CrossMark

SURVEY

Are Robots Ready to Deliver Autism Interventions? A Comprehensive Review

Momotaz Begum[1] · Richard W. Serna[2] · Holly A. Yanco[1]

Abstract This article presents a review of the contemporary robotics research with respect to making robots and human–robot interaction (HRI) useful for autism intervention in clinical settings. Robotics research over the past decade has demonstrated that many children with autism spectrum disorders (ASDs) have a strong interest in robots and robot toys and can connect with a robot significantly better than with a human. Despite showing great promise, research in this direction has made minimal progress in advancing robots as clinically useful for ASD intervention. Moreover, the clinicians are generally not convinced about the potential of robots. A major reason behind this is that a vast majority of HRI studies on robot-mediated intervention (RMI) do not follow any standard research design and, consequently, the data produced by these studies is minimally appealing to the clinical community. In clinical research on ASD intervention, a systematic evaluation of the evidence found from a study is performed to determine the effectiveness of an experimental intervention (e.g., a RMI). An intervention that produces a stable positive effect is considered as an evidence-based practice (EBP) in autism. EBPs enable clinicians to choose the best available treatments for an individual with ASD. The ultimate goal of RMI, therefore, is to be considered as an EBP so that they can actually be used for treating autism. There are several criteria to measure the strength of evidence, and they are mostly geared toward rigorous research design. The research on RMI, therefore, needs to follow standard research design to be acceptable by the clinical community. This paper reviews the contemporary literature on robotics and autism to understand the status of RMI with respect to being an EBP in autism treatment. First, a set of guidelines is reported which is considered as a benchmark for research design in clinical research on ASD intervention and can easily be adopted in HRI studies on RMI. The existing literature on RMI is then reviewed with respect to these guidelines. We hope that the guidelines reported in this paper will help the robotics community to design user studies on RMI that meet clinical standards and thereby produce results that can lead RMI toward being considered as an EBP in autism. Note that the paper is exclusively focused on the role of robots in ASD intervention/therapy. Reviews on the use of robots in ASD diagnosis are beyond the scope of this paper.

Keywords Robots · HRI · Autism spectrum disorders · Therapeutic intervention

✉ Momotaz Begum
momotaz.begum.onu@gmail.com

Richard W. Serna
Richard_Serna@uml.edu

Holly A. Yanco
holly@cs.uml.edu

[1] Department of Computer Science, University of Massachusetts Lowell, Lowell, USA

[2] Department of Psychology, University of Massachusetts Lowell, Lowell, USA

1 Introduction

Anecdotal evidence from numerous human–robot interaction (HRI) studies reported in the literature suggests that many individuals with autism spectrum disorders, ASDs (IwASDs) connect noticeably better with robots than humans. Two recent surveys covering a vast majority of HRI studies with different kinds of robots and IwASDs in varying contexts are available in [9,61]. Almost all of these studies demonstrate that many IwASDs express elevated enthusiasm (e.g., increase in attention [33], imitation ability [22], verbal utterances [34], social activities [67], etc.) while interacting with

robots. IwASDs may even have cognitive and/or biological biases toward robots over humans [17]. A number of recent studies suggest neurobiological evidence in favor of such a claim. For example, an fMRI study suggested that adults with ASD may perceive a humanoid robot as a social interaction partner the same way a typically developed adult perceives a fellow human being [10]. Another study showed that robotic movements elicit visuomotor priming in children with ASD (visual priming is a precondition for automatic imitation, a behavior generally absent in children with ASD) [47]. A long line of research is dedicated to the design of robots with appropriate physical features [36,56], control architectures [20], evaluation metrics [60], and HRI algorithms [7,21] that can be used in ASD intervention.

Despite these efforts, the potential end-users of this technology (i.e., IwASDs, their caregivers, and clinicians) are neither aware nor convinced of the role of robots in ASD intervention [17,18]. Recently, a number of systematic reviews and a meta-analysis of the technology-based interventions for IwASDs (that reviewed the literature published before December 2011) have concluded that the robot-based studies with IwASDs fail to meet a set of criteria commonly observed to assess the outcome of an ASD intervention [26,51]. The problem lies in the fact that the vast majority of robotics research in this domain shows the 'likability' of robots to IwASDs but fails to demonstrate a robot's utility in ASD intervention [17,33]. Demonstration of 'likability' is never sufficient to formally allow a robot to co-locate and interact with a protected population such as IwASDs. The necessity of robotics research to understand end-users' requirements and the inadequacy of current research to improve the utility of robots have been discussed thoroughly in a number of recent publications [17,18,33].

