

The Social Validity of Video Modeling Versus Virtual Reality for Improving Students' Social Communication

Maggie A. Mosher
University of Kansas

Kathleen L. Lane
University of Kansas

Amber L. Rowland
University of Kansas

Adam C. Carreon
Georgia Southern University

Sean J. Smith
University of Kansas

Bruce B. Frey
University of Kansas

Wayne S. Sailor
University of Kansas

Neal M. Kingston
University of Kansas

Haidee A. Jackson
University of Texas Permian Basin

Samantha R. Goldman
University of Kansas

Angela A. Williams
University of Kansas

Ankita Bhattashali
University of Kansas

Abstract

A mixed methods randomized control trial (RCT) was employed to investigate the social validity ratings (i.e., acceptability, appropriateness, and feasibility) of middle-school students using an innovative program of virtual social scenarios called VOISS (Virtual reality (VR) Opportunities to Integrate Social Skills) compared to an evidence-based, video modeling intervention, PEERS (Program for the Education and Enrichment of Relational Skills). Participants within ten classrooms in four states were randomly assigned to VOISS (N=60) and PEERS (N=60). In both conditions, participants experienced an estimated 300 minutes of the intervention spread over two to four months. All participants were given an adapted Children's Intervention Rating Profile (CIRP), the Intervention Appropriateness Measure (IAM), and the Feasibility of Intervention Measure (FIM) to determine their ratings of each intervention's acceptability, feasibility, and appropriateness. Results of statistical significance reveal that VOISS provides an effective and socially valid means of delivering social communication instruction to middle-school students.

The extent to which an intervention is perceived as socially valid significantly influences whether the intervention is selected, implemented, and maintained (Kern & Manz, 2004; Mosher & Carreon, 2021). Social-Emotional-Behavioral (SEB) interventions (SEBI) and evidence-based practices (EBPs) are often ranked with low social validity by adolescents (McCoy et al., 2016, Mosher & Carreon, 2021). Interventions delivered through virtual reality (VR) report increased social validity with this population due to life-like features improving motivation and engagement (Hew & Cheung, 2010; Mikropoulos & Natsis, 2011; Mosher et al., 2024). Despite evidence of positive feelings, there is limited research on the effectiveness of VR-delivered instruction for building SEB competence in students.

Meta-analytic research reveals that explicit SEB instruction in schools (e.g., directly taught and rehearsed) improves elementary students' SEB competencies, particularly when a program's theory, climate, assessments, and progress monitoring are well aligned (Jones et al., 2017). However, findings for comprehensive well-aligned SEBI for adolescents suggest they are less likely, overall, to produce expected gains in competencies (Yeager, 2017). The SEB pressures adolescents face continue to rise. Between 2011 and 2015, emergency room visits related to depression, anxiety, and similar conditions for adolescents in the US rose 28%. Between 2019 and 2021, emergency room visits for suicide attempts increased by 51% for adolescent girls (Richtel, 2021). The National Center for Health Statistics reported an estimate of 6,600 deaths by suicide among Americans, age 10-24 in 2020. It is essential to identify SEBI that utilizes evidence-based practices (EBPs) and are socially valid for adolescents.

The Committee for Children (2019) review found acceptable and applicable evidence-based SEBI mitigates youth suicide risk factors. A public school district in Utah implemented applicable SEBI to all elementary and middle school students and two years later noted decreasing rates of youth substance abuse and suicidality despite an increase of both in neighboring counties with similar demographics (Posamentier et al., 2023). Adolescents must be provided with quality SEBI using EBPs, but it is often difficult to know where to start.

The brain's method of processing emotions during adolescence undergoes a dramatic transformation (Blakemore & Mills, 2014), providing an ideal time for meaningful SEBI. The neural and hormonal changes at the onset of puberty offer a second opportunity for development in all SEB domains (Blakemore & Mills, 2014; Crone & Dahl, 2012). However, finding quality programs to assist adolescents in dealing with SEB struggles and life transitions can be difficult. Most programs tend to be reconstructed initiatives created for younger children and do not provide the flexibility necessary for a dynamic and continuous transaction embedded within the context of the student's cultural environment (Sawchuk, 2021). Technology may help reduce educational barriers as it possesses the following unique capabilities:

- It can provide the flexibility necessary to embark on experiences that are not easily constructed within a classroom.
- It aids students of varying abilities in increasing learning productivity within and outside the classroom.
- It contributes to improving feelings of social acceptance by peers.
- It increases engagement and motivation to learn.

- It mitigates frustration associated with learning.
- It offers opportunities for confidential skill practice.
- It allows the intervention to be tailored to the student (Alghazo & Al-Otaibi, 2016; Glantz et al., 2003; Mosher et al., 2020).

With the increasing comfort of adolescents in using technology-delivered instruction, particularly students with ASD (Kuznekoff & Titsworth, 2013), emerging forms of technology should be further explored as a viable SEBI delivery option.

Social Validity Framework

Traditional EBPs addressing social skill deficits (e.g., role-playing, video modeling, direct instruction) have not been as motivating for adolescents as elementary-age students (McCoy et al., 2016). Taylor et al. (2017) showed that evidence-based SEBI for adolescents has the potential to produce positive long-term outcomes for students with diverse prior experiences, needs, and cultural priorities. SEBI for adolescents is reported to be ineffective unless they go beyond building individual competencies and consider whether the skills, environment, and instruction within the intervention are appropriate and acceptable (Berg et al., 2017; Jennings & Greenberg, 2009). It is necessary, then, to understand ways to assist in making SEBI more interactive and motivating for adolescents.

Determining what aspects of an intervention are appropriate, desired, generalized, and maintained is critical to an intervention's success and is known as social validity (Fox & McEvoy, 1993). The social validity framework provides a measure in which to look at three elements of an intervention: (a) the goals (i.e., importance/justification), (b) the procedures (i.e., appropriate/acceptability), and (c) the outcomes (i.e., meaningful/importance; Armstrong et al., 1997; Kazdin, 1977). Social validity is not something an intervention has or lacks but a multidimensional process consisting of numerous variables, including intervention acceptability and importance (Finney, 1991; Mosher & Carreon, 2021). Social validity is an important predictor of the acceptability of an intervention by participants (Baer et al., 1987). Understanding primary aspects of social validity (i.e., technology preferences, use, knowledge acquisition) within interventions is essential to determining the method most likely to be available, selected, implemented, and maintained by students and their educators (Mosher & Carreon, 2021).

Various forms of immersive learning are currently being used to assist students in experiencing and interacting at all levels of immersion when these real-life experiences are not otherwise accessible (Radianti et al., 2020; Smith et al., 2022). Amongst the most popular and commercially available immersive technologies for learning is virtual reality (VR). VR, by definition, is an artificial or digital environment that can be accessed through a variety of sensory stimuli provided by a computing device (Merriam-Webster, 2022). VR exists as a continuum, with one end allowing participants to interact and experience the simulation with non-immersive screen-based technology to the other end where the participant is fully immersed in the technology environment through head-mounted display technology (Mosher et al., 2024).

VR has recently become more affordable and attainable for student use, leading to the development of applications suitable for schools. Students, particularly

those with disabilities, may benefit from VR features available to schools at low to no cost in academic, behavioral, and social-emotional instruction (Mosher et al., 2022). VR can allow a student to practice, learn, and engage with skills in a safe and authentic environment (Bellani, 2011). Through VR, the environment and skill practice can be replicated in an authentic manner with accuracy, repetition, and individualization (Carreon et al., 2023a). The task of replicating the instruction in a systematic manner in multiple environments without the use of VR can be time consuming and costly (Mosher, 2022). For example, a program can create a situation where a student is bumped into, breaks an iPad, and needs to ask for assistance. This task would be difficult to replicate in person without sacrificing costly technology and the time necessary to enlist and train peers to provide accurate feedback and responses. Therefore, we must determine if innovative forms of technology (e.g., augmented reality, VR, extended reality, generative artificial intelligence, machine learning) have the potential to systematically replicate instruction in a manner acceptable and easily usable by students while being beneficial to improving student outcomes.

The Social Validity of VR for Intervention Delivery

Two separate literature reviews (Mosher & Carreon, 2021; Mosher et al., 2022) explored the social validity of VR to provide systematic and individualized social skill instruction to students with autism spectrum disorders (ASD). These systematic reviews pointed to virtual technology improving social skills for students with ASD. However, the reviews also made apparent the need for conclusive research on the ability of VR to improve the targeted social skills of students. Current VR research tends to rely on perceived improvements without considering quantitative measures (Howard & Gutworth, 2020). VR literature reviews also point to the need to understand the preferences of students and implementers on the choice of technology to deliver the intervention (Mosher et al., 2022), as this preference is shown to influence the intervention's continued use (Kim et al., 2020; Mosher & Carreon, 2021). The prior lack of research in these areas is partly due to the limited number of virtual technologies designed to teach SEB skills to middle school students and the absence of the ability to use the same intervention within varying technologies.

A strong correlation exists between beliefs about an intervention and the use of that intervention (Hew & Brush, 2007; Mosher, 2022). VR offers significant advantages for enhancing classroom learning, due to the reported positive beliefs from VR users about technology's content delivery (Carreon et al., 2022; Hew & Cheung, 2010; Mikropoulos & Natsis, 2011; Rajendran, 2013). Prior to VR's use in classrooms to improve middle school students' SEB competencies, the social validity and efficacy of such an intervention versus a research-based intervention would be beneficial.

