

## Engagement with Electronic Screen Media Among Students with Autism Spectrum Disorders

Beth A. Mineo · William Ziegler · Susan Gill ·  
Donna Salkin

Published online: 15 July 2008  
© Springer Science+Business Media, LLC 2008

**Abstract** This study investigated the relative engagement potential of four types of electronic screen media (ESM): animated video, video of self, video of a familiar person engaged with an immersive virtual reality (VR) game, and immersion of self in the VR game. Forty-two students with autism, varying in age and expressive communication ability, were randomly assigned to the experimental conditions. Gaze duration and vocalization served as dependent measures of engagement. The results reveal differential responding across ESM, with some variation related to the engagement metric employed. Preferences for seeing themselves on the screen, as well as for viewing the VR scenarios, emerged from the data. While the study did not yield definitive data about the relative engagement potential of ESM alternatives, it does provide a foundation for future research, including guidance related to participant profiles, stimulus characteristics, and data coding challenges.

**Keywords** Autism spectrum disorders · Video · Visual media · Engagement · Virtual reality

The recent proliferation of technology applications designed to promote learning among individuals with autism spectrum disorders (ASD) has been met with both optimism and concern. Although well-designed studies are revealing technology's promise in facilitating the development of a variety of skills, there is a tendency in this field to embrace new treatment approaches with an enthusiasm disproportionate to the evidence for their applicability across the population (Jacobson et al. 2005; National Research Council 2001). Studies comparing the relative appeal and impact of technology-enabled interventions are few.

There is abundant anecdotal and research-based evidence for the strength of the visual modality among many individuals with ASD (Grandin 1996; Joseph et al. 2002; Pierce and Schreibman 1994; Quill 1997). Many parents report their children's fascination with and propensity for learning from videos and computer games (Heiman et al. 1995; Nally et al. 2000). In addition, a preference for visual stimuli, particularly those delivered via electronic screen media (ESM),<sup>1</sup> has been noted by numerous researchers (Bernard-Opitz et al. 2001; Buggey 2005; Charlop-Christy and Daneshvar 2003; Schreibman et al. 2000; Shane and Albert 2008; Tissot and Evans 2003). This has led to speculation that such preferences might be exploited to further the development of a variety of skills. Indeed, it has been suggested that the very nature of electronic screen media renders them ideal for the delivery of information to this population. The relatively constrained viewing area limits the attentional frame, helping those with ASD to focus their attention on relevant stimuli and ignore

---

B. A. Mineo (✉)  
University of Delaware, 42 East Delaware Avenue, Newark, DE  
19716, USA  
e-mail: mineo@asel.udel.edu

W. Ziegler  
Bucks County Intermediate Unit, Doylestown, PA, USA

S. Gill · D. Salkin  
Pennsylvania Training and Technical Assistance Network, King  
of Prussia, PA, USA

---

<sup>1</sup> Electronic screen media (ESM) include media for the television screen, as used for television and movie viewing, and the computer monitor, as used for computer applications (Shane and Albert 2008).

irrelevant ones (Charlop-Christy and Daneshvar 2003; Sherer et al. 2001; Shipley-Benamou et al. 2002). The predictability across repeated viewings enables the viewer to anticipate upcoming scenes (Shane and Albert 2008). The auditory “soundtrack” tends to be closely synchronized with the visual stimulation, and this multimodal information flow may assist the viewer in the coordinated processing of information (Quill 1997). Viewing of ESM can be accomplished without interference from another person, and ESM options typically do not make “demands” on an individual with ASD (Bernard-Opitz et al. 2001; Shipley-Benamou et al. 2002).

Many have exploited this preference for ESM for instructional purposes. Among the most fruitful applications is the use of video modeling, which involves the presentation of videotaped samples of models engaged in behaviors targeted for change. Video modeling has been used with individuals with ASD to teach a wide variety of skills. In some instances, adults are portrayed engaging in the target behaviors (MacDonald et al. 2005), while in other cases peers are recruited as models (Sherer et al. 2001; Simpson et al. 2004). Ogletree and colleagues (1995) even used illustrative segments of animated Disney videos as models in an intervention targeting semantic/pragmatic language deficits. In another approach, “self” models, generated through creative response elicitation and video editing, yield an image of the learner engaged in the behavior targeted for change (Buggey 2005; Coyle and Cole 2004; Dowrick 1999; Wert and Neisworth 2003).

The use of video modeling with those having ASD is appealing for a number of reasons. First, many individuals with ASD routinely and spontaneously imitate the actions they see depicted on television or in videos (Charlop-Christy and Daneshvar 2003; Shane and Albert 2008), and video modeling interventions exploit these natural proclivities. Second, model footage may be obtained relatively easily; Alberto and colleagues (2005) found development of video materials to be less time-consuming than the development of static pictorial materials. Unobtrusive video cameras are now ubiquitous, and tools such as Apple’s iMovie® make video editing relatively easy and affordable. Capturing footage of models engaged in target behavior is straightforward, assuming the cooperation of the model. A bit trickier may be capturing the individual him/herself engaged in instances of the target behavior. If these video segments are elicited using prompts and models, the elicitation techniques must be edited out and the desired footage spliced together to depict a seamless stream of behaviors. This can be more challenging and time-consuming than the simple edits required to obtain models of “others” engaged in target behaviors, yet the benefits may outweigh the costs if the “self” models are more effective in eliciting the desired responses. Third,

implementation of video modeling interventions has been found to be more economical in terms of staff resources than are interventions reliant on in vivo instruction (Gena et al. 2005; Shipley-Benamou et al. 2002). Finally, a recent meta-analysis (Bellini and Akullian 2007) demonstrates that video modeling and video self modeling are effective in establishing new behaviors and facilitating generalization of those behaviors beyond the training context.

Another genre of video applications recently harnessed for the benefit of individuals with ASD is virtual reality (VR), in which the individual becomes a participant in the action depicted on the screen. Virtual environments are enabled by three different constellations of technologies. In the first, desktop systems employ a computer monitor displaying a graphical environment. The interactant typically controls an “avatar,” an actor within this environment, using standard input devices such as keyboards, mice, and joysticks. Parsons and colleagues used this type of technology to assess adherence to social conventions within a “virtual café” (Parsons et al. 2005). Of course, this type of application can only be considered a legitimate reflection or extension of the individual’s skills if the individual identifies with the virtual actor and expresses his intentions through the avatar. In this regard, the tendency of people with ASD to interpret situations literally, and to have difficulty with their sense of reality in relation to the external environment, may impede VR’s potential as an instructional tool (Beardon et al. 2001; Parsons et al. 2005; Sherer et al. 2001).

Another variation on virtual reality is the immersive type. Some total immersion systems use head-mounted stereoscopic displays, sound systems, and various devices that both transmit and receive data (such as specially designed gloves with “force feedback” that generate the sensation of encountering surfaces within the virtual world) (Polys et al. 2008). These systems give the user the sense of being completely surrounded by another “reality.” Predictably, many individuals with ASD are not amenable to being outfitted with contraptions that occlude their vision or provide additional weight or aversive sensory stimulation. Other total immersion systems project the virtual world onto the walls of a room, giving the sensation of being surrounded by the virtual environment without the need to don extensive head gear (DeFanti et al. 1993).

A third type of system—a hybrid of sorts—achieves the sense of immersion by portraying the interactant him/herself within the display on a typical monitor. Known as immersive video, the individual’s image is captured with a camera and superimposed on a scene in which the individual’s movements impact the action on the screen. This type of technology has intrigued visitors to museums for many years, and has recently impacted the consumer electronics market with such products as the PlayStation

EyeToy<sup>®</sup>. The use of immersive video has been explored as a means of motivating individuals with motoric limitations to engage in physical activity and extend their range of motion (Harris and Reid 2005; Weiss et al. 2003). Witnesses to these types of applications typically report high levels of interest and engagement among those interacting with the systems, although the participants in the Harris and Reid (2005) study did not find all VR applications equally motivating. Most current applications of this type are designed for entertainment purposes, although their potential as vehicles for instruction is extremely high, given their reputed power to attract and maintain the attention of the interactants (Holden 2005; Parsons et al. 2005; Weiss et al. 2003).