Based on these findings and observations, this paper suggests that a promising way to prove the utility of robots in ASD interventions is to establish robot-mediated interventions (RMIs) as an evidence-based practice (EBP) in autism. EBP has become a benchmark for clinicians involved with the research and treatment of autism [28,52]. The clinical literature on ASD intervention has set up guidelines to determine the strength of evidence from an experimental intervention in order to consider it an EBP. On the one hand, many robotics researchers are not aware of such clinical literature (which is an inherent challenge in cross-disciplinary research), and on the other hand, many of the guidelines might be difficult to directly adopt in HRI studies (a research challenge thoroughly discussed in [33]). This paper, based on an thorough cross-disciplinary survey, reports a set of guidelines that should be observed by and can easily be adopted in HRI studies on RMI in order to generate clinically acceptable data and enhance the probability of establishing RMI as an EBP in autism. The paper then presents a review of the existing literature based on these guidelines to understand where the contemporary robotics research stands with respect to making robots useful for ASD intervention in clinical settings.

1.1 Contribution

The only article that performs a critical review of the contemporary literature on the clinical use of robots in ASD therapy and diagnosis was published in a clinical journal [17]. The article [17] reviewed literature published before March 2011 based on four inclusion criteria; 20 articles met those criteria. Findings from these 20 RMI were reviewed based on the following five methodological characteristics: (1) number of participants, (2) report on diagnostic condition, (3) age of the participants, (4) matching of participants in the case of a group-based design, and (5) the method of robot–IwASD interaction during the study. The major differences of our review from the review presented in [17] are: (1) this article exclusively focuses on the use of robots in ASD intervention (i.e., the use of robots in ASD diagnosis is beyond its scope), (2) literature published between 1990 and September 2014 has been reviewed for this article, and (3) this article presents a set of guidelines to drive the HRI research in a direction where robots can establish their clinical utility in ASD intervention. The guidelines, while being based on the clinical literature on ASD intervention, considers the issues and challenges faced by robotics researchers while designing HRI studies on RMI. The article uses these guidelines as the review criteria. It is our intention with this article to help the robotics community understand a number of deficits in contemporary robotics research on RMI and design HRI studies in a way that will help to clearly demonstrate the utility of robots in ASD intervention.

1.2 Review Methodology

The present review utilizes the following methodology.

(1) The review included articles published between 1990 and September 2014 in peer-reviewed conferences, journals, or technical magazines.
(2) To be included in the review, an article must have presented a HRI study where a physical robot (not a computer avatar) was used in an intervention with at least one IwASD. Such an intervention is aimed at improving any aspect of the behavior of an IwASD, such as improving a behavior related to a life-skill or eliminating an interfering behavior, etc.
(3) HRI studies whose sole focus is to gauge an IwASD's response (interest, aversion, etc.) to robots or particular physical features and/or behaviors of robots were not included in this review (e.g., the studies reported in [21,35,56], etc.).

(4) HRI studies whose sole focus is to demonstrate that IwASDs like robots more than their typically developing (TD) peers (e.g., the studies reported in [6,47]) or are capable of interacting with robots the same way as their TD peers (e.g., the study reported in [35]) were excluded from the review.

(5) In cases where an article was published in multiple venues by the same author(s) and that presented results from the same study, the article with the most comprehensive results was considered as the primary article and was included in the review.

(6) Articles whose primary focus is to describe the development of robots, sensors, and software/algorithms that can be helpful in RMI for IwASDs (e.g., [20,22]) were not included in this review.

Criteria 3–6 were used to filter out articles that, we believe, were either more focused on the robotic technology itself, rather than the possible ways to make it useful in ASD intervention, or intended to prove the 'likability' of a robot instead of its demonstrated effectiveness. We defined these criteria because we strongly believe that the robotics research in the past decade has collected sufficient anecdotal evidence to establish the fact that some IwASDs have a strong fascination with robots. Although there are still many unanswered research questions (such as which IwASDs have preferences for robot and which IwASDs do not? Do all robots elicit the same level of interest or a robot's form can modulate the level of interest? etc.), we believe it is time to direct research on robot-mediated ASD intervention toward actual deployment in clinical settings to serve the clinicians and IwASDs.

The rest of the article is organized as follows: Sect. 2 discusses the clinical literature on EBP. Section 3 presents a set of guidelines to establish RMI as an EBP in autism. Section 4 presents a review of the existing literature on RMI with respect to the guidelines reported in Sect. 3. Finally, Sect. 5 provides a discussion on the reviewed articles and Sect. 6 concludes this article. Note that the rest of the article will use the terms 'intervention' and 'therapy' interchangeably.