Previous research shows that VR-delivered interventions can be implemented with minimal teacher preparation and professional development. This enables educators to provide tailored, real-world, standardized interventions in controlled environments, allowing students to better personalize their learning experience (Charlop-Christy & Daneshvar, 2003; Glantz et al., 2003). Prior studies of the social

validity of video modeling also show highly favorable responses to VR interventions by student participants (King et al., 2014). Research reveals students with and at risk for social-behavioral difficulties have greater social validity toward interventions when the intervention takes up little classroom time (i.e., around 30 minutes a session) and provides a way for students to covertly self-regulate in a manner that does not draw unwanted attention (Felver et al., 2017). Virtual-reality Opportunities to Integrate Social Skills (VOISS) and the Program for the Education and Enrichment of Relational Skills (PEERS) both allow for covert self-regulation and occupy little classroom time. Therefore, it is predicted that the social validity perceptions (i.e., acceptability, feasibility, and appropriateness) of students will remain high for both PEERS and VOISS and that only a slight increase may be shown in the VOISS intervention over PEERS in acceptability, due to the novelty of a game-like VR program delivering the instruction instead of a teacher or peer delivering instruction, as is common in many current SEB instructional models.

SEBI for adolescents has not been found to be as effective as interventions targeting earlier ages (Heckman & Kautz, 2012). Rarely do middle school students report SEBI and SEB programs to be motivating or effective (Yeager, 2017). Investigating the social validity of a VR-delivered intervention versus an evidence-based technology-delivered intervention may illuminate potential barriers (e.g., perceived ease of use, motivation, direct versus indirect instruction) to the intervention's use and effectiveness. Such information may be useful in supporting current SEB practices within middle schools as well as in shaping future SEBI. Therefore, this study seeks to understand further the social validity of VR for delivering SEBI to adolescents by answering the following research questions:

1. Is there a statistically significant difference in the pre-and post-acceptability ratings of a virtual reality-based social skill intervention versus an evidence-based video modeling social skill intervention for middle school students?
2. Is there a statistically significant difference in middle school student ratings of feasibility of a virtual reality-based social skill intervention versus an evidence-based video modeling social skill intervention?
3. Is there a statistically significant difference in middle school student ratings of appropriateness of a virtual reality-based social skill intervention versus an evidence-based video modeling social skill intervention?

Methods

This study was designed to understand and compare the social validity of a VR intervention, VOISS and the highly evidence-based PEERS intervention for delivering SEBI to middle school-aged students. A mixed methods group experimental randomized control trial was conducted with four stages: pretests, practice, intervention, and posttests. These stages are expanded upon in the outlined sections below.

Participants

The participants for this study were students with and without disabilities attending middle school in public, private, and charter institutions across the United States. Participants were recruited via email and through a call presented at four national and regional educators' conferences. All participants had to meet the following criteria to participate: (a) be middle school-aged (10-15), (b) be identified by an educator or practitioner to need expressive or pragmatic social skills determined by a reliable and valid assessment (e.g., Clinical Assessment of Pragmatics), (c) be able to complete perception rating scales, (d) be willing to participate for the duration with follow-up, (e) be willing to use technology for intervention, (f) have an educator to oversee the technology usage, (g) have an educator willing to complete rating scales about student progress, and (h) have the language (i.e., English) and reading ability (i.e., third grade) to participate. Disability diagnosis, if any, was not a prerequisite for participation, but it was documented. Parental informed consent and verbal student assent were obtained prior to intervention.

A total of 152 participants were recruited. After applying the inclusionary criteria, 120 participants identified as having a pragmatic social skill deficit remained. Table 1 provides detailed characteristics of these participants.

Table 1*Participant Characteristics*

Characteristics	Total (N=120)	Percentage
Student Age		
10 years old	48	33.1%
11 years old	51	35.2%
12 years old	14	9.7%
13 years old	7	4.8%
Gender		
Female	54	45%
Male	66	55%
Race & Ethnicity		
African American	7	4.8%
American Indian/Alaska Native	6	4.1%
Asian	11	7.6%
Hispanic/Latino	12	8.3%
More Than One Race	1	0.7%
White	85	58.6%
Diagnosed Disability		
Attention Deficit Hyperactivity Disorder	7	5.8%
Autism	15	12.4%
<i>Autism Level 1</i>	10	8.3%
<i>Autism Level 2</i>	4	3.3%
<i>Autism Level 3</i>	1	0.8%
Dual Diagnosis	24	19.9%
<i>ADHD & LD</i>	7	5.8%
<i>ADHD, AD, & LD</i>	1	0.8%
<i>ADHD, Autism L1, & LD</i>	10	8.3%
<i>ADHD, Autism L1, Depression, & OCD</i>	4	3.3%
<i>BD, LD, & SPD</i>	2	1.7%
Intellectual Disability	3	2.5%
Learning Disability	8	6.7%
No Known Diagnosis	63	52.5%
Student Plan Type		
504	1	0.8%
IEP	25	20.8%
IEP & Behavior Intervention Plan	3	2.5%
IEP & Social Skills Plan (SSP)	5	4.1%
SSP & Pull-Out Social Intervention	17	14.2%
Student Improvement Team Plan	7	5.8%
No Formal Plan	62	51.7%

Research Design

This study utilized a randomized control trial design to evaluate the social validity of VOISS and PEERS for the SEB skill of expressive communication. Once all participants were recruited, they were randomized into matched pairs. Pairs were matched based on the following hierarchical criteria, with priority taken for the higher criteria: (a) student's primary teacher was identical, to ensure the same instruction throughout the school day by each match; (b) similar teacher ratings of student social skill performance (i.e., Clinical Evaluation of Language Fundamentals-Pragmatic Profile (CELF-5 PP)); (c) similar student ratings of their social skill performance (i.e., CELF-5 PP); (d) scores of student answers to the Social Communication Knowledge Questions (SCKQ); and (e) student demographic information (i.e., gender, age,

disability, experience). Following the pairing of students, each was randomly assigned by statistical software (SPSS) to either the VOISS or PEERS intervention. These groups were analyzed in SPSS to ensure no statistical variance between the pretest expressive communication knowledge and application ratings for each student pair. After all students were paired appropriately, an additional assessment of group traits (i.e., age, race, gender, educational plan, diagnosed disability) were considered to ensure proportional groups. Priority was given to gender and age, as eliminating variance amongst 120 participants was not possible. Each group completed a pretest and presurvey before intervention. Each group then completed a one-week training session to allow for independent understanding and navigation of the individual interventions. Following training, participants were expected to complete intervention sessions independently. After intervention, students completed the post knowledge test and post-survey measures.

Setting and Materials

The study was conducted in multiple settings to comply with school COVID protocols. However, all participants assigned to either group received the given intervention in the same room, at the same time a day, from the same device, and with the same teacher. Each student completed all sessions with the same researcher on a virtual video conference (i.e., Zoom) with validity coders randomly assigned to sessions. All sessions occurred for the participants in their typical classroom, at their typically assigned tables, with their one-to-one student-issued device (i.e., Apple iPad or Chromebook), a large screen displaying the video conference software, and a teacher desktop with the ability to speak one-on-one with any student needing assistance. All participants had the same teacher who provided SEBI to them across all sessions. The sessions all occurred during the students' typically scheduled SEBI time. All interruptions (i.e., field trips, school events) that may cause a lapse in participation were controlled by randomly assigning paired participants in the same school with the same teacher. The study began in October 2022 and ended in March 2023.

Participants spent two to three sessions, for a total of 90 minutes, within a two-week period being trained to navigate both the technology devices (i.e., Chromebook, iPad) and application (i.e., PEERS, VOISS). With 20 years of SEBI experience, the first author implemented all training sessions, including pre-and-post-assessment questions. All questions were responded to via online survey software (i.e., Qualtrics). Intervention sessions were conducted during the participating school's SEBI time and ranged from 20-60 minutes per day. Students experienced one to four sessions per week, varying by participant schedule. The varying intervention length occurred because the intervention was designed and intended to be delivered during the teacher's normal instructional time. No additional time was utilized for intervention. In total, all participants received an estimated five hours (300 minutes) of intervention time over a period of two to four months.

VOISS

VOISS is a stand-alone, interactive, social skills, VR application designed to enhance the social skills of participants (Carreon et al., 2023b). VOISS was

developed by experts in education, special education, and SEBI. VOISS was selected due to its cross-platform availability and the reliability data supporting its SEB skill competency development. Therefore, participants could receive randomly assigned intervention on their familiar device. VOISS is available on many popular devices that run Android (i.e., Chromebooks, Android Phones), Apple iOS (i.e., iPad, iPhone), Meta/Oculus (i.e., Quest 2/3), and Windows-based laptops.