This engagement potential, combined with the preference of many people with ASD for electronic screen media, suggests that immersive VR holds promise as a vehicle for delivery of instruction for those on the autism spectrum (Coyle and Cole 2004; Harris and Reid 2005; Kinney et al. 2003). In addition, these technologies carry a certain “cool” factor, and represent non-stigmatizing intervention approaches because they have been embraced by society at large (Ward 2005). We have no empirical evidence, however, that immersive VR is a more attractive and compelling electronic screen medium than other types of instructional delivery mechanisms in more abundant supply in our schools.

Video-based interventions of all types offer some distinct advantages over in vivo training. They enable instruction to be delivered with absolute consistency, and they can be repeated an infinite number of times without degradation of fidelity. Rules and concepts can be conveyed through action, eliminating the need for language or other symbol systems as the vehicle for information exchange (Gena et al. 2005; Shane et al. 2005). VR offers these, and additional, advantages as it is capable of depicting a wide range of environments, many of which may not be practical to capture on video. Within a VR application, the levels of task presentation, response requirements, and nature and pattern of feedback can be systematically modified to suit the interactant’s preferences and level of functioning. Further, VR applications can simulate scenarios that are too dangerous or impractical to experience in real life, affording numerous opportunities for practice without risk of real world consequences for mistakes (Parsons et al. 2005; Rose et al. 2005; Trepagnier 1999).

Despite the promise of instructional delivery via ESM, there have been no published studies examining the *relative* appeal of various types of stimuli delivered using these technologies (Nikopoulous and Keenan 2004; Shipley-Benamou et al. 2002). Even the studies pertaining to more traditional types of video applications—such as those

employing video modeling—are of limited use in addressing this question because of their small numbers of participants, their tendency to involve those with relatively high levels of functioning, and their restriction of instructional stimuli to a single type.

For screen-based technology to have instructional horsepower, it must engage the student and maintain the student’s attention (Woltersdorf 1992). As a foundation for future intervention research, this study was constructed to enable examination of the engagement potential of four different types of screen-based information presentation techniques across a relatively heterogeneous population of individuals with ASD. Specifically, engagement was measured as the students were presented with animated video footage, video footage of themselves engaged in a familiar routine, video footage of a familiar person engaged with an immersive VR game, and an opportunity to view themselves as they interact with the same VR game in real time. We sought to document student preferences relative to the nature of the screen-based media (video or VR) and the “actors” depicted on the screen (self or other).

## Methods

### Research Design

Participants ( $n = 42$ ) were assigned to one of three groups, defined by the visual presentation format to which they were exposed. All participants first were exposed to the same baseline condition (an animated video). Subsequently, they were exposed to one additional screen-based segment, the nature of which depended on the group to which they were assigned. In the Self Video condition, participants saw a video of themselves engaged in a non-aversive activity. In the Other VR condition, participants saw a video of a person familiar to them engaged with an immersive VR system. In the Self VR condition, participants were exposed to a real-time immersive VR application with which they could interact; it was the same application used to generate the Other VR stimuli. Exposure to the second condition occurred within 2–10 days of exposure to the baseline condition. Participants viewed all conditions on a 20-inch screen.

Participants were randomly assigned to groups to reduce the likelihood that confounding variables of age and language ability would account for differences in the outcomes. The study controlled for any such difference that might have existed between the groups by covarying both age and language level as well as baseline performance. An a priori power analysis was conducted to determine the minimum number of participants that would need to be enrolled. The criterion for significance was set at  $p = .05$ .

It was assumed that the analysis of covariance would be non-directional (i.e., two-tailed), thus an effect in either direction would be interpreted. A medium effect size was specified (i.e.,  $f = .25$ , as per Cohen 1988). The study included two covariates (age and an expressive communication indicator) that accounted for 80% of the variance in the dependent variable (i.e.,  $r = .90$  with the dependent variables). Lastly, power was set to  $.80$ , meaning that the study would have an 80% probability of finding differences among the three groups if such differences exist in the population. The power analysis indicated that the study required a minimum of 13 cases per group for a total of 39 cases. Actual enrollment was 14 cases per group, for a total of 42 cases.

### Participants

All participants attended school programs for which a formal diagnosis of autism was the criterion for entry. Youngsters were recruited from eight schools in two intermediate units (regional education service agencies) in Southeastern Pennsylvania. Letters were sent home to parents explaining the study and seeking permission for the students to participate and to be videotaped. Those students returning permission forms marked in the affirmative were enrolled in the study, and a survey seeking more detailed information about the students' communication behaviors, behavioral patterns, and experiences and preferences relative to visual stimuli was distributed to the parent(s) and the students' primary teachers.

Participants were assigned to one of three experimental conditions when their completed permission forms were returned. Assignment occurred as forms were returned, such that the first student with a completed form was assigned to the Self Video condition, the next student with a completed form was assigned to the Other VR condition, and so on. The only exception to this random assignment was in the case of female participants. As the number of female students in the targeted programs was significantly fewer than the number of males, a separate assignment to conditions was reserved for females in an attempt to achieve approximately equivalent distribution of the females across conditions.

Baseline data were secured for a total of 62 students. Several participants were eliminated prior to collection of data in the second condition. Additional participants were eliminated because of problems revealed during data coding. Ultimately, the data from 42 students were included in the data analysis. Student attrition was attributable to: (a) the lack of completed survey information from parents and/or teachers ( $n = 3$ ); (b) the presence of significant discrepancies between parent and teacher report that could not be reconciled with the use of additional data ( $n = 6$ );

(c) student request to withdraw from the study ( $n = 1$ ); (d) identification of a student characteristic that had the potential to confound the results (student had a cochlear implant and his level of auditory acuity was not verified) ( $n = 1$ ); (e) the parent/teacher survey reported that the student had never sustained attention to a video for 5 min or more ( $n = 1$ ); (f) student absence ( $n = 1$ ); (g) student responses during the study were contaminated by procedural violations or distractions in the room during data collection ( $n = 5$ ); and (h) digital video data files were compromised ( $n = 2$ ).

A summary of participant characteristics can be found in Table 1. In addition to chronological age (years; months) at the time of baseline data collection, the table also includes participant status relative to five defining communication characteristics. None of the participants completing the research protocol demonstrated other physical or sensory disabilities that had the potential to interfere in their performance on the experimental tasks.