2 Evidence-Based Practices in Autism and Robot-Mediated Interventions

The concept of EBP was introduced by the field of medicine to minimize the gap between research and practice. EBPs enable physicians to choose methods that have strong scientific evidence from carefully controlled research studies [31]. The definition of EBP is somewhat unique in different disciplines. The American Psychological Association has defined EBP as "the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences" [1]. In the autism literature,

EBP is generally defined as "intervention practices that have been tested in high quality research designs and found efficacious" [45]. Despite the lack of an universal definition of EBP, definitions from diverse areas of professional practice share the core theme that EBP require careful assessment of current research with the goal of identifying interventions that have demonstrated effectiveness. Thus the basic foundation of EBP is the systematic review of evidence from scientific research. With the information on the best available EBPs, clinicians analyze the characteristics of the person with ASD and his/her support network and apply their judgment for making decision regarding interventions to consider and those to avoid. EBP has become a benchmark in ASD intervention. Clinical researchers have established rubrics to evaluate the strength of evidence from experimental ASD interventions and criteria for considering an intervention an EBP [28,42,44,52,58].

Based on these criteria many federal government agencies [e.g., National Professional Development Center (NPDC) on Autism Spectrum Disorders] and nationally recognized non-profit organizations [e.g., National Autism Center (NAC)] perform rigorous systematic reviews of research to identify ASD interventions that can be considered an EBP [68,69]. For example, a review of the research literature from 1997 to 2011 by the NPDC in 2014 yielded 27 EBPs for autism treatment [69]. Among these 27 EBPs, a noteworthy EBP category was *technology-aided instruction and intervention* (TAII). TAIIs are ASD interventions that use technologies to facilitate a positive outcome of an ASD intervention. The term 'technology' was defined as "any electronic item /equipment /application /or virtual network that is used intentionally to increase/maintain, and/or improve daily living, work/productivity, and recreation/leisure capabilities of adolescents with autism spectrum disorders" [43]. The examples of technology included speech-generating devices, smart phones, tablets, computed-assisted instructional programs, and virtual networks. Unfortunately, no RMI was included in the review; the robot was not considered as a 'technology' capable of producing positive outcomes in an ASD intervention. A slightly more positive picture, for HRI researchers, can be derived from the 2009 NAC's National Standards report. This report yielded 33 EBPs (11 practices as 'established' treatment and 22 practices as 'emerging' treatment) for ASD after examining the literature on ASD intervention published between 1957 and the Fall of 2007 [68]. Among the 33 EBPs, *technology-based treatment* was considered an 'emerging' EBP, "although one or more studies suggest that a treatment produces beneficial treatment effects for individuals with ASD, additional high quality studies must consistently show this outcome before we can draw firm conclusions about treatment effectiveness" [68]. Nineteen studies on technology-based ASD intervention, published from 1993 to 2005, qualified to be included in NAC's sys-

tematic review. However, there was only one HRI study on robot-mediated ASD intervention (reported in [55]) that met the inclusion criteria established by the NAC; although the IEEE Xplore and ACM Digital library alone published nearly 300 peer-reviewed articles on robot-mediated ASD intervention within 1993–2005. Compared to RMI, computer-aided interventions (e.g., virtual reality programs), also a new technology similar to robots, have quickly become a promising EBP in autism [68,69]. This may suggest a lack of efforts in the robotics community to conduct HRI studies on robot-mediated ASD intervention that meet clinical standards and generate data that prove the clinical utility of robots in ASD therapy.

As noted above, clinical researchers have proposed guidelines for an experimental intervention to be considered an EBP. Due to the nature of clinical disciplines, these guidelines cover a wide range of issues and factors related to autism. It might not be possible for robotics researchers to directly adopt these guidelines in HRI studies on ASD intervention. This may be because a robot, unlike all other technologies being used in ASD interventions, is a complex piece of technology, and many aspects of its use in ASD interventions are not yet fully understood. For instance, processing power, sensors, the Internet, and artificial intelligence (AI) are far from being in a state where a robot can mimic the role of a human therapist. Wizard-of-Oz (WoZ) control [32] might still be used to explore the utility of robots, but many questions remain regarding such control [53], e.g., who would provide the technical and financial support for deploying a WoZ-controlled robot in clinical settings? Would an autonomous robot be able to mimic the abilities of a WoZ-controlled robot? etc.

The next section describes a set of guidelines that can be adopted by robotics researchers to design RMI that could demonstrate the utility of robots in ASD intervention and could increase the possibility of RMI qualifying as an EBP in autism. The contemporary HRI literature is then reviewed in an effort to understand the current state-of-the art with respect to observing these guidelines.

3 A Road Map to Establish Robot-Mediated Interventions as an EBP in Autism

The guidelines described in this section are based on a comprehensive review of the clinical literature on ASD intervention. Based on these guidelines, HRI studies on RMI for ASD should be designed systematically, while focusing on the following six methodological elements:

(1) *Goal of intervention* What specific clinical goal does the study/intervention seek to achieve and why?
(2) *Participants* What type of participants should be allowed to participate in the study?