In VOISS, participants are presented with scenarios wherein they interact with similar-age peer avatars and familiar school adults (i.e., teachers, administrators, cafeteria workers, and paraprofessionals). Participants navigate the scenarios and multiple locations using a touch screen or pointing device. Within scenarios, participants are presented with an authentic social skill situation and need to use critical thinking skills to complete the scenario. To move to the next situation within each scenario, participants must select correct multiple-choice responses, move to correct locations, or orally respond correctly to a situation. In the event of an incorrect response, a natural consequence is displayed, based on the selected choice, and narration within VOISS uses direct instruction to reteach the skill and elicit an appropriate response. Participants complete the scenarios by obtaining all correct responses. Each scenario varies in length, making completion of scenarios dependent on student competence in the targeted SEB skill.

PEERS

PEERS is an evidence-based social skill intervention for adolescents with ASD, attention deficit hyperactivity disorder (ADHD), anxiety, depression, and others who are at risk for challenges in SEB competency development. PEERS was selected due to its availability across platforms, its development by experts in education, special education, and SEBI, and the extensive research supporting its ability to improve social competencies through video modeling. PEERS was developed by Dr. Elizabeth Laugeson in 2005 at UCLA and has since been used in over 100 countries. The PEERS videos used in this study are validated through research and are accessible via the same platforms as VOISS (i.e., Chromebook, iPad). Many of the PEERS videos are available for free at <https://www.semel.ucla.edu/peers/resources/role-play-videos>. Although the videos are labeled “role play,” this study only utilized the video-modeling portion. It was determined that only the video modeling, not the role play portion, followed the fidelity checklist with acceptable procedures to be considered an EBP. Videos utilized in this study that were not free were obtained from the PEERS trainer with a curriculum guide to inform instruction.

In PEERS, participants are presented with social skill scenarios via video modeling. Participants watch adolescents and adults take part in problem-solving authentic social skill scenarios. Like VOISS, participants participate in various environments such as classrooms, libraries, and offices. During the videos, participants watch a video with both examples and non-examples with associated natural consequences for each. After watching the videos, participants are provided an opportunity to imitate the task seen in the video through recall or guided questioning of the curriculum. Students then discuss the scenario with the teacher and design their video. The scenarios presented in PEERS were selected due to the

nearly identical targeted SEB competency instructed within VOISS, to ensure instruction of the same measured skill.

Technology to Deliver SEBI

Each participant experiences the interventions of VOISS and PEERS through their typical school-given device. These devices included multiple models of iPads and Chromebooks. While there were multiple models, all devices used one of two operating systems (i.e., iOS, Android) and ran the two-intervention software identically across platforms. Matched peers used the same platform in identical classroom settings, with 102 students accessing the interventions via a Chromebook and 18 students via an Apple iPad. Each device provided identical access to visuals on the screen and audio from the included speaker. It was decided to use the device the student typically used versus a new or chosen device to reduce the time needed to train and familiarize the participant with the device, decrease results from a technology novelty effect, and ensure participants had access to their accessibility needs.

Social-Emotional-Behavioral Skills

A variety of SEB skills are available in the VOISS and PEERS interventions. To compare the effectiveness reliably, we determined that the VOISS Expressive Communication (EC) domain and the PEERS Social Communication (SC) domain were compatible matches. The VOISS EC domain contained 24 EC skills and 26 scenarios. These scenarios were sent to four specialists in expressive communication and SEB skills (i.e., special educators who provide SEBI and speech-language pathologists) to identify and exclude skills that build or may be vulnerable to pretest effects, history, and/or maturation. A total of 22 SEB skills and 24 scenarios were recommended. These skills were then sent to the same four experts for their alignment with the PEERS videos. It was determined that 20 skills within VOISS and PEERS covered identical skills. The accompanying scenarios and PEERS videos were identified and utilized for intervention and comparison of effects. These same skills were also the skills assessed in the Social Communication Knowledge Questions (used as a screener for those deficient in social skills) and the Clinical Evaluation of Language Fundamentals-Pragmatic profile (measures verbal and non-verbal contextual communication).

Social Validity Measures

Individual surveys containing rating scales were selected as the instrument for social validity data collection, rather than focus groups or interviews, because they allow students to share their views about the intervention without the influence of outside voices, which research shows causes less biased responses than responses given when in a group of peers or directly to a researcher (Creswell, 2002). Rating scales were chosen over other instruments because subjective measurements are more appropriate to assess social acceptance, feasibility, and appropriateness (Kazdin, 1977; Wolf, 1978). Surveys were also selected because they produce information about beliefs and attitudes, which are otherwise difficult to measure using observational techniques (McIntyre, 1999).

There are a number of empirically validated scales for measuring social validity, such as the Treatment Evaluation Inventory (TEI; Kazdin 1980), Intervention Rating Profile-20 (IRP-20; Witt & Marstens 1983); Children's Intervention Rating Profile (CIRP; Witt & Elliott 1985); Behavior Intervention Rating Scale (BIRS; Von Brock & Elliott 1987); Treatment Acceptability Rating Form—Revised (TARF-R; Reimers et al. 1992); and the Abbreviated Acceptability Rating Profile (AARP; Tarnowski & Simonian 1992). These rating scales are primarily developed as a questionnaire with a Likert-type scale completed by either the parent or teacher. An adaptation of the Intervention Rating Profile (Adapted IRP; Lane et al., 2015), similar to the IRP-15 (a brief version of the IRP; Martens et al., 1985), was first chosen over other acceptability rating forms, because the IRP is commonly used in educational settings, assesses acceptability of interventions, determines risks, and allows for a measure on acceptability of length of treatment as well as effects on the educator and fellow students.

The targeted questions in this study are on students' feelings of social validity rather than educators' feelings. This caused the Children's Intervention Rating Profile (CIRP; Witt & Elliott, 1985) to be considered. It was noted that this measure was created for the acceptability of an intervention. Interventions for adolescence are reported to be ineffective unless they consider whether the skills, environment, and instruction are appropriate and acceptable (Berg et al., 2017; Jennings & Greenberg, 2009). An appropriateness measure was determined to be needed in addition to acceptance. Finally, interventions that are not feasible are not likely to be maintained (Proctor et al., 2011). This is particularly true when considering interventions delivered through technology (Lorenzo et al., 2016). Therefore, a feasibility measure was also included. This led to selecting three areas of needed measurements: (a) acceptability, (b) appropriateness, and (c) feasibility in the measures outlined below.

Adapted Children's Intervention Rating Profile (RQ 1)

The Adapted Children's Intervention Rating Profile (Germer et al., 2011; Lane et al., 2015) was chosen as the student measure of acceptability because it was written at a third grade reading level to allow students to complete the intervention ratings. The Adapted CIRP was modified slightly from the CIRP (Witt & Elliott, 1985) to maintain the readability, validity, and reliability level of the CIRP while modifying vocabulary to better fit current school-age raters. The underlying construct of acceptability measured within the Adapted CIRP was well-defined and supported by a comprehensive theoretical framework and prior research. The definition of acceptability to be measured is how well an intervention will be received or is received by a target person or population and the extent to which the intervention meets the needs of the target population and context (Briesch et al., 2013; Lane et al., 2015; Martens et al., 1985).

The CIRP was additionally modified by authors of the study based on research behind visuals. This change included the addition of pictures accompanying the ratings to thumbs up and thumbs down, rather than just the original numbers or happy and sad face, to gain a more accurate picture of agreement and disagreement rather

than if the question made the student happy or sad. Also, a word was placed with every number as students with disabilities in the age group in past assessments required additional vocabulary to understand the difference between a 4 and a 5. Students completed the measure at Time 1 (the session prior to the start of the intervention) and Time 2 (the session immediately after the end of the intervention). The measure contained seven questions on a 5-point Likert scale (1=Strongly Disagree, 5=Strongly Agree) and was created to assist in determining whether an intervention should be selected for use within a classroom. Total scores range from seven to 42 with scores of 24.5 or higher considered acceptable (Turco & Elliot, 1986). Higher total scores indicate greater levels of intervention acceptability.

Adapted Intervention Appropriateness Measure (RQ 2)

After considering multiple feasibility and appropriateness surveys, the Intervention Appropriateness Measure (IAM), and Feasibility of Intervention Measure (FIM) were selected due to their ability to accurately assess appropriateness and feasibility within the targeted population as well as the survey length necessary for a thorough understanding while considering time and attention span of the target population. The IAM and FIM contain response selection on a Likert scale, which ranges from completely disagree (i.e., score 1) to completely agree (i.e., score 5) in which higher scores indicate a greater sense of appropriateness or feasibility toward the intervention (Weiner et al., 2017). The scales have a Flesch reading ease score of 95.15, which is a fifth-grade reading level. There are no specialized skills or training needed to administer, score, or analyze the IAM or FIM (Weiner et al., 2017). The combined measures take less than five minutes to complete. The IAM and FIM received the highest validity and reliability ratings of all student rating scales with a fifth-grade reading level or below according to the Implementation Outcome Repository. They were the chosen methods of middle school student evaluation measures by *Program Fit Measures*, a California Evidence-Based Clearinghouse for Child Welfare.