### Communication Characteristics

Heterogeneity is a hallmark of the population of persons with ASD (National Research Council 2001), and individuals often demonstrate “splintered” patterns of skills within a single domain, such as communication (Tager-Flusberg 2004; Wilkinson 1998). As several studies of visual behavior and video-based interventions found that expressive language was a major factor differentiating between high-performing and low-performing groups (Charlop-Christy and Daneshvar 2003; Parsons et al. 2005; Spiker and Ricks 1984), we sought to define a communication-related characteristic that could function as a covariable in the data analysis. Several researchers consider expressive language level to be a key predictor of outcome in the ASD population (Joseph et al. 2002; Kobayashi et al. 1992; Lord et al. 2000), so an item was included on the parent and teacher surveys asking the informants to estimate the number of meaningful words that students produced. The survey clarified that “meaningful words” are ones that the student uses spontaneously, not just in imitation or as a part of a repetitive phrase. Informants could select from the following ranges: 0–9, 10–25, 26–50, 51–75, 76–100, and 100+. For descriptive purposes, this information was translated into a binary response, with a plus (+) designation indicating that the participant's expressive lexicon was more than 50 words, and a minus (–) indicating that the expressive lexicon was 50 or fewer words. This cut point was not arbitrary; rather, it reflects the point at which typically-developing children begin to combine words, or what Lord and colleagues (2000) refer to as “phrase speech.” The presence of phrase speech in a youngster's communicative repertoire is

**Table 1** Participant characteristics

Condition/participant	CA	Communication characteristics				
		Receptive lexicon > 50	Expressive lexicon > 50	Yes/No gestures	Echolalia	Conversation
<i>Self video</i>						
1	8;4	+	+	+	-	+
2	6;1	+	+	+	-	+
3	6;2	+	-	+	+	-
4	6;2	+	+	+	+	+
5	14;4	+	+	+	+	+
6	13;6	+	+	+	+	+
7	14;2	+	+	+	-	+
8	13;4	+	+	-	+	+
9	12;6	+	-	-	+	-
10	7;11	+	+	+	+	+
11	8;8	+	+	-	-	+
12	13;3	+	-	-	+	+
13	10;5	+	-	-	+	-
14	6;9	+	+	+	-	+
<i>Other VR</i>						
15	8;9	+	-	+	+	-
16	12;9	+	-	+	+	+
17	13;7	+	-	+	-	-
18	18;8	-	-	-	-	-
19	14;2	+	-	+	-	-
20	6;11	+	-	-	+	-
21	7;5	+	-	-	-	-
22	8;4	+	-	+	-	-
23	12;6	+	+	-	+	-
24	10;4	+	+	-	-	-
25	8;10	+	+	+	+	+
26	6;9	+	+	+	+	-
27	6;5	+	+	+	+	+
28	7;5	+	+	+	+	-
<i>Self VR</i>						
29	6;0	+	-	+	+	+
30	6;10	+	+	+	+	+
31	14;2	+	+	+	-	+
32	8;5	+	+	+	+	+
33	13;9	+	+	+	-	+
34	12;6	+	-	+	+	+
35	12;8	+	+	-	+	-
36	13;8	+	-	+	-	-
37	16;0	-	-	-	-	-
38	6;8	+	-	+	-	+
39	12;1	-	-	-	+	-
40	17;1	+	-	-	+	+
41	10;10	+	+	-	+	-
42	11;8	-	-	+	-	-

considered by Lord and colleagues (2000) to distinguish between children who are “verbal” from those who are “pre-verbal.” In this study, we consider the participants with expressive lexicons of 50 words or less as “low language” rather than pre-verbal, as the vast majority was reported to produce some meaningful words.

It should be noted that school records yielded no uniform language measure that could be used to characterize participants’ communication competence. On the other hand, the parent/teacher surveys offered a wealth of communication-related information. Although the expressive communication measure was selected as a covariable for data analysis, four additional elements are presented in Table 1, as the expressive measure alone is not adequate for characterization of their communication profiles. These items, which were selected with guidance from Seltzer and colleagues (2003), include a measure of receptive lexicon size, as well as indications of the extent to which participants use yes/no gestures, are echolalic, and engage in conversation. Informants were asked to estimate the size of the youngsters’ receptive lexicon using the same approach described relative to the expressive lexicon. With regard to the remainder of the communication elements, informants were asked to characterize the frequency of the specified behavior using a four-point scale that included “always,” “sometimes,” “rarely,” and “never”. In instances in which parent and teacher responses were markedly discrepant, teacher responses were used for the purpose of characterizing student behavior.

All five communication dimensions are represented in a binary manner, such that:

- (a) A plus (+) rating for *Receptive Lexicon > 50* indicates that the student was reported to understand more than 50 spoken words.
- (b) A plus (+) rating for *Expressive Lexicon > 50* indicates that the student was reported to produce more than 50 words using speech, sign, or picture.
- (c) A plus (+) rating for *Yes/No Gestures* indicates that the student was reported to use a conventional gesture (head nod/head shake) to express yes and/or no “always” or “sometimes.”
- (d) A plus (+) rating for *Echolalia* indicates that the student was reported to immediately echo the utterances of others “always” or “sometimes.”
- (e) A plus (+) rating for *Conversation* indicates that the student was reported to engage in conversation with the teacher informant “always” or “sometimes.”

#### *Experiences With/Preferences for ESM*

The survey instrument also solicited data about participants’ experience with and preferences for ESM.

Regarding experience, informants were asked to indicate whether the students had ever: (a) watched a video for more than 5 min; (b) watched a video for more than 30 min; (c) watched a video for more than 1 h; (d) played a video game; (e) played a video game that pictured the student as a part of the action; (f) imitated the movements of a character on TV or in a video; (g) recognized him/herself in pictures; and (h) recognized him/herself in videos.

As attention was a necessary prerequisite for legitimate participation in this study, we imposed the same 5-min inclusion criterion used by Shipley-Benamou and colleagues (2002) in their video modeling study. One participant was eliminated from the study when the informants indicated that he had never watched a video for more than 5 min. A summary of student experiences, again clustered by group assignment, can be found in Table 2. A plus (+) designation signifies that at least one informant indicated that the student had engaged in the specified activity. A minus (–) indicates that the student had not engaged in the activity.

Regarding preferences, informants were asked to rate the students’ typical reactions to a variety of visual media and social and recreational activities. The 5-point rating scale ranged from 5—*dislikes very much* to 1—*likes very much*. A rating of 3 indicated that the youngster neither liked nor disliked the activity. Informants also had options to indicate that the student had never experienced the activity or that they were not sure of the students’ preference. Nineteen items were included in the preference section of the survey, the most relevant for this study being: (a) still pictures of him/herself; (b) videos of him/herself, (c) still pictures of familiar others; (d) videos of familiar others; (e) still pictures of TV/movie characters; (f) videos of TV/movie characters; and (g) playing video games.

Student preferences for these activities are depicted in Table 3, with participants again clustered by group assignment. A plus (+) designation was assigned when both informants indicated that the student either “liked” or “liked very much” the specified item or activity. A minus (–) designation was assigned when both informants indicated that the student either “disliked” or “disliked very much” the specified item or activity. A designation of 0 was assigned when the child was reported to “neither dislike nor like” the item or activity, and a designation of “D” was assigned when there was a discrepancy between parent and teacher report. A question mark (?) indicates that both informants either claimed no knowledge of the student’s preference or that the student had never experienced the activity.

**Table 2** Participant experience with visual media

Condition/ participant	Type of visual media experience						
	Video > 30 min	Video > 60 min	Video game	Immersive video game	Imitate character	Recognize self- pictures	Recognize self- video
<i>Self video</i>							
1	+	+	+	+	+	+	+
2	+	+	+	+	+	+	+
3	+	+	+	-	+	+	+
4	+	+	+	-	+	+	+
5	+	+	+	-	+	+	+
6	+	+	+	+	+	+	+
7	+	+	+	-	+	+	+
8	+	+	+	-	+	+	+
9	+	+	-	-	-	+	+
10	+	-	+	-	+	+	+
11	+	+	+	-	+	+	+
12	+	+	+	-	+	+	+
13	+	+	+	-	-	+	-
14	+	+	+	-	-	+	+
<i>Other VR</i>							
15	-	-	+	+	+	+	+
16	+	+	-	-	+	+	+
17	+	+	+	-	-	+	+
18	+	+	-	-	-	+	-
19	+	+	+	-	+	+	+
20	-	-	+	-	+	+	+
21	+	-	-	-	-	-	+
22	+	+	+	-	+	+	+
23	+	+	-	-	-	+	+
24	+	+	+	-	+	+	+
25	+	+	+	-	-	+	+
26	+	+	+	+	+	+	+
27	+	+	+	+	+	+	+
28	+	+	+	+	+	+	+
<i>Self VR</i>							
29	+	+	+	-	+	+	+
30	+	+	+	+	+	+	+
31	+	+	-	-	-	+	+
32	+	+	+	-	+	+	+
33	+	+	-	-	+	+	+
34	+	+	+	+	+	+	+
35	+	+	+	-	+	+	+
36	+	+	-	-	-	+	+
37	+	+	-	-	-	+	+
38	+	+	+	+	+	+	+
39	+	+	+	-	-	+	+
40	+	+	+	-	+	+	+
41	+	+	+	+	+	+	+
42	+	+	-	-	-	+	+