(3) *Independent variables* with RMI, the robot itself and its behaviors are independent variables. The hardware of the robot and the supporting software/algorithm/AI should be described with replicable precision.
(4) *Dependent variables* What specific ASD behaviors will be modified by the RMI?
(5) *Research design* What research design is suitable to evaluate the goal of the study given the number or type of participants?
(6) *Generalization training* What is the plan to help the participants generalize the skill trained by a robot to humans?

The following sections will elaborate on these six required design elements of a HRI study on RMI.

3.1 Goal of Intervention

The general purpose of an ASD therapy is to ensure a long-term effect in independent functioning, health and well-being, and quality of life for an IwASD [39]. Thus, therapeutic goals are designed to improve necessary life-skills or eliminate/reduce behaviors that interfere with life functioning so that an IwASD can live an independent, meaningful, and socially active life. HRI studies on RMI must develop clinical goals that have social significance and fulfill the general purpose of an ASD therapy. A recent meta-analysis of six comprehensive systematic reviews on the clinical ASD literature suggests four possible goals for ASD interventions [39]: *social* (teaching life-skills required for social interaction such as joint attention, friendship skills, pretend play, social engagement, social problem solving skills, appropriate participation in group activities, interpersonal skills, etc.), *communication* (teaching life-skills required to convey information to others in a verbal or non-verbal manner such as requesting, labeling, receptive and expressive language, conversation, greetings, speech, pragmatics, etc.), *maladaptive behavior* (eliminating/reducing the behaviors that interfere with the learning or life functioning such as aggression, repetitive behaviors, depression, anxiety, non-functional patterns of behaviors, interest, or activity, etc.), and *academic* (teaching life-skills related to school readiness such as learning readiness, higher cognitive functions, sensor and motor skills, skills required for a specific job, etc.). Robotics researchers should consider these four categories when choosing a goal for the RMI. Whatever behavior/skill an RMI aims to modify, it must be done so with an eye toward contributing to the life and well-being of the IwASDs.

3.2 Participants

After determining the goal of an intervention, it is important for a RMI to define precisely the eligibility criteria for IwASDs to participate in the study (commonly known as

inclusion criteria). This step is important because significant information about the treatment effect can be derived only if all participants meet carefully designed inclusion criteria. For example, a RMI to teach a social skill (e.g., pretend play) must ensure that each participant begins with a low level of performance in executing that specific skill prior to introducing the RMI. Well-established diagnostic tools and procedures should be used to evaluate how a participant performs on that particular class of social skill prior to recruiting him/her in the study. The type of study also influences the inclusion criteria. For example, in the case of single-subject (SS) design (where the focus is to compare the individual improvement of each participant as a result of the intervention) the inclusion criteria are generally less restrictive than group based design (where the focus is on understanding the effect of an intervention on a large population).

After participants are selected based on the inclusion criteria, it is important to collect and report participants' information. Reporting detailed demographic and diagnostic information of the study participants is a standard practice in clinical research. It greatly facilitates drawing statistically valid conclusions from a study on the effect of a treatment on a certain population. In general, a RMI should, at least, report the following information about participants: (1) age, (2) gender, and (3) diagnostic information (diagnosis of autism confirmed by at least one psychometrically solid instrument, e.g., childhood autism rating scale (CARS) [62], social communication questionnaire (SCQ) [59], social responsiveness scale [11,12], autism diagnostic observation schedule (ADOS/ADOS-G) [37,38], etc.). Other information such as IQ, mental age, co-occurring medical conditions, personal history, etc., can also be provided, if available, to present a complete picture of the participant pool. It is important to ensure that participants are described using standard terms so that any other RMI can recruit participants of similar nature based on that description.

3.3 Independent Variable

The independent variable in any RMI is the RMI itself. The common goal of all RMIs is to show how the use of a robot (as a co-therapist, therapist, or simply as an intervention tool) can improve the therapeutic outcomes. There are two important factors to consider with regard to the independent variable: design and reproducibility. They are discussed below.

(1) *Design* A robot-mediate intervention is a novel type of ASD intervention only in the sense that it uses a new tool (i.e., the robot) to replace or augment a human therapist. Few recent articles [16,18] discuss possible roles of a robot in robot mediated therapies, e.g., robot as a sole therapist, robot as a mediator or assistant in a therapy, etc. Irrespective of the role of a robot, the therapy itself (i.e.,

how the robot will play its designated role to achieve a therapeutic goal) should conform to standard, clinically established methodologies. Close collaboration with a domain expert (e.g., psychologist, behavioral scientist, therapist, etc.) could prove useful in this regard. There are a number of approaches that have proven to be effective in ASD intervention, e.g., applied behavior analysis (ABA) [27], early start Denver model [57], structured teaching (TEACCH) [41], etc. Use of any of such established approaches or a combination of them to design a RMI will ensure, on one hand, that the fidelity of the intervention itself will not be questionable to the clinical community, and on the other hand, that the robot can be seamlessly integrated in the existing clinical practices on ASD interventions. It will greatly enhance the probability of robots to be deployed in clinical settings.

