interaction in shared virtual environments* (pp. 127–145). Springer. https://doi.org/10.1007/978-1-4471-0277-9_8.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 13*(2), 103–124 (JSTOR).

Bozgeyikli, L., Raij, A., Katkoori, S., & Alqasemi, R. (2018). A survey on virtual reality for individuals with autism spectrum disorder: Design considerations. *IEEE Transactions on Learning Technologies, 11*(2), 133–151. https://doi.org/10.1109/TLT.2017.2739747

Bozgeyikli, E., Raij, A., Katkoori, S., & Dubey, R. (2016). Locomotion in virtual reality for individuals with autism spectrum disorder. In *Proceedings of the 2016 symposium on spatial user interaction* (pp. 33–42).

Brondino, N., Fusar-Poli, L., Miceli, E., Di Stefano, M., Damiani, S., Rocchetti, M., & Politi, P. (2019). Prevalence of medical comorbidities in adults with autism spectrum disorder. *Journal of General Internal Medicine, 34*(10), 1992–1994.

Brown, H. M., Stahmer, A. C., Dwyer, P., & Rivera, S. (2021). Changing the story: How diagnosticians can support a neurodiversity perspective from the start. *Autism, 25*(5), 1171–1174. https://doi.org/10.1177/13623613211001012

Carter, E. J., Williams, D. L., Hodgins, J. K., & Lehman, J. F. (2014). Are children with autism more responsive to animated characters? A study of interactions with humans and human-controlled avatars. *Journal of Autism and Developmental Disorders, 44*(10), 2475–2485. https://doi.org/10.1007/s10803-014-2116-8

Cengher, M., Budd, A., Farrell, N., & Fienup, D. M. (2018). A review of prompt-fading procedures: Implications for effective and efficient skill acquisition. *Journal of Developmental and Physical Disabilities, 30*(2), 155–173. https://doi.org/10.1007/s10882-017-9575-8

Cobb, S., Beardon, L., Eastgate, R., Glover, T., Kerr, S., Neale, H., Parsons, S., Benford, S., Hopkins, E., & Mitchell, P. (2002). Applied virtual environments to support learning of social interaction skills in users with Asperger's Syndrome. *Digital Creativity, 13*(1), 11–22.

Coleman, D. M., & Adams, J. B. (2018). Survey of vocational experiences of adults with autism spectrum disorders, and recommendations on improving their employment. *Journal of Vocational Rehabilitation, 49*(1), 67–78.

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1), 10–32. https://doi.org/10.1111/j.1467-8535.2009.01038.x

Deka, D., Feeley, C., & Lubin, A. (2016). Travel patterns, needs, and barriers of adults with autism spectrum disorder: Report from a survey. *Transportation Research Record, 2542*(1), 9–16.

Dixon, D. R., Bergstrom, R., Smith, M. N., & Tarbox, J. (2010). A review of research on procedures for teaching safety skills to persons with developmental disabilities. *Research in Developmental Disabilities, 31*(5), 985–994.

Dixon, D. R., Miyake, C. J., Nohelty, K., Novack, M. N., & Granpeesheh, D. (2019). Evaluation of an immersive virtual reality safety training used to teach pedestrian skills to children with autism spectrum disorder. *Behavior Analysis in practice. Journal Article*, 1–10.

Doyle, P. M., Wolery, M., Ault, M. J., & Gast, D. L. (1988). System of least prompts: A literature review of procedural parameters. *The Journal of the Association for Persons with Severe Handicaps, 13*(1), 28–40. https://doi.org/10.1177/154079698801300104

Duvekot, J., van der Ende, J., Verhulst, F. C., Slappendel, G., van Daalen, E., Maras, A., & Greaves-Lord, K. (2017). Factors influencing the probability of a diagnosis of autism spectrum disorder in girls versus boys. *Autism, 21*(6), 646–658.

Eaves, L. C., & Ho, H. H. (2008). Young adult outcome of autism spectrum disorders. *Journal of Autism and Developmental Disorders, 38*, 739–747.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of convenience sampling and purposive sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1–4.

Feeley, C., Deka, D., Lubin, A., & McGackin, M. (2015). *Detour to the right place: A study with recommendations for addressing the transportation needs and barriers of adults on the autism spectrum in New Jersey*. Rutgers Center for Advanced Infrastructure and Transportation *[Technical Report]* (p. 57).