Appropriateness is the perceived fit, relevance, or alignment of an intervention or practice in a specific context for a specific issue with the expectation or current role (Weiner et al., 2017). Appropriateness is a necessary measure to attain whether stakeholders' feelings about the intervention align with their expectations and current needs (Proctor et al., 2011). Appropriateness is a similar construct to acceptability but remains distinct in that it can ascertain resistance in implementing or partaking in an intervention by stakeholders. For example, an intervention may be suitable or appropriate for a particular need, but the intervention's features may make the intervention unacceptable to the rater (e.g., too much deviation from the original intervention method intent; Proctor et al., 2011). The Intervention Appropriateness Measure (IAM) is a four-item scale with excellent internal consistency and strong psychometric properties. Cut-off scores for interpretation of FIM results are not yet available; however, higher scores indicate greater feasibility. Still, this survey was selected for use as an accurate measure of acceptability with the highest scored

scale of the three social validity scales developed by Weiner et al. (2017) with a Cronbach's alpha of 0.91.

The Feasibility of Intervention Measure (RQ3)

Feasibility is the extent to which an intervention or practice can be or has been successfully implemented within a given context (Weiner et al., 2017). Feasibility is connected to the construct of appropriateness but varies conceptually (Weiner et al. 2017). For example, an intervention may be appropriate (i.e., relevant in a classroom) but at the same time not feasible because the classroom setting may not allow for access to the time necessary to complete the intervention (Proctor et al., 2011). Feasibility assists in measuring both the practical component of the intervention implementation (i.e., how easily the intervention can be implemented) in each context in which it will be delivered by the student and those assisting the student. The Feasibility of Intervention Measure (FIM) is a four-item scale with good internal consistency and reliability with a Cronbach's alpha score of 0.89 (Weiner et al., 2017). Cut-off scores for interpretation of FIM results are not yet available; however, higher scores indicate greater feasibility.

Survey Implementation Reasoning

Acceptability is believed to be a dynamic concept, which can change within a short period of time. For this reason, acceptability ratings may vary before and after intervention implementation. As a result, the student acceptability measures will be given at Time 1 and Time 2 (i.e., pre-and-post-intervention). However, appropriateness and feasibility are most effectively assessed retrospectively to allow raters to have experiences to draw on to form their opinions (Proctor et al., 2011). Therefore, IAM and FIM will only be given at Time 2. See Appendix F for a full list of the items on each of the rating scales.

Written surveys can be subject to coverage error and item nonresponse, where some questions can be inadvertently or intentionally skipped (Salant et al., 1994). To resolve the possibility of coverage error, the questions of the survey were electronically randomized by classroom to help limit biased context results and ensure that if people quit partway through the survey, the data collected would not be substantially affected. Randomization also limited the possibility of order influencing the participants' responses.

The surveys were distributed to all matched participants within the same timeframe to ensure the surveys do not reflect seasonal or temporal differences. Data was analyzed immediately following collection. Qualtrics (Provo, UT) was chosen for the survey platform because of its accessibility, data security, and randomization features. Experts were consulted to ensure appropriate language and response options as well as to assess whether the surveys measured the target construct (Browne & Keeley, 1998; Fowler, 1995).

Pretest

Before beginning the intervention stage, students completed an Adapted CIRP through Qualtrics on their preferred device (Chromebook, iPad) as well as a knowledge-based test to determine pre-knowledge scores. The test and surveys were read aloud to the student by the same person and in the same classroom with their matched peers.

Posttest

The post surveys (IAM, FIM, and CIRP) and knowledge assessment were presented to students through Qualtrics on their preferred device (Chromebook, iPad). Effect size estimates for each intervention condition were calculated using the partial eta squared effect (Gray & Kinnear, 2012) from ANOVA Repeated Measure and Cohen's d (1988) from the independent samples t test. Partial eta square effect sizes are categorized as small (.01), medium (.06) and large (.14 or higher). Cohen's d effect sizes are categorized as small (0.2), medium (0.5) and large (0.8 or higher).

Results

A 2-by-2 mixed-design analysis of variance (ANOVA) was performed to evaluate whether there were significant effects between pre-and-post-CIRP survey ratings and pre-and- post-knowledge assessments. An independent samples t test was performed on measures with post ratings only (i.e., IAM and FIM) using Levene's test for equality of variance prior to calculating effect sizes for each intervention condition to determine if ratings between groups were statistically significant. Finally, statistical significance ($p < .05$) was calculated for all measured variables.

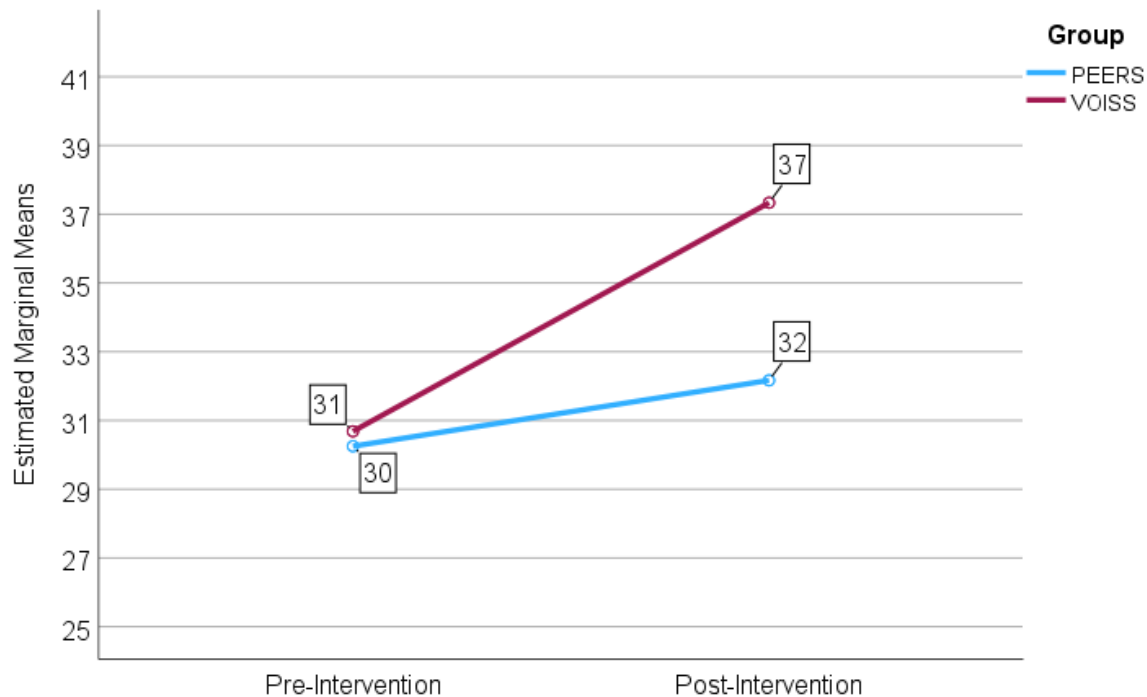
Acceptability. To evaluate whether there were significant effects between pre-and-post- CIRP, an ANOVA was performed to answer the following question: Is there a difference in the acceptability ratings of a VR based social skill intervention (VOISS) versus an evidence-based video modeling social skill intervention (PEERS) for middle school students? It was predicted that acceptability of students will remain high for both PEERS and VOISS and only a slight increase may be shown in the VOISS intervention due to the novelty of VR.

As predicted, the repeated measures analysis of variance with student CIRP ratings of intervention acceptability as the dependent variable found a significant effect ($F[1, 118] = 46.54, p < .001$) with a large effect size (partial eta squared of 0.28). Both interventions were found highly acceptable to students pre ($M = 30.47$) and post intervention ($M = 34.75$). There was also a significant interaction when looking at each group ($F[1, 118] = 14.21, p < .001$) revealing that the VOISS intervention was significantly more acceptable than PEERS with a medium effect size (partial eta squared of 0.11) pre-to-post ratings. When isolating the interventions, those receiving the VOISS intervention ($F[1,59] = 40.17, p < .001$) increased their ratings of intervention acceptability significantly with a very large effect size ($\eta^2 = .41$). Those receiving the PEERS intervention also increased their ratings of intervention acceptability significantly ($F[1,59] = 6.14, p = .016$) with a medium effect size ($\eta^2 = .09$). The prediction that both interventions would be seen

acceptable by students was accurate as was the prediction that VOISS would be rated more acceptable than PEERS by students. See Figure 1 for student CIRP ratings of acceptability.

Figure 1

Student CIRP Ratings of Intervention Acceptability



The largest increase in ratings on the CIRP pretest to posttest was for the rating on the question “this program could help other kids too,” which started with a 63% completely agree response and rose to an 85% completely agree response. Although acceptability remained high for both interventions, the mean acceptability for specific questions for the PEERS intervention saw a decrease in three questions. After the PEERS intervention, the ratings of liking being in the program, believing the program will be helpful in school performance, and believing this program is the best method for the participant went down. Table 2 provides the questions and the mean responses for each group. Within Table 3, in bold, is the mean difference pretest to posttest for each question by group.