(2) *Reproducibility* Reproducibility indicates how well an intervention can be reproduced by a third-party, given a documentation of the intervention process and a similar clinical population. Reproducibility, therefore, is a crucial aspect of ASD intervention design, which could contribute greatly to the popularity of an intervention among broader community and, ultimately, to its assessment as an EBP in autism. The important considerations for robotics researchers to design a replicable RMI are as follows:

– *Hardware and software* It is important to provide detailed specifications of all the hardware and software used to implement the intervention, e.g., the robot, sensors (on-board or external to the robot), robot-control algorithm/interface, programs to activate the sensor and collect and/or process the sensor data. If the robot is custom-built, then the process of building that robot should be documented in such a way that a third-party can make a similar robot based only on that documentation. The programs for robot control should be open sourced or, at the very least, be accessible to registered users for free.

– *Settings* The physical settings where the intervention was conducted, the placement of the robot and sensors, relative positioning of the robot and participants, etc., should be precisely documented. For example, "the intervention was conducted in a $15m \times 15m$ room painted in white, the robot was placed on top of a $2m \times 2m$ white table, four cameras were installed on four walls at a height 6m from the ground, the participant was sitting on a chair placed $3m$ away in front of the robot, etc."

– *Actions* All actions required to be performed by the robot or any other person involved in the intervention should be documented with replicable precision. For example: "the robot therapist executed the interven-

tion in a 1 : 1 context, 5 min per day, 3 days per week, the robot therapist delivered a prompt 5 seconds after delivering the command if there was no response", etc.

Reproducibility also facilitates direct comparison of results with other research on the same topic.

3.4 Dependent Variables

Dependent variables are quantities of behavior which are assumed to be modified (increased, in case of skill learning and reduced in case of interfering behaviors) by the independent variable (i.e., the intervention). Accordingly, dependent variables are chosen based on their social significance and are naturally related to the target of an intervention. Dependent variables should be defined in such a way that they clearly show strong link to the outcome measure of the intervention. Well-defined dependent variables will help to clearly indicate the effectiveness of an intervention. It is a standard practice to use observable behaviors of the participants which can be measured using crisp metrics as dependent variables (e.g., the number of times an IwASD repeated a certain alphabet, in an alphabet learning task). The dependent variables are measured repeatedly before the intervention (baseline measurement), during the intervention, and after the intervention is removed in order to clearly understand the effect of the intervention on the participants. Accordingly, the process of measuring dependent variables should be documented with replicable precision.

3.5 Research Design

Both group-based designs and SS designs are widely used for ASD-interventions in clinical research. Group-based design [especially, randomized controlled trial (RCT)], although considered by some as the gold standard in clinical research, might not be the only choice to prove the effectiveness of an ASD intervention. SS research designs also have unique value in autism research [28,40]. Accordingly, robotics researchers can choose either of these two approaches for research design depending on the number and type of the participants and the nature of the intervention. Irrespective of research design, one important consideration for HRI studies is the number of intervention sessions. In clinical research, an intervention is inherently assumed to be a multi-session process. In the HRI domain, however, the majority of the studies on RMI are single-session studies where an IwASD interacts with a robot only once for a limited time (generally less than an hour). Changing a behavior through an intervention is a long process, especially for a complex population like IwASDs. It is unlikely that anyone can draw a meaningful conclusion about behavioral change from a single-session

intervention. Consider the 'likability' of a robot, a major focus of many HRI studies. Many of those studies included only a single session of observation. Thus the likability of the robot may be confounded with a 'novelty effect' on the IwASD during the first few sessions. Therefore, RMI should be arranged in multiple sessions over a reasonably long period of time.

Both SS and group-based designs have unique methodological characteristics and follow unique approaches to ensure proper experimental control. They are discussed below with respect to designing RMI.

3.5.1 Single-Subject Design

SS designs focus on the therapeutic improvement of each individual who serves as his/her own control during the study. Although an SS study might consist of only one subject, three to eight subjects significantly strengthen the external validity of the study [28]. Generally, the guidelines for designing a SS study on RMI are as follows.