Felce, D. (1997). Defining and applying the concept of quality of life. *Journal of Intellectual Disability Research, 41*(2), 126–135.

Gisbergen, M. van, Kovacs, M., Campos, F., Heeft, M., & Vugts, V. (2019). What we don't know. The effect of realism in virtual reality on experience and behaviour. In *Augmented reality and virtual reality* (pp. 45–57). Springer.

Glaser, N., Newbutt, N., Palmer, H., & Schmidt, M. (2022). *Video-based virtual reality technology for autistic users: An emerging technology report. Technology, Knowledge and learning*. https://doi.org/10.1007/s10758-022-09594-x

Glaser, N. J., & Schmidt, M. (2018). Usage considerations of 3D collaborative virtual learning environments to promote development and transfer of knowledge and skills for individuals with autism. *Technology, Knowledge and Learning*, 1–8.

Glaser, N., & Schmidt, M. (2021). Systematic literature review of virtual reality intervention design patterns for individuals with autism. *International Journal of Human-Computer Interaction, 37*.

Glaser, N., Schmidt, M., Schmidt, C., Palmer, H., & Beck, D. (2021). The centrality of interdisciplinarity for overcoming design and development constraints of a multi-user virtual reality intervention for adults with autism: A design case. In B. Hokanson, M. Exter, A. Grincewicz, M. Schmidt, & A. A. Tawfik (Eds.), *Intersections across disciplines: Interdisciplinarity and learning* (pp. 157–171). Springer International Publishing. https://doi.org/10.1007/978-3-030-53875-0_13.

Grynszpan, O., Weiss, P. L., Perez-Diaz, F., & Gal, E. (2014). Innovative technology-based interventions for autism spectrum disorders: A meta-analysis. *Autism, 18*(4), 346–361.

Hale, K. S., & Stanney, K. M. (2014). *Handbook of virtual environments: Design, implementation, and applications* (2nd ed.). CRC Press.

Hull, L., Petrides, K., & Mandy, W. (2020). The female autism phenotype and camouflaging: A narrative review. *Review Journal of Autism and Developmental Disorders, 7*(4), 306–317.

Izuma, K. (2017). Chapter 16—the neural bases of social influence on valuation and behavior. In J.-C. Dreher, & L. Tremblay (Eds.), *Decision neuroscience* (pp. 199–209). Academic Press. https://doi.org/10.1016/B978-0-12-805308-9.00016-6.

Jang, S., Vitale, J. M., Jyung, R. W., & Black, J. B. (2017). Direct manipulation is better than passive viewing for learning anatomy in a three-dimensional virtual reality environment. *Computers & Education, 106*, 150–165.

Josman, N., Ben-Chaim, H. M., Friedrich, S., & Weiss, P. L. (2008). Effectiveness of virtual reality for teaching street-crossing skills to children and adolescents with autism. *International Journal on Disability and Human Development, 7*(1). https://doi.org/10.1515/IJDHD.2008.7.1.49

Jung, S., & Lindeman, R. W. (2021). Perspective: Does realism improve presence in VR? Suggesting a model and metric for VR experience evaluation. In *Frontiers in virtual reality* (Vol. 2). https://www.frontiersin.org/articles/10.3389/frvir.2021.693327.

Karami, B., Koushki, R., Arabgol, F., Rahmani, M., & Vahabie, A.-H. (2021). Effectiveness of virtual/augmented reality–based therapeutic interventions on individuals with autism spectrum disorder: A comprehensive meta-analysis. *Frontiers in Psychiatry, 12*, Article 665326.

Khowaja, K., Banire, B., Al-Thani, D., Sqalli, M. T., Aqle, A., Shah, A., & Salim, S. S. (2020). Augmented reality for learning of children and adolescents with autism spectrum disorder (ASD): A systematic review. *IEEE Access, 8*, 78779–78807.

Kraemer, H. C., Kupfer, D. J., Clarke, D. E., Narrow, W. E., & Regier, D. A. (2012). DSM-5: How reliable is reliable enough? *American Journal of Psychiatry, 169*(1), 13–15.

Lemire, C., Rousseau, M., & Dionne, C. (2022). A comparison of fidelity implementation frameworks used in the field of early intervention. *American Journal of Evaluation*, Article 10982140211008978.

Lombard, M., Bolmarcich, T., & Weinstein, L. (2009). *Measuring presence: The Temple presence inventory*.