Table 2

CIRP Question Response Means Pretest and Posttest with Mean Difference for Each Group

	The program we will use sounds fair. ...we used was fair.			This program could help other kids too. ...will help other kids, too.			I think I will like being in this program. I liked the program we used.			I think being in this program will help me do better in school. Being in this program helped me...		
PEER S	4.50	4.80	0.30	4.07	4.93	0.87	4.33	4.30	-0.03	4.22	3.98	-0.23
VOISS	4.63	5.42	0.78	4.27	5.52	1.25	4.80	5.28	0.48	4.67	5.05	0.38
Total	4.57	5.11	0.54	4.17	5.23	1.06	4.57	4.79	0.23	4.44	4.52	0.08
Reverse Score Rating Report Here*	I think my teacher will be (was) too harsh on me. ...was too harsh on me.			Being in this program may cause problems with my friends. ...caused problems with my friends.			There are better ways to teach me. ...were better ways to teach me.			<i>Total Pre</i>	<i>Total Post</i>	<i>Mean Difference</i>
PEER S	4.50	5.13	0.63	4.53	5.40	0.87	3.87	3.62	-0.25	30.02	32.17	2.15
VOISS	4.72	5.55	0.83	4.65	5.63	0.98	4.23	4.88	0.65	31.97	37.33	5.36
Total	4.61	5.34	0.73	4.59	5.52	0.93	4.05	4.25	0.20	31.00	34.75	3.75

*Questions in the bottom portion of the table show scores after reverse scoring. For example, "I think my teacher will be too harsh on me" increase pre to post in bold shows that they are less likely to believe their teacher will be harsh on them after intervention.

Appropriateness. Inspection of Q-Q Plots revealed that IAM scores were normally distributed for both groups and that there was homogeneity of variance as assessed by Levene's Test for Equality of Variances. Therefore, an independent samples *t* test was performed on the data with a 95% confidence interval (CI) for the mean difference to answer the following question Is there a difference in appropriateness ratings between interventions? It was predicted that both groups would indicate high appropriateness ratings for the PEERS and VOISS interventions. It was found that appropriateness ratings interaction was statistically significant with a large effect size ($t[118] = 5.44, p < .001, d = 0.99$). Middle school students' ratings of appropriateness for VOISS ($M = 18.22$) were significantly higher than those for PEERS ($M = 14.53$).

The prediction that both interventions would be seen appropriate by students was not accurate. Student participants rated VOISS as "completely agree" on 91 to 92% of questions on acceptability with "the program seems suitable" as the highest rated question. Student participants rated PEERS as "completely agree" on only 66 to 81% of questions on acceptability. The areas which student participants did not find acceptable in relation to the PEERS intervention were on whether the intervention seemed "fitting" and was "a good match" to their wants and needs. Table 3 provides

the means and standard deviations for each intervention group on all social validity measures administered.

Table 3*Mean Acceptability, Appropriateness, and Feasibility of the Interventions*

Pretest	Group	Mean	SD	Posttests	Group	Mean	SD
Pre	PEERS	30.25	5.739	IAM ^b	PEERS	14.53	3.92
CIRP ^a	VOISS	30.68	5.369		VOISS	18.22	3.48
	<i>Total</i>	<i>30.47</i>	<i>5.538</i>		<i>Total</i>	<i>16.38</i>	<i>3.7</i>
				FIM ^b	PEERS	18.45	3.31
Posttests	Group	Mean	SD		VOISS	18.82	2.00
Post	PEERS	32.17	4.396		<i>Total</i>	<i>18.64</i>	<i>2.66</i>
CIRP ^a	VOISS	37.33	4.725				
	<i>Total</i>	<i>34.75</i>	<i>5.233</i>				

^a Total scores range from 7 to 42 with scores of 24.5 or higher considered acceptable (Turco & Elliot, 1986)

^b Total scores range from 4 to 20 with higher scores considered higher social validity (Weiner et al., 2017).

Feasibility. Inspection of Q-Q Plots revealed that FIM scores were normally distributed for both groups and there was homogeneity of variance as assessed by Levene's Test for Equality of Variances. Therefore, an independent samples *t* test was performed on the data with a 95% confidence interval (CI) for the mean difference to answer the following question: Is there a difference in middle school student ratings of feasibility of a VR based social skill intervention (VOISS) versus an evidence-based video modeling social skill intervention (PEERS)? It was predicted that the feasibility of students toward both interventions would be high. Higher scores on the FIM indicate greater feasibility. It was found that both interventions were highly feasible, with a mean score between 18 and 19 out of 20 for both intervention groups. One intervention was not statistically different than the other intervention in ratings of feasibility ($t[118] = 0.73, p = 0.465, d = 0.13$). Both interventions received between 91 to 98% "completely agree" responses to feasibility questions. The VOISS intervention had the highest ratings on the question "the program seems easy to use" at 98% of participants giving this question a 5 rating of "completely agree." There were no neutral or negative ratings on "the program seems easy to use" and "the program seems possible" for the VOISS intervention. See Table 4 for mean responses to feasibility and appropriateness questions.

Table 4*Mean Responses to Intervention Feasibility and Appropriateness Questions*

IAM	The program seems fitting.		The program seems suitable.		The program seems applicable.		The program seems like a good match.	
VOISS	â:273 <i>M: 5</i>	91% ca 5% cd	â:275 <i>M:5</i>	92% ca 3% cd	â:273 <i>M: 5</i>	91% ca 7% cd	â:272 <i>M: 5</i>	91% ca 7% cd
PEERS	209 <i>M:2</i>	69% ca 3% cd	222 <i>M:4</i>	74% ca 3% cd	244 <i>M:4</i>	81% ca 0 cd	197 <i>M:2</i>	66% ca 5% cd
FIM	The program seems implementable.		The program seems possible.		The program seems doable.		The program seems easy to use.	
VOISS	â:273 <i>M: 5</i>	91% ca 3% cd	â:288 <i>M: 5</i>	96% ca 0 cd	â:273 <i>M: 5</i>	91% ca 5% cd	â:295 <i>M: 5</i>	98% ca 0 cd
PEERS	â:272 <i>M:5</i>	91% ca 5% cd	â:282 <i>M:5</i>	94% ca 5% cd	â:274 <i>M:5</i>	91% ca 5% cd	â:279 <i>M:5</i>	93% ca 5% cd

â60: Raw score out of 300 possible points; % sa: Percent of students rating a 5 “completely agree” on this question

% sd: Percent of students rating a 1 “completely disagree” on this question; *M*: Closest mean rating

Efficacy. Although the effectiveness of the intervention was not a research question

within this portion of the study, it is helpful to understand that both interventions were found effective in improving student knowledge of expressive communication skills.

This is an important finding as it allows us to state that we are looking at the social validity of two interventions, which were both found to be significantly effective in improving expressive communication knowledge, an essential aspect of all SEBI.

The expressive communication knowledge test was analyzed before and after the intervention and showed a significant interaction ($F[1, 118] = 46.45, p < .001$) with a large effect size (partial eta squared of 0.28). Additionally, the interaction between the groups was significant ($F[1, 118] = 235.9, p < .001$) with a very large effect size (partial eta squared of 0.67). After the intervention, the social communication knowledge means increased by 9.4 points, representing a 24% improvement compared to the preintervention scores.

Discussion

Adolescents often prefer technology-based interaction to address areas of social communication weakness over face-to-face (Sweeney et al, 2019). However, there is limited research as to whether a VR intervention that improves social communication is acceptable, appropriate, and feasible for middle school students with varying disabilities and from a variety of backgrounds. This study examined the social validity of a VR intervention for social communication knowledge acquisition and application. Study findings indicated high ratings of acceptability, appropriateness, and feasibility for the VR intervention among middle school students. This finding is consistent with previous research revealing high social validity of VR interventions presented through iPads and Chromebooks with adolescent students (Mosher et al., 2022; Lozano-Álvarez et al., 2023). Interestingly, the acceptability and feasibility of PEERS was also high. This is in line with research,

which reveals students with and at-risk for social-behavioral difficulties have greater acceptance of interventions when the intervention takes up little classroom time (i.e., less than 30-min a session), is presented through technology (Wong et al., 2020), and does not draw unwanted attention (Felder et al., 2017). The acceptability and appropriateness of the VOISS intervention was significantly higher than PEERS, a program known for being enjoyed and valued by adolescents (Gilmore et al., 2023).

Rating results showed that although both interventions were found acceptable and feasible, PEERS was not found appropriate by several adolescents, whereas VOISS was found both appropriate and accessible. This finding should be explored further through mixed methods research to understand the reasons for the high levels of feasibility, appropriateness, and acceptability for students receiving the VOISS intervention over those receiving PEERS.

Social Validity of a VR Intervention to Improve Social Communication Skills

Social validity is a critical component of social communication interventions (Carter & Wheeler, 2019; Hansen et al., 1989). Study findings agree with Halabi et al. (2017), who found VR interventions not only improve skill performance for students with pragmatic delays, but also that they have greater acceptability than other instructional methods. The largest increase in acceptability ratings for both interventions was on the question of agreement as to whether this program could help other kids, which started with a 63% “complete agreement rating” and rose to 85% rating “complete agreement” to the question. This suggests that students recognize benefits after the interventions they had not expected before the interventions. Although acceptability remained high for both interventions, the mean acceptability for three questions for the PEERS intervention saw a decrease (i.e., liking being in the program, believing the program will be helpful in school performance, and believing the program is the best method for the participant). This may indicate that students felt less favorably about aspects of the PEERS intervention, particularly related to the interventions’ helpfulness and fit, than they did prior to intervention implementation. The term “fit” within the acceptability scale is also similar to terms used in the appropriateness scale, which found the PEERS intervention ratings substantially lower than VOISS. In future research, it would be helpful to conduct a year-long study utilizing all the additional aspects of both programs beyond the social narratives and video modeling (i.e., PEERS intervention’s role plays and parent generalization support strategies and VOISS intervention’s activities throughout SEB domains and teacher generalization tactics) to determine how this may influence student acceptability and appropriateness ratings.