**Table 3** Participant visual preferences

Condition/ participant	Type of visual stimulus						
	Picture- Self	Video- Self	Picture-familiar Other	Video-familiar Other	Picture–TV/movie character	Video–TV/movie character	Video game
<i>Self video</i>							
1	+	+	+	+	+	+	+
2	+	+	+	+	+	+	+
3	+	+	+	–	+	+	+
4	+	+	+	+	+	+	+
5	+	0	+	+	+	+	+
6	+	+	+	0	+	+	+
7	+	+	+	+	+	+	+
8	+	+	+	+	+	+	+
9	+	+	+	+	+	+	?
10	+	+	+	+	+	+	+
11	+	+	+	+	+	+	D
12	+	+	+	+	+	+	0
13	+	+	+	?	+	+	0
14	+	+	+	+	+	+	D
<i>Other VR</i>							
15	+	+	+	+	+	+	+
16	+	+	+	+	+	+	0
17	+	+	+	+	+	+	D
18	+	?	+	?	+	+	?
19	+	+	+	+	+	+	–
20	+	+	+	+	+	+	?
21	+	+	+	+	+	+	+
22	+	+	+	+	+	+	+
23	+	?	+	?	+	+	0
24	+	+	+	+	+	+	+
25	+	+	+	+	+	+	+
26	+	+	+	+	+	+	+
27	+	+	+	D	+	D	+
28	+	+	+	+	+	+	+
<i>Self VR</i>							
29	+	+	+	+	+	+	+
30	+	+	+	+	+	+	+
31	+	+	+	+	+	+	–
32	+	+	+	+	+	+	+
33	+	+	+	+	+	+	–
34	+	+	+	+	+	+	+
35	+	+	+	+	+	+	+
36	+	+	+	+	–	–	–
37	+	+	+	+	+	+	?
38	+	+	+	+	+	+	+
39	+	+	+	+	+	+	+
40	+	+	+	+	+	+	+
41	+	+	+	+	+	+	+
42	+	+	+	+	+	+	–

## Dependent Measures

As the purpose of the research was to examine the degree to which various ESM engaged the participants, two behaviors associated with engagement were measured. The first of these was gaze, specifically the amount of time that the student looked at the screen during a 2-min presentation of screen-based media. This was measured in seconds, with the maximum possible value being 120 s (the duration of the video segment). The second was vocalizations, which is quantified by the number of discrete utterances (either single or multiword) produced by the student during each 2-min interval.

## Equipment and Materials

The research involved the presentation of four distinct types of screen-based segments. In the baseline condition, each participant saw an identical 2-min clip extracted from the Disney movie *Mary Poppins*<sup>®</sup>. The particular segment was chosen for its moderate activity level; that is, it did not include any rousing musical numbers, and focused on the interaction among the human characters as they rode animated carousel horses. In the Self Video condition, the student viewed a video of him/herself engaged in a familiar, non-aversive activity. A personalized clip was needed for each of the 14 students assigned to this condition, requiring the capture of a unique video segment for each participant. Target students were filmed in such a way that they remained the center of attention, occupying the central portion of the screen, with the camera having followed their movements throughout the activity. Depicted in the segments were such activities as morning aerobics, circle time, free play, and recess.

In the Other VR condition, participants viewed a video of a familiar individual engaged in interaction with the “Orbosity” application within the Vtree I-C-Me<sup>®</sup> immersive virtual reality system. In this application, the interactant is pictured on a beach, surrounded by egg-shaped objects moving through the air. When the interactant reaches out and makes virtual contact with one of the orbs, the orb transforms into a bird and flies away. In each of the schools attended by participating students, a familiar person (e.g., the physical education teacher) agreed to be filmed during interaction with the Orbosity application. These individuals were chosen based on teachers’ impressions of the models’ positive relationship with the participants (Apple et al. 2005). The system was configured to record the session directly to videotape. As I-C-Me enables selection of the duration of each interaction segment, the segment was preset at 120 s.

In the Self VR condition, participants engaged directly with the I-C-Me Orbosity application. A small camera

captured the image of the student, which was then displayed on the screen against the Orbosity application’s beach backdrop. When participant movement resulted in contact with one of the egg-shaped objects, the participant witnessed its change into a bird.

In all four viewing conditions, students were seated in a chair placed 55 inches from a 42-inch Luxor AV42 media cart containing the image capture, display, and recording equipment. Behind the students was a large green screen. On the top shelf of the cart was a Toshiba MV20Q41 TV/VCR unit with a 20-inch screen. This unit displayed the visual stimuli in each condition. The VCR component played the video segments used in the baseline condition as well as in the Self Video and Other VR conditions. The TV/VCR was connected to an HP Workstation XW 4300 via an S-video signal converter (Model #15-1238) to an RCA video cable; this configuration was required for display of the Self VR stimuli on the TV monitor. Also on the top shelf of the cart was a small fire wire camera (PYRO 1394 WebCam Version 1) that was used for image capture during the Self VR condition. This camera, mounted on a 3.5-inch tripod, was positioned in front of the TV/VCR unit just below the viewing screen. Mounted on top of the monitor was a Canon Optura 600 DV Camcorder that was used to record participant behavior in each condition for subsequent analysis. The TV/VCR was used to record the screen display generated during the Self VR condition.

Each participant viewed two screen-based segments, a baseline segment (*Mary Poppins*) and a segment dictated by the condition to which they had been assigned (Self Video, Other VR, Self VR). A small piece of black paper was affixed to the screen in the VR conditions to occlude visual access to the on-screen counter built into the Orbosity application; some individuals with ASD are obsessed with such counters, and this distractor needed to be eliminated to ensure the integrity of the data.

Data collection procedures were piloted with four students to enable the researchers to verify the adequacy of the instrumentation and procedures. The data generated during this pilot were not included in the study, but were utilized in the process of establishing reliability during data coding.

Participant data were recorded serially on mini DV tapes, which were subsequently converted to DVD to facilitate data management and coding. Each participant’s segments were isolated and, if necessary, trimmed to include only the 120 s in which the participant viewed the experimental stimuli. Using VideoRedo Plus<sup>®</sup>, a software package that enables digital clips to be viewed in their entirety as well as frame-by-frame, researchers reviewed the clips for instances of the dependent variables. The software enables the onset and offset of target behaviors to be tagged, yielding both a record of segments with onset and offset times and a cumulative record of time spent

engaging in the target behavior. For example, when coding the segments for gaze, the researcher set a marker at the point at which the participant looked away from the screen, and then another when gaze returned to the screen. The system recorded the duration of the off-target looking and generated a cumulative total of those segments. Vocalizations were coded by watching the data segments in real time, recording each instance of a vocalization behavior, and transcribing each utterance.

**Reliability**

Prior to coding the video segments generated by the research participants, two individuals—a state-certified teacher and a nationally-certified speech/language pathologist—independently coded one pilot segment for the occurrence and duration of target behaviors. Discrepancies between the coders relative to the occurrence and duration of behaviors were resolved by simultaneous viewing of the clips by the two coders until discussion yielded 100% agreement about occurrence and duration. Subsequently, each coder independently coded four pilot segments two times, with a three-day timespan between instances. Reliability was established when the coders’ two observations of the cumulative duration of gaze behavior during a 2-min segment agreed within a 0.5 s margin. Once the coders’ reliability reached this criterion, coding of the experimental data commenced.