– *Baseline measurement* Prior to the beginning of an intervention, it is critical to establish a baseline level of the dependent measure (the IwASD's level of behavior) being studied. It is this level of behavior that is being compared to the level achieved after the intervention has been taken place. During the baseline phase, the dependent variables are measured repeatedly at regular intervals until a stable/consistent pattern is achieved. The process of baseline measurement should be described with replicable precision.
– *Experimental control* Experimental control is critical to nullifying threats to internal validity by demonstrating a functional relation between the independent and dependent variables within the same participant. In a SS-design study, experimental control is achieved when the level of the dependent behavior changes only when the independent variable is introduced. For example, in a type of SS-design study known as an ABAB reversal design with a single individual, the level of the behavior might change (increase/decrease, depending on the nature of the study) only after treatment is introduced. Moreover, the subsequent removal of the treatment returns the level of the behavior to what it was prior to treatment. Introducing treatment a second time, along with an observed change in the level of the behavior, significantly improves confidence that it was the independent measure itself that was responsible for the change in the dependent measure. Importantly, whether before, during, or after treatment or the removal of treatment, measurement of behavior must consist of at least three data points. In addition to the reversal design, other SS designs include the multiple baseline and alternating treatments designs. In the case

of a RMI, any of these three approaches can be used to demonstrate experimental control.

– *Presentation of results* The lack of standards in HRI studies for presenting results may be one reason their findings are not convincing to the clinical community on the effectiveness of robots in ASD interventions. The demonstration of experimental control in SS-design studies is very dependent upon the visual presentation of the results. Thus, results from a SS-design study should be presented in such a way that it depicts a clear functional relationship between the independent and the dependent variables. In other words, the results should show that the use of a robot is clearly linked to the positive outcomes of a therapeutic intervention. Figure 1 illustrates a hypothetical set of results from a SS multiple-baseline-across-participants study. In this example, the data show that the behavior does not change until treatment is introduced. Such a design is often used when one does not expect the removal of treatment to result in pre-treatment levels of behavior. This is often the case with learning new skills, such as reciting the alphabet or counting numbers.

– *Reliability of observation* Most behavioral data from HRI studies is collected through the subjective observation of a human, whether done in real time or through the observation of video-recorded data (also known as behavioral coding). Thus, it becomes important to obtain a second and independent set of observed data. It is important to report the inter-observer agreement on the coded data. The calculation through Kappa statistics of the level of agreement between two independent observers is a popular way to report inter-observer agreement.

3.5.2 Group-Based Design

Group-based designs have high value in autism research. Such designs are often used when the experimental questions concern the effects of an intervention on relatively large populations of individuals. A group-based design involves at least two groups of participants. For example, if one is asking whether or not an intervention is effective, a control/comparison group and an intervention/experimental/treatment group are included. The intervention group receives the RMI while the control group does not receive any intervention or a different intervention. One might also compare the effects of an intervention on two different populations, such as IwASD versus TD children. Group-based designs nullify the threat to external validity through high number of participants, preferably $n > 10$ [52]. Important considerations for group based design of RMI are as follows:

– *Experimental control* Experimental control is established through comparing the results of the intervention group with the control group. The threat to internal validity is eliminated through a number of ways, e.g., random assignment of participants to intervention and control group (in case of RCT), group matching (in case of quasi-experimental design), etc.

– *Presentation of results* The results from a group-based design are evaluated based on their effect size and statistical significance. Accordingly, proper statistical analysis should be performed on the dependent variables to show a statistically significant difference between the control and the intervention group. Depending on the type of data, there are several standard statistical analysis methods (e.g., ANOVA, t-test, etc.) that can be used for this purpose.

The issue of reliability of observation is equally applicable to data collected using a group-based design.

3.6 Generalization Training

The purpose of generalization training in an ASD therapy has been explained in [68] as an effort "to spread the treatment effect across time, settings, stimuli, responses, or persons." In the case of RMI, the purpose of generalization training is to train an IwASD to execute a learned behavior or maintain a reduced level of an interfering behavior the same way with humans as (s)he executed/maintained while interacting with the robot. Generalization is an important component of a RMI as there exists a known concern that IwASDs may fail to generalize a skill or behavior learned through robots with other humans [17]. HRI studies on RMI should include a plan for generalization.

4 HRI Studies on ASD Interventions: How Well They Meet Clinical Standard

In light of the guidelines presented in Sect. 3, this section presents a comprehensive review to understand the status of the contemporary robotics research with respect to making RMI an EBP in autism. Table 1 presents a review of the literature published between 1990 and 2014 that meet the inclusion criteria of this review as outlined in Sect. 1.2. The guidelines for understanding Table 1 are as follows:

– *Robot* The second column of the table lists the name of the robot used in the RMI. If the robot is commercially available, the name of the manufacturer is also mentioned. For custom-built robots, a brief description of the type and functionality of the robot is provided.

– *Goal* This column provides a brief description of the clinical behaviors that the RMI planned to achieve.