Appropriateness ratings of the VOISS intervention in comparison to the PEERS intervention was statistically significant with a large effect size. Since both interventions teach the same expressive communication skills and are delivered through the same preferred device to randomly matched peers, this finding suggests an aspect of the intervention (e.g., representation of cultures, method of breaking down skills, response options), rather than the skills themselves or the delivery device, may be the cause. The appropriateness ratings by student participants on individual questions for PEERS was only 66 to 81% in “complete agreement”

compared to the VOISS appropriateness “complete agreement” ratings in the 91 to 92% range. This finding should be investigated further, particularly considering the comments discussed in the acceptability ratings. This finding also raises the question as to whether an intervention can be considered acceptable by middle school students (e.g., convenience, ease of use, meets needs) but not appropriate (e.g., fitting, a good match, best option).

Both interventions were reported as highly feasible, with a mean score between 18 and 19 out of 20. One intervention was not statistically different than the other in feasibility, as both interventions received between 91 to 98% “completely agree” responses to all feasibility questions. Although there were a couple of students who rated some aspects of PEERS as neutral or not feasible, there were no neutral or negative ratings for the VOISS intervention on “the program seems easy to use” and “the program seems possible.” This reveals that both interventions have high ratings for ease of use. Future research should consider if intervention feasibility for students may be higher when the technology delivering the intervention is familiar to students. This knowledge would be impactful for curriculum developers, as understanding what improves the successful implementation of an intervention within a given context is vital for intervention implementation and maintenance (Weiner et al., 2017).

Prior research shows video modeling to be a highly favorable intervention for students (King et al., 2014). Yet, VOISS was rated as significantly more acceptable and appropriate than PEERS. Some researchers attribute greater acceptance of VR instructional programs over other interventions to be due to the pressure-free practice environment within VR, reducing the stress for students (Pizzoli et al., 2019), while others attribute high acceptability to the “real-life” feeling within VR (Halabi et al., 2017). It would be advantageous to understand which aspects of interventions improve acceptability for students who need assistance building SEB competencies.

Although not a part of the original questions presented for examination, the comments section at the end of participants’ CIRP surveys suggest the content of the intervention and how it is presented may be just as important as the element of realness and reduced stress. Two comments, coming from students who rated the highest acceptability and applicability, one from each intervention group, provided information on the benefit found in the way the instruction was given. A student in the VOISS group commented, “The program was funny, had realistic situations and reactions. I liked understanding why I was supposed to respond a certain way.” The student using PEERS commented, “I didn’t like how sometimes they would do the same topic, but I liked that what they talked about sometimes happens to me too. Now I see how to respond next time.” The same phenomenon was discovered in the comments from those with lower acceptability and appropriateness ratings. All three of the 120 students who did not find the intervention acceptable (scored lower than 24.5 CIRP) and had lower ratings on intervention appropriateness (scores of 6, 10, and 15 out of 20) were receiving the PEERS intervention. One student stated, “I would not like to do it again because it is too hard and frustrating. And I am sorry to say but it’s kind of boring.” Another stated, “The picture quality on videos is good and the people we watched are real relatable people, but imitating what they did correctly

after didn't help me understand why I am supposed to do that." A third participant added, "Acting was okay minus screaming one, some good examples, but did they really have to do it again and again, we get it already. It was like class most of the time boring. We talk, share about our day, film each other doing the right action to one of our problems and watch the one who gets it right over. But what right looks like to her is sus."

After asking a follow-up question on one comment, it was discovered "sus" refers to suspect interpretations of something, and the participant felt that sometimes the correct action in the eyes of a teacher is not the correct action to maintain friends for a student. The comments suggest there may be benefits in examining, in future research, the aspects within interventions (e.g., repetition, response options, relevancy of scene) separately, to examine what causes one intervention to be more acceptable than another.

Limitations

The purpose of this study is to determine middle school students' perceived acceptability, feasibility, and appropriateness of a VR-delivered SEBI (VOISS) versus a SEBI delivered through video modeling (PEERS). The interventions within this study present the EBPs of social narratives and video modeling. However, this study does not consider whether social narratives within the VR intervention (VOISS) are consistent with the indicators of EBPs. This study does not seek to determine if all VR SEBI increase student SEB competencies, as this would require a more considerable number of VR programs created for this purpose. Choosing to focus on middle school students does not provide enough information to determine the implications of this research for those younger than ten and older than fifteen.

Although we recruited at national and regional conferences, this study's participants were limited to four states, making it difficult to generalize the diverse population of the United States. The primary method of determining social validity were participant rating surveys. Creswell (2002) states the major disadvantage of surveys are that they report what people think not what they do, may have low response rates, and do not provide participants flexibility in question responding. These disadvantages do not apply to this research because students' beliefs, not their actions, were being analyzed. Also, the selected surveys were all chosen based on their validity and reliability data and current use within education. There was a comment area added to the surveys within the CIRP adaptation for students to provide any additional thoughts.

A final limitation to this study was the way students interacted with the individual group technology. There was more time spent outside the device with teacher interaction for PEERS. VOISS utilizes built-in questions and answers. Prompts from outside the technology by educators were often related to remaining engaged with the device, whereas PEERS requires interaction with a trained educator. However, the results allow effective comparisons of VOISS's ability to intervene as a stand-alone SEBI.

Implications for Research and Practice

This study supports data on the effectiveness of non-immersive VR interventions (Carreon et al., 2023b; Howard & Gutworth, 2020; Mosher & Carreon, 2021; Mosher et al., 2022) by demonstrating that a non-immersive VR intervention presented through a classroom's current technology was a highly acceptable intervention. Participant CIRP responses suggest the intervention within the technology may be as, if not more, important than the technology delivering the intervention. Comments by students on the CIRP, as well as the significantly lower acceptability and appropriateness ratings of PEERS compared to VOISS, suggest having knowledge of why a skill should be performed in a certain manner and in a specific place may be just as important as providing examples of what the skill looks like and a practice environment. Social communication skill application is often performed in combination with multiple other social skills and is contextually dependent (Ke, et al., 2018). The complexity of this dynamic task increases the challenge of understanding why and when to translate social skill knowledge to performance, particularly for students with pragmatic delays. Future research of interventions that contain the "why" and "where" behind pragmatic skills may assist educators in choosing interventions that have higher levels of efficacy and social validity.

Another important discovery in this study that warrants investigation by future researchers is the measurement tools used to determine whether classroom interventions for adolescents should be adopted. Often, the primary student measurement tool to determine if an intervention should be adopted is a measure of intervention acceptability (Common et al., 2018). It is less common that educators add additional ratings of feasibility and appropriateness. However, this study found that, though feasibility and acceptability were adequate for the PEERS intervention, appropriateness was not. Overall, students did not term acceptability to hold the same meaning as appropriateness, as shown by the differing scores in these two areas by the same rater on the same day about the same intervention in this study. Although appropriateness may entail aspects of acceptability, as shown in the two questions within the IAM appropriateness scale that are like those found in the CIRP scale (i.e., "the program seems suitable" and "the program seems applicable"), appropriateness measures a crucial understanding of whether adolescents feel the intervention best fits their needs. PEERS had lower ratings of "completely agree" on the intervention's perceived "fit" and "match." Many adolescents report SEB programs to be "unmotivating," "irrelevant," and "out of date" (Heckman & Kautz, 2012; Yeager, 2017). Meta-analyses reveal varying degrees of effectiveness of social skill programs for adolescents, which may be due to a feeling of "mismatch" by students (Corcoran et al., 2018; Gates et al., 2017; Wolstencroft et al., 2018). These terms of "irrelevant" or "not fitting" are terms more often associated with an intervention's appropriateness rather than acceptability (Weiner et al., 2017). It would be helpful in future studies to look at interventions being seen as "acceptable," which contain research-based methods but are not making significant growth to determine if these interventions are rated appropriate by their user.

Finally, this study points to the need for future researchers to determine the

cultural fit of an intervention prior to the intervention's implementation. The expected norms and behaviors of cultures are embedded within social skill acquisition. However, the educators' expected norms may not be an appropriate fit to the student's cultural norms. For example, evidence shows positive outcomes when using social narratives and video modeling to develop social communication (Smith et al., 2020; Wong et al., 2015). However, identifying appropriate responses can be subjective and thus challenging, especially when educators create these without guides and examples that fit the student's needs. Both the PEERS and VOISS interventions provide these examples for the student, so all that is required is that the teacher follow the implementation guide rather than create content. Having this validated content within an intervention improves consistency, regardless of the educator implementing the program.