Interrater reliability was assessed for 27 of the 84 data segments (roughly one-third of total segments). Reliability was assessed relative to the number of vocalizations in each video segment as well as to the onset and offset times for the gaze behaviors. Relative to the number of vocalizations, reliability was calculated by dividing the number of agreements by the number of agreements and disagreements and multiplying by 100. Reliability for vocalizations was 94%. For gaze, a point-by-point comparison of onset and offset times recorded by each coder was conducted. Agreement for gaze behaviors was recorded if the onset/offset times differed by no more than .90 s between coders; reliability was 86.1% (519 out of 603 onsets/offsets).

**Procedures**

During collection of baseline data, participants were brought individually to the room housing the data collection equipment. They were seated in a chair facing the monitor. A researcher said, “[participant’s name], watch,” and the Mary Poppins video segment was played. Participants received no other cues during the session unless they left their seat, at which time they were instructed to

return to their seat by the researcher. At the conclusion of the session, participants were escorted back to their classrooms.

Regarding collection of the second data set, the procedures paralleled those followed during baseline data collection. Participants were brought individually to the data collection room and were seated in a chair facing the monitor. In the Self Video or Other VR conditions, a researcher said, “[participant’s name], watch,” and the appropriate video segments were played. Participants received no other cues during the session unless they left their seat, at which time they were instructed to return to their seat by the researcher. In the Self VR condition, the initial procedure was identical, with the exception of the provision of one additional prompt to the participants. If participants did not immediately begin interacting with the virtual environment, the researcher provided a prompt to “touch the eggs.” No other prompts were provided. At the conclusion of the 2-min segment, participants were escorted back to their classrooms.

**Results**

The results are presented in Tables 4–6. For reader convenience, data relative to age and expressive language are repeated in these tables. The tables present the data relative

**Table 4** Results: gaze and vocalization behaviors in baseline and self video conditions

Participant	Exp Lang 50+	CA	Condition			
			Baseline		Self video	
			Gaze	Vocalization	Gaze	Vocalization
1	+	8;4	110	9	93	4
2 <sup>a</sup>	+	6;1	114	1	114	3
3 <sup>a</sup>	–	6;2	107	7	111	6
4	+	6;2	115	0	104	0
5	+	14;4	118	0	116	0
6 <sup>a</sup>	+	13;6	120	1	111	7
7	+	14;2	109	1	98	0
8	+	13;4	93	0	113	0
9	–	12;6	75	1	94	0
10	+	7;11	102	0	109	0
11	+	8;8	111	0	92	2
12 <sup>a</sup>	–	13;3	116	4	110	1
13	–	10;5	40	2	58	0
14	+	6;9	104	0	87	0
Unadjusted mean			102.4	1.9	100.7	1.6

<sup>a</sup> The participant produced meaningful vocalizations during data collection

**Table 5** Results: gaze and vocalization behaviors in baseline and other VR condition

Participant	Exp Lang 50+	CA	Condition			
			Baseline		Other VR	
			Gaze	Vocalization	Gaze	Vocalization
15	–	8;9	66	7	53	4
16 <sup>a</sup>	–	12;9	114	0	72	14
17	–	13;7	101	0	87	1
18	–	18;8	15	11	28	4
19	–	14;2	107	0	85	12
20	–	6;11	87	6	79	1
21	–	7;5	111	0	92	2
22	–	8;4	109	6	105	2
23	+	12;6	100	0	93	1
24	+	10;4	111	0	110	6
25	+	8;10	118	0	116	0
26 <sup>a</sup>	+	6;9	55	2	64	21
27 <sup>a</sup>	+	6;5	89	1	84	12
28 <sup>a</sup>	+	7;5	104	7	115	9
Unadjusted mean			91.9	2.9	84.5	16.4

<sup>a</sup> The participant produced meaningful vocalizations during data collection

**Table 6** Results: gaze and vocalization behaviors in baseline and self VR condition

Participant	Exp Lang 50+	CA	Condition			
			Baseline		Self VR	
			Gaze	Vocalization	Gaze	Vocalization
29 <sup>a</sup>	–	6;0	119	1	119	0
30 <sup>a</sup>	+	6;10	114	5	117	11
31	+	14;2	66	5	106	3
32	+	8;5	111	0	120	0
33	+	13;9	104	0	117	0
34	–	12;6	111	0	120	5
35	+	12;8	117	2	119	7
36	–	13;8	97	0	88	0
37	–	16;0	41	0	15	1
38	–	6;8	95	7	107	12
39	–	12;1	115	0	116	0
40	–	17;1	70	0	81	4
41	+	10;10	110	0	116	5
42	–	11;8	67	0	80	5
Unadjusted mean			95.5	1.4	101.5	3.8

<sup>a</sup> The participant produced meaningful vocalizations during data collection

to the dependent measures (gaze, vocalization) for both conditions to which the participants were exposed. In the gaze column, the number of seconds the participant looked

**Table 7** Means and standard deviations for unadjusted and adjusted means among the three groups on gaze

Group	Unadjusted (observed) posttest mean	Adjusted (covaried) mean at posttest
Self video	100.7	94.8
Other VR	84.5	88.5
Self VR	101.5	103.5

*Note:* Values rounded at first decimal point for convenient presentation

at the screen (out of a possible 120) is reported. The number of utterances produced by the participant in a 120-s time period is represented in the vocalization column. Those with a notation next to their participant number produced meaningful words during either the baseline or second conditions.

The data were analyzed with regard to group assignment—Self Video, Other VR, and Self VR—and the dependent variables of gaze (measured in seconds) and vocalization (the number produced). Each dependent variable was evaluated on the interval level of measurement. For each participant, the dependent variables were measured during the viewing of the animated video (baseline) and during exposure to one of three comparison conditions (Self Video, Other VR, or Self VR).

Data were analyzed using analyses of covariance (ANCOVA). This method of analysis permitted us to control for variation among the participants due to age, language levels, and performance on the baseline measure. An ANCOVA was completed relative to each dependent variable. For each ANCOVA, baseline scores for the dependent variable under consideration served as covariate. In addition, children's age and their expressive language level served as the second and third covariate. Adjusted posttest means served as the dependent variables. Post hoc comparisons were apportioned using the Bryant-Paulson adjustment.

Results are presented separately according to the dependent variable under consideration. Table 7 presents adjusted (i.e., covaried) and unadjusted (i.e., obtained) posttest means for the three groups on the dependent variable of gaze. The ANCOVA was significant ( $F = 4.64$ ,  $df [2, 36]$ ,  $p = .001$ ). Adjusted posttest means revealed that children in the Self VR condition obtained the highest gaze scores. Post hoc comparisons revealed that the performance of those in the Self VR condition was significantly higher than children in the Other VR condition ( $p = .014$ ). There was no significant difference between children in the Self VR condition and those in the Self Video condition ( $p = .28$ ). Lastly, there was no difference between children in the Self Video condition and those in the Other VR condition ( $p = .67$ ).

**Table 8** Means and standard deviations for unadjusted and adjusted means among the three groups on vocalizations

Group	Unadjusted (observed) posttest mean	Adjusted (covaried) mean at posttest
Self video	1.6	1.3
Other VR	6.4	6.2
Self VR	3.8	4.3

*Note:* Values rounded at first decimal point for convenient presentation

Table 8 presents adjusted and unadjusted posttest means for the three groups on the dependent variable of vocalization. The ANCOVA was significant ( $F = 3.81$ ,  $df [2, 36]$ ,  $p = .031$ ). Adjusted posttest means revealed that children in the Other VR condition obtained the highest vocalization scores. Post hoc comparisons revealed that their performance was significantly higher than children in the Self Video condition ( $p = .03$ ). There was no significant difference between children in the Other VR condition and those in the Self VR condition ( $p = .32$ ). In addition, there was no difference between children in the Self Video condition and those in the Self VR condition ( $p = .86$ ).