Fig. 1 Illustration of a functional relation between the independent and dependent variables in a standard graphical representation of data from a single-subject multiple-baseline across-participants study
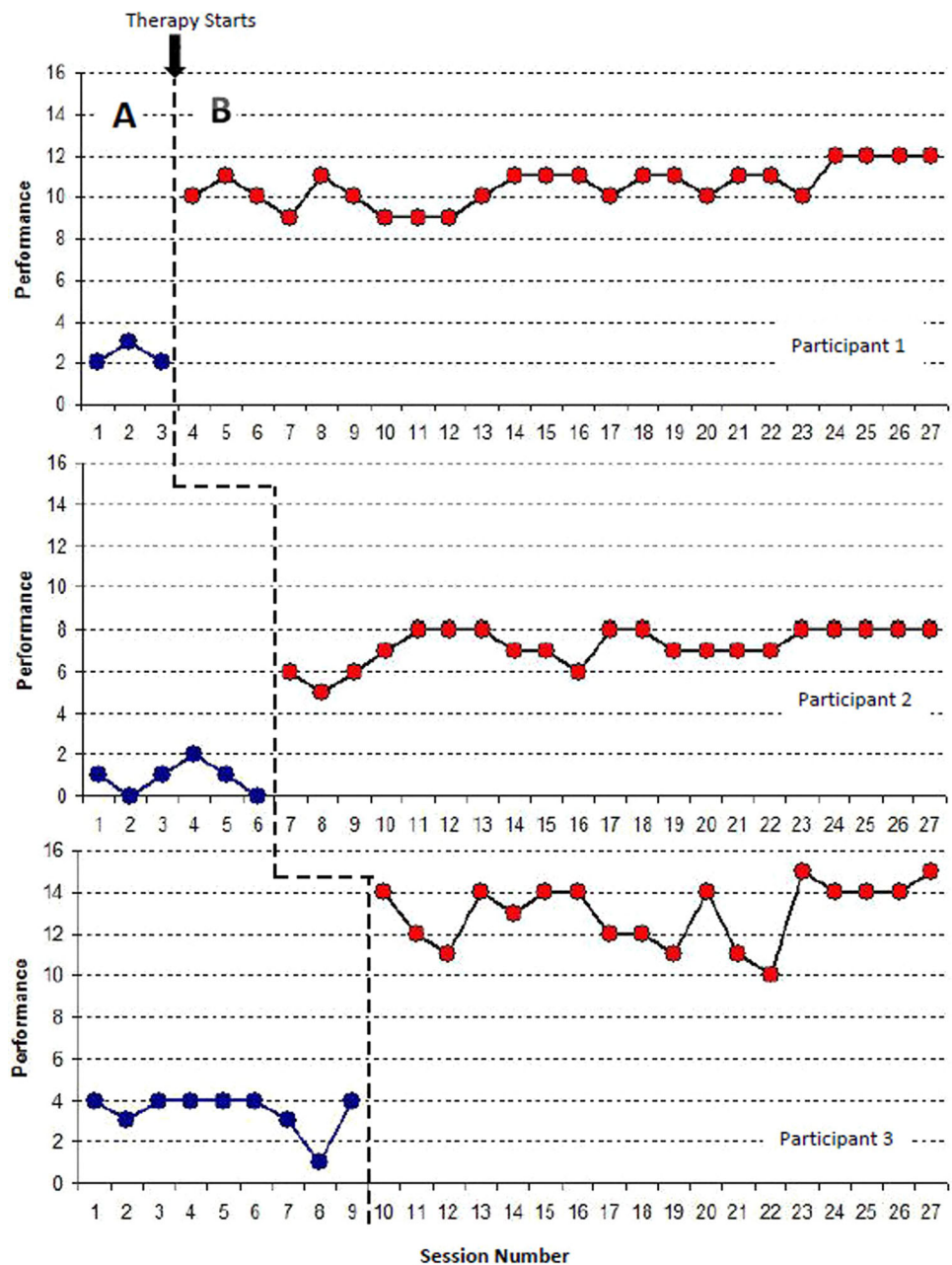
– *Participants* This column provides a brief description of the participants. First, the total number of participants is reported along with whether they were recruited based on any inclusion criteria. This is followed by the diagnosis and the standard tools used to make the diagnosis. Finally, the age range of the participants is reported.

– *Method* This column reports the independent variable of the study: design and reproducibility. If the study used any standard therapeutic approach, it is reported briefly. Otherwise, the therapy is reported as 'custom-designed'. With respect to reproducibility, a RMI is reported as 'sufficient for replication' if all three of the following conditions are met: (1) the robot used is commercially available, (2) the software and/or algorithms used to operate the robot are open-sourced or commercially available and the way to use them are properly documented, and (3) the physical settings used to conduct the RMI are described in detail. Otherwise, it is reported that the materials are not sufficiently described for replication.

– *Outcome measure* This column briefly describes the dependent variable(s) in the RMI.

– *Research design* This column briefly describes the type of research design, the number of human–robot session involved in the RMI, and the tools that were used to analyze the results.

– *Generalization training* This column briefly describes the generalization phase, if any.