CIRP comments from a study participant revealed the created content of his educator, as well as videos within PEERS, contained "suspect" content that was not a correct fit for him to keep and maintain relationships. Inappropriate instruction in skills was also discovered in the PEERS curriculum during the matching process. These skills were not used in this study and were not found within VOISS. For example, PEERS has listed good and bad eye contact as a curriculum skill and have video models to teach students to maintain eye contact with a speaker. For those within the Navajo tribe, this would be offensive instruction, as making eye contact is seen as disrespectful and impolite (National Park Service, 2018). There are additional populations (i.e., adolescents and adults with a diagnosis of ASD), where instruction in making and receiving direct eye contact is extremely uncomfortable and anxiety producing (Trevisan et al., 2017). Assessments such as the CELF also continue to use eye contact as a measured skill for effective social communication. It is imperative that interventions and assessments ensure intercultural sensitivity so that inappropriate skills are not inadvertently taught and reinforced and students are not incorrectly identified with deficits due to cultural differences. Sharing this knowledge with educators who are creating their own classroom content and assessing students is essential to confirming the content does not provide unintended harm to students and students are not mis-identified as in need of remediation who are displaying appropriate skills.

References

- Alghazo, A., & Al-Otaibi, B. (2016). Using technology to promote academic success for students with learning disabilities. *Journal of Studies in Education*, 6(3), 62.
- Armstrong, K. J., Ehrhardt, K. E., Cool, R. T., & Poling, A. (1997). Social validity and treatment integrity data: Reporting in articles published in the Journal of Developmental and Physical Disabilities, 1991-1995. *Journal of Developmental and Physical Disabilities*, 9, 359-367. <https://doi.org/10.1023/A:1024982112859>
- Baer, D. M., Wolf, M. M., & Risley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, 20(4), 313-327. <https://10.1901/jaba.1987.20-313>
- Berg, J., Osher, D., Same, M. R., Nolan, E., Benson, D., & Jacobs, N. (2017). Identifying, defining, and measuring social and emotional competencies. *Washington, DC: American Institutes for Research*.
- Blakemore, S.-J., & Mills, K. L. (2014). Is adolescence a sensitive period for sociocultural processing? *Annual Review of Psychology*, 65, 187-207. <https://doi.org/10.1146/annurev-psych-010213-115202>
- Briesch, A. M., Chafouleas, S. M., Neugebauer, S. R., & Riley-Tillman, T. C. (2013). Assessing influences on intervention implementation: Revision of the Usage Rating Profile-Intervention. *Journal of School Psychology*, 51(1), 81-96. <https://doi.org/10.1016/j.jsp.2012.08.006>
- Browne, M. N., & Keeley, S. M. (2007). *Asking the right questions: A guide to critical thinking*. Pearson Education. <https://doi.org/10.1016/j.jsp.2012.08.006>
- Callahan, K., Hughes, H. L., Mehta, S., Toussaint, K. A., Nichols, S. M., Ma, P. S., Kutlu, M., & Wang, H.-T. (2017). Social validity of evidence-based practices and emerging interventions in autism. *Focus on Autism and Other Developmental Disabilities*, 32(3), 188-197. <https://doi.org/10.1177/1088357616632446>
- Carreon, A., Criss, C., & Mosher M. (2023a). Classroom virtual reality: A preliminary guide to available virtual content. *Journal of Special Education Technology*. 39(1), 143-150. <https://doi.org/10.1177/01626434231170593>
- Carreon, A., Smith, S., Frey, B., Rowland, A., & Mosher M. (2023b). Comparing immersive VR and non-immersive VR on social skill acquisition for students in middle school with ASD, *Journal of Research on Technology in Education*. 56(5), 530-543. <https://doi.org/10.1080/15391523.2023.2182851>
- Carreon, A., Smith, S., Mosher, M., & Rowland, A. (2022). A review of virtual reality intervention research for students with disabilities in K-12 settings. *Journal of Special Education Technology*. 37(1), 82-99. <https://doi.org/10.1177/0162643420962011>

- Carter, S. L., & Wheeler, J. J. (2019). *The social validity manual: Subjective evaluation of interventions*. Academic Press. <https://doi.org/10.1016/B978-0-12-816004-6.00008-4>
- Charlop-Christy, M. H., & Daneshvar, S. (2003). Using video modeling to teach perspective taking to children with autism. *Journal of Positive Behavior Interventions*, 5(1), 12-21. <https://doi.org/10.1177/10983007030050010101>
- Children, C. f. (2019). *Social-Emotional Learning and Youth Suicide Prevention*. <https://www.cfchildren.org/wp-content/uploads/policy-advocacy/sel-youth-suicide-prevention.pdf>
- Creswell, J. W. (2002). *Educational research: Planning, conducting, and evaluating quantitative* (Vol. 7). Prentice Hall Upper Saddle River, NJ.
- Crone, E. A., & Dahl, R. E. (2012). Understanding adolescence as a period of social–affective engagement and goal flexibility. *Nature Reviews Neuroscience*, 13(9), 636-650. <https://doi.org/10.1038/nrn3313>
- Felver, J. C., Felver, S. L., Margolis, K. L., Kathryn Ravitch, N., Romer, N., & Horner, R. H. (2017). Effectiveness and social validity of the soles of the feet mindfulness-based intervention with special education students. *Contemporary School Psychology*, 21, 358-368. <https://doi.org/10.1007/s40688-017-0133-2>
- Finney, J. W. (1991). On further development of the concept of social validity. *Journal of Applied Behavior Analysis*, 24(2), 245. <https://doi.org/10.1901/jaba.1991.24-245>
- Fowler Jr, F. J., & Fowler, F. J. (1995). *Improving survey questions: Design and evaluation*. Sage.
- Fox, J. J., & McEvoy, M. A. (1993). Assessing and enhancing generalization and social validity of social-skills interventions with children and adolescents. *Behavior Modification*, 17(3), 339-366. <https://doi.org/10.1177/01454455930173006>
- Germer, K. A., Kaplan, L. M., Giroux, L. N., Markham, E. H., Ferris, G. J., Oakes, W. P., & Lane, K. L. (2011). A function-based intervention to increase a second-grade student's on-task behavior in a general education classroom. *Beyond Behavior*, 20(3), 19-31.
- Gilmore, R., Ziviani, J., McIntyre, S., Goodman, S., Tyack, Z., & Sakzewski, L. (2023). Exploring caregiver and participant experiences of the Program for the Education and Enrichment of Relational Skills (PEERS®) for youth with acquired brain injury and cerebral palsy. *Disability and Rehabilitation*, 1-9.

- Glanz, K., Rizzo, A. S., & Graap, K. (2003). Virtual reality for psychotherapy: Current reality and future possibilities. *Psychotherapy: Theory, Research, Practice, Training*, 40(1-2), 55. <https://doi.org/10.1037/0033-3204.40.1-2.55>
- Gray, C. D., & Kinnear, P. R. (2012). *IBM SPSS statistics 19 made simple*. Psychology Press. <https://doi.org/10.4324/9780203723524>
- Halabi, O., El-Seoud, S. A., Alja'am, J. M., Alpona, H., Al-Hemadi, M., & Al-Hassan, D. (2017). Design of Immersive Virtual Reality System to Improve Communication Skills in Individuals with Autism. *International Journal of Emerging Technologies in Learning*, 12(5), 50-64. <https://doi.org/10.3991/ijet.v12i05.6766>
- Hansen, D. J., St. Lawrence, J. S., & Christoff, K. A. (1989). Group conversational-skills training with inpatient children and adolescents: Social validation, generalization, and maintenance. *Behavior Modification*, 13(1), 4-31. <https://doi.org/0.1177/01454455890131001>
- Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. *Labour Economics*, 19(4), 451-464. <https://doi.org/10.1016/j.labeco.2012.05.014>
- Hew, K. F., & Brush, T. (2007). Integrating technology into K-12 teaching and learning: Current knowledge gaps and recommendations for future research. *Educational Technology Research and Development*, 55, 223-252. <https://doi.org/10.1007/s11423-006-9022-5>
- Hew, K. F., & Cheung, W. S. (2010). Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology*, 41(1), 33-55. <https://doi.org/10.1111/j.1467-8535.2008.00900.x>
- Howard, M. C., & Gutworth, M. B. (2020). A meta-analysis of virtual reality training programs for social skill development. *Computers & Education*, 144, 103707. <https://doi.org/10.1016/j.compedu.2019.103707>
- Jennings, P. A., & Greenberg, M. T. (2009). The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes. *Review of Educational Research*, 79(1), 491-525. <https://doi.org/10.3102/0034654308325693>
- Jones, S. M., Barnes, S. P., Bailey, R., & Doolittle, E. J. (2017). Promoting social and emotional competencies in elementary school. *The Future of Children*, 27(1), 49-72. <https://doi.org/10.1353/foc.2017.0003>
- Kazdin, A. E. (1977). Assessing the clinical or applied importance of behavior change through social validation. *Behavior Modification*, 1(4), 427-452. <https://doi.org/10.1177/014544557714001>