## Discussion

This study sought to examine the differential impact of a variety of video-based ESM on the engagement of students with autism. The literature suggests that such media are popular among individuals with ASD, yet there are no reports in the literature directly comparing the relative appeal of various ESM forms. Some reports indicate that research participants are drawn to depictions of familiar individuals or characters (Ogletree et al. 1995), some reports suggest that seeing oneself may have value (Mechling et al. 2006), and research with populations of those with physical and intellectual disabilities indicates that immersive virtual reality is effective in capturing and maintaining participant attention (Weiss et al. 2003).

We were particularly interested in exploring the appeal of an emerging genre of ESM—immersive virtual reality—relative to other types of ESM. For comparison purposes, we chose one contrasting condition that was similar in terms of seeing oneself depicted on the screen (Self Video), and one that included the same VR game, albeit with someone else paying it (Other VR). All comparisons were anchored with data relative to an animated video, a form of ESM already known to elicit rapt attention from many with ASD. The comparison conditions were selected to increase our understanding of those characteristics of ESM that individuals with ASD find most compelling. For example, if participants are found to respond robustly to seeing

themselves in the display, the data should be relatively comparable in the Self Video and Self VR conditions. If they merely are attracted to movement within the visual display, data should be similar across conditions. If the participants are drawn to the nature of the VR applications, the Other VR and Self VR data should be similar. If there is something uniquely motivating about the student's real-time interaction with the immersive VR application, that would be reflected in the data from in the Self VR condition.

The fact that the baseline condition was itself so successful in drawing and maintaining attention turned out to be somewhat problematic, as it was difficult to demonstrate marked differences between the baseline and comparison conditions due to ceiling effects. To increase the sensitivity of the study to differences between baseline and the comparison conditions, we might have conducted it over a longer period of time to highlight discrepancies/drop-offs in attention, or we might have conducted it in the presence of a distractor to heighten competition for attention.

The relative appeal of the comparison conditions varied depending on the measure used to gauge engagement. The Other VR condition appeared to have the edge when vocalization was the engagement indicator. This is due in large part to the subset of students assigned to this condition who became verbally effusive at the sight of a favored person on the screen. The number of vocalizations increased from 0 to 14 for one child, 0 to 12 for a second, 1 to 12 for a third, and 2 to 21 for a fourth. Of these four children, two had been characterized as “low language.” Further, dramatically increased vocalization was not characteristic of the youngsters' real-world encounters with these individuals. What we cannot state conclusively, however—due to limitations in the number of comparison conditions—is whether the appeal derived from depiction of a favored person in a video, from visualization of the favored person engaged in the VR game, or a combination of both. The lack of a significant difference between the Other VR and Self VR condition relative to vocalizations suggests that the VR game itself held notable appeal.

The Self VR condition appeared to have the edge when gaze duration was considered. Although real-time interaction with the immersive VR game was only possible when the youngsters were looking at the screen, an alternate explanation is that the youngsters enjoyed seeing themselves on the screen. This interpretation is strengthened by the fact that there was no significant difference between the Self VR and Self Video conditions relative to gaze.

The sheer novelty of the Self VR condition may have drawn the participants' gaze to the display. Only 4 of the 14 participants in this group were reported to have had previous experience with immersive video. Several students appeared to be intrigued by the sight of themselves

on the screen in this condition. All participants in the Self VR condition had been reported by informants to both recognize themselves in videos and enjoy seeing themselves in videos. What we do not know is which aspect of the Self VR condition—seeing themselves, seeing themselves superimposed against a cartoon background, and/or being able to impact the action on the screen—accounted for the notable gaze durations observed. During collection of pilot data, one student was observed to turn his gaze from the display and search in the space surrounding him, presumably for the eggs and birds he had seen depicted on the screen.

### Reflections on the Stimuli

Consistent with conventional wisdom, most participants appeared to be highly interested in all ESM viewing conditions. Even the Mary Poppins video, the most typical of the conditions, generally attracted participants' visual attention. Most of the students were drawn to it initially, although attention flagged for many students later in the segment. We purposefully selected a segment that was neither action-packed nor part of a lively musical number. Many students were noted to increase attention to the video when there was a transition marked by a noticeable change in the music or the introduction of new sounds (e.g., a horn sounding the commencement of a foxhunt and the excited barking of dogs). In these cases, the audio track appeared to have the potential to re-engage diminished visual attention.

Most of the children appeared to accept unquestioningly and without distress the appearance of their images on the screen. Many of the participants exposed to the Self Video and Self VR conditions registered surprise and delight at seeing themselves on the display. A similar response was observed in the Other VR condition. In fact, several participants in the Other VR condition who had produced few or no vocalizations during baseline became verbally effusive at the sight of a favored person on the display, often calling to or naming the individual.

### Data Coding Challenges

Several characteristics of the participants' behavior posed challenges for the data coders. First, several students rocked rapidly in their seats or jumped up during data collection, making determination of their gaze direction difficult. When the participants left the range of the camera recording their behavior, no gaze behaviors were recorded during their period of absence, even if the target behaviors were in evidence when they left the field of view and returned shortly thereafter. Informants indicated that several participants were known to rock, jump, or elope when they found something to be interesting or fun, so it is

possible that these behaviors, in their extreme, precluded us from recording the full extent of the participants' gaze. Indeed, several students were observed to continue fixation on the visual stimulus even when they were far beyond the range of the camera. It tended to be these extreme behaviors that accounted for most of the variability between coders reported in the reliability section.

### Language Variability

In comparison with many studies of the use of video technology with those having ASD, most of the participants in this study had more limited language competence. Twenty participants (48%) had expressive lexicons of 50 words or less, and only 12 (29%) were reported to have expressive lexicons of more than 100 words. Thus, a great many participants who were not designated as "low language" were nonetheless markedly delayed in their expressive communication skills. Further, Table 1 illustrates the variability across communication domains that characterizes the population of individuals with ASD. Among the participants in this study, relatively high receptive and expressive lexicons were not necessarily correlated with the appropriate use of gestures or a tendency to engage in conversation.

Another point of interest is the notable lack of concordance between reported language skills and appropriateness of communication during data collection. Only 10 participants out of 42 were noted to have made meaningful remarks during either baseline or second condition data collection. Forty percent of these participants were those designated as having "low language," further reinforcing the splintered nature of these youngsters' language profiles. Indeed, many participants were completely silent during one or both data collection segments, despite the fact that informants reported that all but one of the participants tended to vocalize during activities they found to be interesting or fun. These seemingly anomalous results are a reminder of the risk inherent in making assumptions about the presence of certain language behaviors based on observation of other—seemingly related—behaviors within the population of those with ASD.

### Keeping the Results in Context

The ideal research design would have matched participants across conditions according to characteristics such as age and language profile. This was not achievable for a variety of reasons, including the notable heterogeneity of the population and the limitations imposed by the prevailing circumstances in the schools in which data collection took place. The decision to utilize a random assignment design was made, in part, by the fact that it was not possible to

equate or match participants using standardized scores, such as those derived from the Autism Diagnostic Interview-Revised (ADI-R) (Lord et al. 1994) or the Autism Diagnostic Observation Schedule (ADOS) (Lord et al. 2000), because such assessment data were not consistently available from the youngsters' educational records. Alternatively, we might have employed stratified sampling, but the nature of the "rolling enrollment" of youngsters precluded this approach. Instead, we accounted for inherent differences among participants statistically by co-varying age and expressive language level in the analysis.