Loomes, R., Hull, L., & Mandy, W. P. L. (2017). What is the male-to-female ratio in autism spectrum disorder? A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry, 56*(6), 466–474.

Lorenz, T., Frischling, C., Cuadros, R., & Heinitz, K. (2016). Autism and overcoming job barriers: Comparing job-related barriers and possible solutions in and outside of autism-specific employment. *PLoS One, 11*(1), Article e0147040.

Lubin, J., & Brooks, J. (2021). College students with autism: Navigating the bumpy road. *College Student Journal, 55*(3), 318–324 (APA PsycInfo).

Massey, N. G., & Wheeler, J. J. (2000). Acquisition and generalization of activity schedules and their effects on task engagement in a young child with autism in an inclusive pre-school classroom. *Education and Training in Mental Retardation and Developmental Disabilities*, 326–335.

McCoy, K., & Hermansen, E. (2007). Video modeling for individuals with autism: A review of model types and effects. *Education & Treatment of Children*, 183–213.

Mesa-Gresa, P., Gil-Gómez, H., Lozano-Quilis, J.-A., & Gil-Gómez, J.-A. (2018). Effectiveness of virtual reality for children and adolescents with autism spectrum disorder: An evidence-based systematic review. *Sensors, 18*(8), 2486. https://doi.org/10.3390/s18082486

Müller, E., Schuler, A., & Yates, G. B. (2008). Social challenges and supports from the perspective of individuals with Asperger syndrome and other autism spectrum disabilities. *Autism, 12*(2), 173–190.

Osgood, T. (2022). *Practice leadership in challenging behaviour services for autism and intellectual disabilities: Practical strategies for supporting people*. Jessica Kingsley Publishers.

Palmen, A., & Didden, R. (2012). Task engagement in young adults with high-functioning autism spectrum disorders: Generalization effects of behavioral skills training. *Research in Autism Spectrum Disorders, 6*(4), 1377–1388.

Palmen, S. J., van Engeland, H., Hof, P. R., & Schmitz, C. (2004). Neuropathological findings in autism. *Brain, 127*(12), 2572–2583.

Parsons, S. (2016). Authenticity in virtual reality for assessment and intervention in autism: A conceptual review. *Educational Research Review, 19*, 138–157. https://doi.org/10.1016/j.edurev.2016.08.001

Parsons, S., Beardon, L., Neale, H. R., Reynard, G., Eastgate, R., Wilson, J. R., Cobb, S., Benford, S. D., Mitchell, P., & Hopkins, E. (2000). *Development of social skills amongst adults with Asperger's Syndrome using virtual environments: The "AS Interactive* (pp. 23–25). *Virtual Reality and Associated Technologies, ICDVRAT*.

Parsons, S., & Cobb, S. (2011). State-of-the-art of virtual reality technologies for children on the autism spectrum. *European Journal of Special Needs Education, 26*(3), 355–366.

Parsons, S., Leonard, A., & Mitchell, P. (2006). Virtual environments for social skills training: Comments from two adolescents with autistic spectrum disorder. *Computers and Education, 47*(2), 186–206. https://doi.org/10.1016/j.compedu.2004.10.003. Scopus.

Rezae, M., Chen, N., McMeekin, D., Tan, T., Krishna, A., & Lee, H. (2020). The evaluation of a mobile user interface for people on the autism spectrum: An eye movement study. *International Journal of Human-Computer Studies, 142*. https://doi.org/10.1016/j.ijhcs.2020.102462. APA PsycInfo.

Rojas-Andrade, R., & Bahamondes, L. L. (2019). Is implementation fidelity important? A systematic review on school-based mental health programs. *Contemporary School Psychology, 23*(4), 339–350.

Rubenstein, E., & Furnier, S. (2021). # Bias: The opportunities and challenges of surveys that recruit and collect data of autistic adults online. *Autism in Adulthood, 3*(2), 120–128.

Schmidt, M. M. (2014). Designing for learning in a three-dimensional virtual learning environment: A design-based research approach. *Journal of Special Education Technology, 29*(4), 59–71.

Schmidt, M., & Glaser, N. (2021). Investigating the usability and learner experience of a virtual reality adaptive skills intervention for adults with autism spectrum disorder. *Educational Technology Research and Development, 69*(3), 1665–1699.