- Kazdin, A. E. (1980). Acceptability of alternative treatments for deviant child behavior. *Journal of Applied Behavior Analysis*, 13(2), 259-273.
<https://doi.org/10.1901/jaba.1980.13-259>
- Kern, L., & Manz, P. (2004). A look at current validity issues of school-wide behavior support. *Behavioral Disorders*, 30(1), 47-59.
<https://doi.org/10.1177/019874290403000102>
- King, B., Radley, K. C., Jenson, W. R., Clark, E., & O'Neill, R. E. (2014). Utilization of video modeling combined with self-monitoring to increase rates of on-task behavior. *Behavioral Interventions*, 29(2), 125-144. <https://doi.org/10.1002/bin.1379>
- Kuznekoff, J. H., & Titsworth, S. (2013). The impact of mobile phone usage on student learning. *Communication Education*, 62(3), 233-252.
<https://doi.org/10.1080/03634523.2013.767917>
- Lane, K. L., Royer, D. J., Messenger, M. L., Common, E. A., Ennis, R. P., & Swogger, E. D. (2015). Empowering teachers with low-intensity strategies to support academic engagement: Implementation and effects of instructional choice for elementary students in inclusive settings. *Education and Treatment of Children*, 38(4), 473-504.
<https://doi.org/10.1353/etc.2015.0013>
- Lorenzo, G., Lledó, A., Pomares, J., & Roig, R. (2016). Design and application of an immersive virtual reality system to enhance emotional skills for children with autism spectrum disorders. *Computers & Education*, 98, 192-205.
<https://doi.org/10.1016/j.compedu.2016.03.018>
- Lozano-Álvarez, M., Rodríguez-Cano, S., Delgado-Benito, V., & Mercado-Val, E. (2023). A Systematic Review of Literature on Emerging Technologies and Specific Learning Difficulties. *Education Sciences*, 13(3), 298.
<https://doi.org/10.3390/educsci13030298>
- Martens, B. K., Witt, J. C., Elliott, S. N., & Darveaux, D. X. (1985). Teacher judgments concerning the acceptability of school-based interventions. *Professional Psychology: Research and Practice*, 16(2), 191. <https://doi.org/10.1037/0735-7028.16.2.191>
- McCoy, A., Holloway, J., Healy, O., Rispoli, M., & Neely, L. (2016). A systematic review and evaluation of video modeling, role-play and computer-based instruction as social skills interventions for children and adolescents with high-functioning autism. *Review Journal of Autism and Developmental Disorders*, 3, 48-67.
<https://doi.org/10.1007/s40489-015-0065-6>
- McCoy, D. C., Peet, E. D., Ezzati, M., Danaei, G., Black, M. M., Sudfeld, C. R., Fawzi, W., & Fink, G. (2016). Early childhood developmental status in low-and middle-income countries: national, regional, and global prevalence estimates using predictive

- modeling. *PLoS medicine*, 13(6), e1002034.
<https://doi.org/10.1371/journal.pmed.1002034>
- McIntyre, L. (1999). *The practical skeptic: core concepts in sociology*. Mountain view, CA: *Mayfield Publishing International*, 311(6997), 109-112.
- Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education*, 56(3), 769-780.
<https://doi.org/10.1016/j.compedu.2010.10.020>
- Mosher, M. (2022). Technology tools available for implementing social skill instruction. *TEACHING Exceptional Children*. 55(1), 60-71.
<https://doi.org/10.1177/00400599211041738>
- Mosher, M. & Carreon, A. (2021). *Teaching social skills to students with autism spectrum disorder through augmented, virtual, and mixed reality*. *Research in Learning Technology*, 29. <https://doi.org/10.25304/rlt.v29.2626>
- Mosher, M. A., Carreon, A. C., Craig, S. L., & Ruhter, L. C. (2022). Immersive technology to teach social skills to students with autism spectrum disorder: a literature review. *Review Journal of Autism and Developmental Disorders*, 9, 1-17. <https://doi.org/10.1007/s40489-021-00259-6>
- Mosher, M. A., Carreon, A. C., & Sullivan, B. J. (2020). A step-by-step process for selecting technology tools for students with ADHD. *Journal of Special Education Technology*, 37(2), 310-317. <https://doi.org/10.1177/0162643420978570>
- Mosher, M., Frey, B., Carreon, A., Smith, S., Rowland, A., & Lowrey, A. (2024). The Technology Immersive Presence Scale (TIPS) Proof of Concept: Creating a Student Accessible Measure of Presence in Virtual Environments. *Journal of Interactive Learning Research*, 35(1), 101-133.
- National Park Service (2018) *Hubbel Trading Post Traveling Among the Navajo*.
<https://www.nps.gov/hutr/planyourvisit/upload/Traveling-Among-the-Navajo.pdf>
- Posamentier, J., Seibel, K., & DyTang, N. (2023). Preventing youth suicide: A review of school-based practices and how social–emotional learning fits into comprehensive efforts. *Trauma, Violence, & Abuse*, 24(2), 746-759.
<https://doi.org/10.1177/15248380211039475>
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: conceptual distinctions, measurement challenges, and research agenda. *Administration and policy in mental health and mental health services research*, 38, 65-76.
<https://doi.org/10.1007/s10488-010-0319-7>

- Rajendran, G. (2013). Virtual environments and autism: a developmental psychopathological approach. *Journal of Computer Assisted Learning*, 29(4), 334-347. <https://doi.org/10.1111/jcal.12006>
- Reimers, T. M., Wacker, D. P., Cooper, L. J., & De Raad, A. (1992). Acceptability of behavioral treatments for children: Analog and naturalistic evaluations by parents. *School Psychology Review*, 21(4), 628-643. <https://doi.org/10.1016/j.invent.2016.12.001>
- Richtel, M. (2021). Children's screen time has soared in the pandemic, alarming parents and researchers. *The New York Times*, 17. <https://static1.squarespace.com/static/59dba6d3e3df281768a63220/t/606f1cb849a3137b7548dcbd/1617894584991/Childrens+Screen+Time+Has+Soared+in+the+Pandemic.pdf>
- Salant, P., Dillman, I., & Don, A. (1994). *How to conduct your own survey*.
- Sawchuk, S. (2021, October 12). Why high school SEL programs feel 'lame'—and how to fix them. Education Week. <https://www.edweek.org/leadership/why-high-school-sel-programs-feel-lame-and-how-to-fix-them/2021/10>
- Smith, S., Cheatham, G., & Mosher, M. (2020). Evidence-based Practices to Promote Inclusion in Today's Catholic School. *Journal of Catholic Education*, 23(2), 111-134. <http://dx.doi.org/10.15365/joce.2302102020>
- Smith, S. J., Mosher, M. A., Lowrey, K. A. (2022). Advances in the use of technology and online learning to improve outcomes for students with disabilities. In Lemons, C. J., Powell, S. R., Lane, K. L., & Aceves, T. C. (Eds.). *Handbook of Special Education Research, Volume II: Research-Based Practices and Intervention Innovations Volume II* (pp. 178-189). Routledge. <https://doi.org/10.4324/9781003156888>
- Sweeney, G. M., Donovan, C. L., March, S., & Forbes, Y. (2019). Logging into therapy: Adolescent perceptions of online therapies for mental health problems. *Internet Interventions*, 15, 93–99. <https://doi.org/10.1016/j.invent.2016.12.001>
- Tarnowski, K. J., & Simonian, S. J. (1992). Assessing treatment acceptance: The abbreviated acceptability rating profile. *Journal of Behavior Therapy and Experimental Psychiatry*, 23(2), 101-106. [https://doi.org/10.1016/0005-7916\(92\)90007-6](https://doi.org/10.1016/0005-7916(92)90007-6)
- Taylor, R. D., Oberle, E., Durlak, J. A., & Weissberg, R. P. (2017). Promoting positive youth development through school-based social and emotional learning interventions: A meta-analysis of follow-up effects. *Child Development*, 88(4), 1156-1171. <https://doi.org/10.1111/cdev.12864>
- Trevisan, D. A., Roberts, N., Lin, C., & Birmingham, E. (2017). How do adults and teens with self-declared Autism Spectrum Disorder experience eye contact? A qualitative analysis of first-hand accounts. *PLoS One*, 12(11). <https://doi.org/10.1371/journal.pone.0188446>

- von Brock, M. B., & Elliott, S. N. (1987). Influence of treatment effectiveness information on the acceptability of classroom interventions. *Journal of School Psychology, 25*(2), 131-144. [https://doi.org/10.1016/0022-4405\(87\)90022-7](https://doi.org/10.1016/0022-4405(87)90022-7)
- Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., Boynton, M. H., & Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science, 12*, 1-12. <https://doi.org/10.1186/s13012-017-0635-3>
- Witt, J. C., & Elliot, S. N. (1985). Acceptability of classroom intervention strategies. In T. R. Kratochwill (Ed.), *Advances in School Psychology* (Vol. 4, pp. 251-288). Mahwah, NJ: Erlbaum.
- Wolf, M. M. (1978). Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*(2), 203-214. <https://doi.org/10.1901/jaba.1978.11-203>
- Wong, C. A., Madanay, F., Ozer, E. M., Harris, S. K., Moore, M., Master, S. O., ... & Weitzman, E. R. (2020). Digital health technology to enhance adolescent and young adult clinical preventive services: affordances and challenges. *Journal of Adolescent Health, 67*(2), S24-S33. <https://doi.org/10.1016/j.jadohealth.2019.10.018>
- Yeager, D. S. (2017). Social and emotional learning programs for adolescents. *The Future of Children, 73-94*. <https://doi.org/10.1353/foc.2017.0004>