We developed a multidimensional scheme for characterizing the youngsters' communication profiles which relied on parent and teachers completion of an extensive questionnaire. A limitation of the study is its reliance on informant reports. It would have been preferable for the research team to have had the opportunity to corroborate these reports through home and/or classroom observation, but that was not feasible given the relatively large number of participants, limited personnel resources, and the fact that the looming conclusion of the school year imposed severe time constraints. The elicitation of identical data from both teachers and parents provided a source of corroboration for data that were used to describe participant characteristics.

This study was a first attempt to examine the relative engagement potential of various types of ESM, and the data reflect differences in responding across experimental conditions. It is not the case, however, that one type of ESM is necessarily "better" at engaging children than another. Rather, different ESM types occasioned different patterns of responding across the key variables of gaze and vocalizations. It will be important for future research to include additional conditions that enable more systematic variation of the many dimensions on which the comparison conditions varied: reality vs. fantasy, novelty, degree of active engagement, and viewing of self versus another person (varying in familiarity as well as appeal).

It is important to remember that this study was a fundamental comparison of engagement potential, not of the instructional potency of the various types of video-based ESM. As engagement is inextricably linked with instructional power, however, it is vital to consider an instructional application's engagement quotient. The most exquisitely designed instructional paradigm will be ineffective if students are not compelled to attend to it. Likewise, emerging technologies can generate tremendous excitement among parents and teachers because of students' utter fascination, if not obsession, with them. As essential as engagement is, however, it does not guarantee that the application will facilitate the learning and generalization of new skills. We must ensure that the allure of new technology does not eclipse the focus on instructional value.

## Conclusion

The engagement potential of video-based technology among people with ASD is legendary. This study was the first to compare the relative appeal of four types of video-based electronic screen media using a systematic protocol, and its findings document that all types of stimuli included in the study generally held student visual attention reasonably well for a short period of time, with opportunities to see oneself portrayed on the screen generating greater gaze durations than seeing another person. With regard to vocalization, the VR conditions occasioned more vocalizations than did the more traditional video, even when it portrayed the youngster himself/herself. The extent to which these findings would be maintained under viewing conditions of greater duration has yet to be demonstrated; the relatively robust responses observed during this study may diminish as the novelty of the application wears off. More importantly, this study only considered engagement, and the much more important question is whether engagement, once garnered and maintained, will impact learning in a positive way. It appears from numerous video modeling studies that this is the case. Currently, the vast majority of immersive VR applications are for entertainment or physical conditioning purposes. Whether this technology can be harnessed as a vehicle for effective instruction for those with ASD remains to be seen.

**Acknowledgments** This work was conducted on behalf of the Pennsylvania Department of Education, Bureau of Special Education. The authors are grateful to the students, families, and educational personnel associated with the following programs of the Bucks County Intermediate Unit #22: Maple Point Middle School, McDonald Elementary School, Pearl S. Buck Elementary School, Pennwood Middle School, Tawanka Learning Center, William Penn Middle School, and Willowdale Elementary School, and the Francis Harvey Green School of Delaware County Intermediate Unit #25.

## References

- Alberto, P. A., Cihak, D. F., & Gama, R. I. (2005). Use of static picture prompts versus video modeling during simulation instruction. *Research in Developmental Disabilities, 26*, 327–339. doi:10.1016/j.ridd.2004.11.002.
- Apple, A. L., Billingsley, F., & Schwartz, I. S. (2005). Effects of video modeling alone and with self-management on compliment-giving behaviors of children with high-functioning ASD. *Journal of Positive Behavior Interventions, 7*, 33–46. doi:10.1177/10983007050070010401.
- Beardon, L., Parsons, S., & Neale, H. (2001). An interdisciplinary approach to investigating the use of virtual reality environments for people with Asperger syndrome. *Educational and Child Psychology, 18*, 53–62.
- Bellini, S., & Akullian, J. (2007). A meta-analysis of video modeling and video self-modeling interventions for children and adolescents with autism spectrum disorders. *Exceptional Children, 73*, 264–287.