Schmidt, M., & Glaser, N. (2021). Piloting an adaptive skills virtual reality intervention for adults with autism: Findings from user-centered formative design and evaluation. *Journal of Enabling Technologies*. https://doi.org/10.1108/JET-09-2020-0037. ahead-of-print(ahead-of-print).

Schmidt, M., Glaser, N., Schmidt, C., Beck, D., Palmer, H., & Lim, M. (2020). Promoting acquisition and generalization of skills for individuals severely impacted by autism using immersive technologies. In B. Hokanson, G. Clinton, A. A. Tawfik, A. Grincewicz, & M. Schmidt (Eds.), *Educational technology beyond content* (pp. 71–84). Springer International Publishing. https://doi.org/10.1007/978-3-030-37254-5_6.

Schmidt, M., Schmidt, C., Glaser, N., Beck, D., Lim, M., & Palmer, H. (2019). Evaluation of a spherical video-based virtual reality intervention designed to teach adaptive skills for adults with autism: A preliminary report. *Interactive Learning Environments*, 1–20. https://doi.org/10.1080/10494820.2019.1579236

Scott, M., Falkmer, M., Girdler, S., & Falkmer, T. (2015). Viewpoints on factors for successful employment for adults with autism spectrum disorder. *PLoS One, 10*(10), Article e0139281.

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

Sharma, S. R., Gonda, X., & Tarazi, F. I. (2018). Autism spectrum disorder: Classification, diagnosis and therapy. *Pharmacology & Therapeutics, 190*, 91–104. https://doi.org/10.1016/j.pharmthera.2018.05.007

Shier, M., Graham, J. R., & Jones, M. E. (2009). Barriers to employment as experienced by disabled people: A qualitative analysis in calgary and Regina, Canada. *Disability & Society, 24*(1), 63–75.

Slater, M., Lotto, B., Arnold, M. M., & Sanchez-Vives, M. V. (2009). How we experience immersive virtual environments: The concept of presence and its measurement. *Anuario de Psicología, 40*, 18.

Smith, K. A., Ayres, K. A., Alexander, J., Ledford, J. R., Shepley, C., & Shepley, S. B. (2016). Initiation and generalization of self-instructional skills in adolescents with autism and intellectual disability. *Journal of Autism and Developmental Disorders, 46*(4), 1196–1209.

Steuer, J. (1992). Defining virtual reality: Dimensions determining telepresence. *Journal of Communication, 42*, 73–93.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization 1. *Journal of Applied Behavior Analysis, 10*(2), 349–367.

Stokes, T. F., & Osnes, P. G. (2016a). An operant pursuit of generalization – republished article. *Behavior Therapy, 47*, 720–732.

Stokes, T. F., & Osnes, P. G. (2016b). An operant pursuit of generalization – republished article. *Behavior Therapy, 47*, 720–732.

Stokes, T. F., & Osnes, P. G. (2016). An operant pursuit of generalization–Republished article. *Behavior Therapy, 47*(5), 720–732.

Strickland, D. (1996). A virtual reality application with autistic children. *Presence: Teleoperators and Virtual Environments, 5*(3), 319–329. https://doi.org/10.1162/pres.1996.5.3.319

Strickland, D. (1997). Virtual reality for the treatment of autism. *Studies in Health Technology and Informatics*, 81–86.

Strickland, D. C., McAllister, D., Coles, C. D., & Osborne, S. (2007). An evolution of virtual reality training designs for children with autism and fetal alcohol spectrum disorders. *Topics in Language Disorders, 27*(3), 226.

Taylor, J. L., Smith, L. E., & Mailick, M. R. (2014). Engagement in vocational activities promotes behavioral development for adults with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 44*(6), 1447–1460. https://doi.org/10.1007/s10803-013-2010-9

Tsirgiotis, J. M., Young, R. L., & Weber, N. (2021). A mixed-methods investigation of diagnostician Sex/Gender-bias and challenges in assessing females for autism spectrum disorder. *Journal of Autism and Developmental Disorders*, 1–16.

Wallace, S., Parsons, S., & Bailey, A. (2017). Self-reported sense of presence and responses to social stimuli by adolescents with autism spectrum disorder in a collaborative virtual reality environment. *Journal of Intellectual & Developmental Disability, 42*(2), 131–141. https://doi.org/10.3109/13668250.2016.1234032

Zener, D. (2019). Journey to diagnosis for women with autism. *Advances in Autism, 5*(1), 2–13.