- Bernard-Opitz, V., Sriram, N., & Nakhoda-Sapuan, S. (2001). Enhancing social problem solving in children with autism and normal children through computer-assisted instruction. *Journal of Autism and Developmental Disorders*, *31*, 377–398. doi: [10.1023/A:1010660502130](https://doi.org/10.1023/A:1010660502130).
- Buggey, T. (2005). Video self-modeling applications with students with autism spectrum disorder in a small private school setting. *Focus on Autism and Other Developmental Disabilities*, *20*, 52–63.
- Charlop-Christy, M. J., & Daneshvar, S. (2003). Using video modeling to teach perspective taking to children with autism. *Journal of Positive Behavior Interventions*, *5*, 12–21. doi: [10.1177/10983007030050010101](https://doi.org/10.1177/10983007030050010101).
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Coyle, C., & Cole, P. (2004). A videotaped self-modelling and self-monitoring treatment program to decrease off-task behaviour in children with autism. *Journal of Intellectual & Developmental Disability*, *29*, 3–15. doi: [10.1080/08927020410001662642](https://doi.org/10.1080/08927020410001662642).
- DeFanti, T. A., Sandin, D. J., & Cruz-Neira, C. (1993). A 'room' with a 'view'. *IEEE Spectrum*, *30*, 30–33, 39. doi: [10.1109/6.237582](https://doi.org/10.1109/6.237582)
- Dowrick, P. W. (1999). A review of self-modeling and related interventions. *Applied & Preventive Psychology*, *8*, 23–39. doi: [10.1016/S0962-1849\(99\)80009-2](https://doi.org/10.1016/S0962-1849(99)80009-2).
- Gena, A., Couloura, S., & Kymissis, E. (2005). Modifying the affective behavior of preschoolers with autism using in-vivo or video modeling and reinforcement contingencies. *Journal of Autism and Developmental Disorders*, *35*, 545–556. doi: [10.1007/s10803-005-0014-9](https://doi.org/10.1007/s10803-005-0014-9).
- Grandin, T. (1996). *Thinking in pictures and other reports from my life*. New York: Vintage.
- Harris, K., & Reid, D. (2005). The influence of virtual reality play on children's motivation. *Canadian Journal of Occupational Therapy*, *72*, 21–29.
- Heiman, M., Nelson, K. E., Tjus, T., & Gilberg, C. (1995). Increasing reading and communication skills in children with autism through an interactive multimedia program. *Journal of Autism and Developmental Disorders*, *25*, 459–480. doi: [10.1007/BF02178294](https://doi.org/10.1007/BF02178294).
- Holden, M. (2005). Virtual environments for motor rehabilitation. *Cyberpsychology & Behavior*, *8*, 187–211. Review. doi: [10.1089/cpb.2005.8.187](https://doi.org/10.1089/cpb.2005.8.187)
- Jacobson, J. W., Foxx, R. M., & Mulick, J. A. (2005). *Controversial therapies for developmental disabilities: Fad, fashion and science in professional practice*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Joseph, R. M., Tager-Flusberg, H., & Lord, C. (2002). Cognitive profiles and social-communicative functioning in children with autism spectrum disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *43*, 801–821. doi: [10.1111/1469-7610.00092](https://doi.org/10.1111/1469-7610.00092).
- Kinney, E. M., Vedora, J., & Stromer, R. (2003). Computer-presented video models to teach generative spelling to a child with an autism spectrum disorder. *Journal of Positive Behavior Interventions*, *5*, 22–29. doi: [10.1177/10983007030050010301](https://doi.org/10.1177/10983007030050010301).
- Kobayashi, R., Murata, T., & Youshinaga, K. (1992). A follow-up study of 201 children with autism in Kyushu and Yamaguchi areas, Japan. *Journal of Autism and Developmental Disorders*, *22*, 395–411. doi: [10.1007/BF01048242](https://doi.org/10.1007/BF01048242).
- Lord, C., Risi, S., Lambrecht, L., Cook, E. H., Leventhal, B. L., DiLavore, P. C., et al. (2000). The Autism Diagnostic Observation Schedule–Generic: A standard measure of social and communication deficits associated with the spectrum of autism. *Journal of Autism and Developmental Disorders*, *30*, 205–223. doi: [10.1023/A:1005592401947](https://doi.org/10.1023/A:1005592401947).
- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview–Revised: A revised version of the diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, *24*, 659–685. doi: [10.1007/BF02172145](https://doi.org/10.1007/BF02172145).
- MacDonald, R., Clark, M., Garrigan, E., & Vangala, M. (2005). Using video modeling to teach pretend play to children with autism. *Behavioral Interventions*, *20*, 225–238. doi: [10.1002/bin.197](https://doi.org/10.1002/bin.197).
- Mechling, L. C., Gast, D. L., & Cronin, B. A. (2006). The effects of presenting high-preference items, paired with choice, via computer-based video programming on task completion of students with autism. *Focus on Autism and Other Developmental Disabilities*, *21*, 7–13.
- Nally, B., Houlton, B., & Ralph, S. (2000). The management of television and video by parents of children with autism. *Autism*, *4*, 331–337. doi: [10.1177/136236130004003008](https://doi.org/10.1177/136236130004003008).
- National Research Council. (2001). *Educating children with autism*. Committee on Educational Interventions for Children with Autism. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.
- Nikopoulous, C. K., & Keenan, M. (2004). Effects of video modeling on social initiations by children with autism. *Journal of Applied Behavior Analysis*, *37*, 93–96. doi: [10.1901/jaba.2004.37.93](https://doi.org/10.1901/jaba.2004.37.93).
- Ogletree, B. T., Fischer, M., & Sprouse, J. (1995). An innovative language treatment for a child with high-functioning autism. *Focus on Autistic Behavior*, *10*, 1–10.
- Parsons, S., Mitchell, P., & Leonard, A. (2005). Do adolescents with autistic spectrum disorders adhere to social conventions in virtual environments? *Autism*, *9*, 95–117. doi: [10.1177/1362361305049032](https://doi.org/10.1177/1362361305049032).
- Pierce, K., & Schreibman, L. (1994). Teaching daily living skills to children with autism in unsupervised settings through pictorial self-management. *Journal of Applied Behavior Analysis*, *27*, 471–481. doi: [10.1901/jaba.1994.27.471](https://doi.org/10.1901/jaba.1994.27.471).
- Polys, N. F., Brutzman, D., Steed, A., & Behr, J. (2008). Future standards for immersive VR. *IEEE Computer Graphics and Applications*, *28*, 94–104. doi: [10.1109/MCG.2008.29](https://doi.org/10.1109/MCG.2008.29).
- Quill, K. (1997). Instructional considerations for young children with autism: The rationale for visually cued instruction. *Journal of Autism and Developmental Disorders*, *27*, 697–714. doi: [10.1023/A:1025806900162](https://doi.org/10.1023/A:1025806900162).
- Rose, F. D., Brooks, B. M., & Rizzo, A. A. (2005). Virtual reality in brain damage rehabilitation. *Cyberpsychology & Behavior*, *8*, 241–262. Review. doi: [10.1089/cpb.2005.8.241](https://doi.org/10.1089/cpb.2005.8.241)
- Schreibman, L., Whalen, C., & Stahmer, A. (2000). The use of video priming to reduce disruptive transition behavior in children with autism. *Journal of Positive Behavior Management*, *2*, 3–11. doi: [10.1177/109830070000200102](https://doi.org/10.1177/109830070000200102).
- Seltzer, M. M., Krauss, M. W., Shattuck, P. T., Orsmond, G., Swe, A., & Lord, C. (2003). The symptoms of autism spectrum disorders in adolescence and adulthood. *Journal of Autism and Developmental Disabilities*, *33*, 565–581. doi: [10.1023/B:JADD.000005995.02453.0b](https://doi.org/10.1023/B:JADD.000005995.02453.0b).
- Shane, H. C., & Albert, P. D. (2008). Electronic screen media for persons with Autism Spectrum Disorders: Results of a survey. *Journal of Autism and Developmental Disorders*, Published online February 22, 2008. doi: [10.1007/s10803-007-0527-5](https://doi.org/10.1007/s10803-007-0527-5).
- Shane, H. C., Sorce, J., Cordeiro, R. F., Duggan, M., & Weiss-Kopp, S. (2005). *Video technology for language instruction for children with ASD*. Presentation at the annual convention of the American Speech-Language-Hearing Association, San Diego, CA
- Sherer, M., Pierce, K. L., Paredes, S., Kisacky, K. L., Ingersoll, B., & Schreibman, L. (2001). Enhancing conversation skills in children with autism via video technology: Which is better, "self" or "other" as a model? *Behavior Modification*, *25*, 140–158. doi: [10.1177/0145445501251008](https://doi.org/10.1177/0145445501251008).
- Shipley-Benamou, R., Lutzker, J. R., & Taubman, M. (2002). Teaching daily living skills to children with autism through

- instructional video modeling. *Journal of Positive Behavior Interventions*, 4, 165–175. doi:[10.1177/10983007020040030501](https://doi.org/10.1177/10983007020040030501).
- Simpson, A., Langone, J., & Ayres, K. M. (2004). Embedded video and computer based instruction to improve social skills for students with autism. *Education and Training in Developmental Disabilities*, 39, 240–252.
- Spiker, D., & Ricks, M. (1984). Visual self-recognition in autistic children: Developmental relationships. *Child Development*, 55, 214–225. doi:[10.2307/1129846](https://doi.org/10.2307/1129846).
- Tager-Flusberg, H. (2004). Strategies for conducting research on language in autism. *Journal of Autism and Developmental Disorders*, 34, 75–80. doi:[10.1023/B:JADD.0000018077.64617.5a](https://doi.org/10.1023/B:JADD.0000018077.64617.5a).
- Tissot, C., & Evans, R. (2003). Visual teaching strategies for children with autism. *Early Child Development and Care*, 173, 425–433. doi:[10.1080/0300443032000079104](https://doi.org/10.1080/0300443032000079104).
- Trepagnier, C. G. (1999). Virtual environments for the investigation and rehabilitation of cognitive and perceptual impairments. *NeuroRehabilitation*, 12, 63–72.
- Ward, C. D. (2005). Commentary on Standen, P. J., & Brown, D. J., Virtual reality in the rehabilitation of people with intellectual disabilities. *Cyberpsychology & Behavior*, 8, 288. Review.
- Weiss, P. L., Bialik, P., & Kizony, R. (2003). Virtual reality provides leisure time opportunities for young adults with physical and intellectual disabilities. *Cyberpsychology & Behavior*, 3, 335–342. doi:[10.1089/109493103322011650](https://doi.org/10.1089/109493103322011650).
- Wert, B. Y., & Neisworth, J. T. (2003). Effects of video self-modeling on spontaneous requesting in children with autism. *Journal of Positive Behavior Interventions*, 5, 30–34. doi:[10.1177/10983007030050010501](https://doi.org/10.1177/10983007030050010501).
- Wilkinson, K. M. (1998). Profiles of language and communication skills in autism. *Mental Retardation and Developmental Disabilities Research Reviews*, 4, 73–79. doi:[10.1002/\(SICI\)1098-2779\(1998\)4:2<73::AID-MRDD3>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1098-2779(1998)4:2<73::AID-MRDD3>3.0.CO;2-Y).
- Woltersdorf, M. A. (1992). Videotape self-modeling in the treatment of attention-deficit hyperactivity disorder. *Child & Family Behavior Therapy*, 14, 53–73. doi:[10.1300/J019v14n02\\_04](https://doi.org/10.1300/J019v14n02_04).