– *Findings* This column summarizes the major findings of the RMI.

5 Discussion

Table 1 shows the 22 articles published before September 2014 that met the inclusion criteria for this review. Almost all of the articles are related to, at-least, one other published article on the same study. In some cases, there were several other publications on the same or a slightly different study with a different analysis. This is particularly true for the studies that use custom-built robots (e.g., KASPAR, Probo, Robota, FACE, etc.). Articles were published in a variety of venues, not merely in conference proceedings or journals whose focus is robotics and HRI. Most importantly, some of these articles originated from non-robotics research laboratories [2,23,46]. These are indications that the potential of robots as a tool for ASD intervention is attracting non-robotics researchers.

One important trend to note in Table 1 is that a majority of the articles (19 out of 22) that met our inclusion criteria were published after 2010. A vast majority of HRI studies on RMI published prior to 2010 were not included in this review primarily because their focus was to investigate either the 'likability' of a robot or the features/characteristics of a robot that trigger interest in IwASDs. An increasing number of publications in recent years is a good sign that more and more robotics research is now focused on proving the effectiveness of a robot in ASD intervention, which will eventually may help RMI to be an EBP in autism.

Overall, the studies also show much promise for RMI. Although none of the studies reviewed here claimed that their RMI was definitively able to teach/modify a behavior, many studies reported significant improvement (often, through statistical validation) of the participants' target behavior during or after the RMI [2,24,34,46,50,66,67]. It should be noted, however, that the quality of this evidence could be significantly increased through rigorous research design so that it reaches the status where RMI can be considered an EBP in autism.

The rest of this section will provide an overall discussion of the research reported in Table 1 based on the design guidelines presented in Sect. 3.

5.1 Goal of Intervention

HRI researchers historically have focused on skills/behaviors related to *social* and *communication* deficits (specific behaviors under these two categories are discussed in Sect. 3.1). For example, a vast majority of RMIs focused on training imitation and turn-taking behaviors [19,25,30,49,55,64,65], group play [24,67], and joint attention [5]. There are some

recent studies that focused on training a number of important *social* and *communication* life-skills and behaviors such as improving an IwASD's touch-sensitivity [54], teaching how to write a text-message [46] and ask a question [29].

As discussed in Sect. 3.1, there are other target areas for ASD intervention that have extremely important social significance and could be suitable for RMI. For example, robots could play a significant role in teaching/improving behaviors in the academic category. Similarly, the ability of a robot to precisely repeat actions and behaviors can serve a significant role in teaching verbal behaviors, such as how to initiate and/or continue a conversation in a socially acceptable manner in different contexts of life (e.g., with peers, with colleagues at the work place, etc.). A recent study has also presented some preliminary results on the use of RMI for improving cognitive flexibility in IwASD [14].

Finally, similar to the role of socially assistive robots as a companion of the elderly and people with disabilities [8], robots may have much potential to help an IwASD to eliminate/reduce behaviors that interfere with the normal functioning of his/her life, e.g., depression, anxiety, etc. No RMI focuses on these target behaviors.

5.2 Participants

Recruiting participants based on well-defined inclusion criteria is not a common practice in HRI research. A majority of the studies reported in Table 1 recruited participants without investigating how well they served the goal of the study. This makes it difficult to draw any clear conclusion about the effect of the RMI on the participants. There were only a few RMIs published in recent years that recruited participants based on well-defined inclusion criteria [29,34,49,50,65,66]. In addition, the practice of reporting participants' diagnostic information also was uncommon in HRI research on RMI. A vast majority of HRI studies reported their participants merely as "children/person diagnosed with ASD". Fortunately, many more recent studies reported detailed information about participants' demographics and diagnostic conditions using standard tools such as DSM IV, ADOS, SCQ, CARS, ADI-R, etc. [2,5,13,19,23,29,33,34,46,48–50,63–66].

5.3 Independent Variable

Contemporary HRI research appears to have a major weakness with respect to designing and reporting the independent variable (i.e., the RMI itself). A vast majority of the HRI studies relies on custom-designed therapies which may or may not have been designed in close collaboration with a domain-expert, thereby highly increasing the probability of producing study data that are questionable/less-appealing to the clinical community. A few recent studies, however, were designed

Table 1 Review of HRI studies on robot-mediated ASD interventions

References, years	Robots	Goals	Participants			Methods	Outcome measures	Research design	Generalization	Findings
			#	Diagnosis	Age					
[46] 2014	NAO from Aldebaran Robotics, Inc.	To teach writing text messages that include at least a greeting, a personal narrative and a closing	3	Intellectual disabilities or ASD of participants are confirmed by WISC–III and Stanford–Binet scale	19–21	(1) Custom-designed interaction: the robot provides response prompts	Number of occurrences of a greeting, a personal narrative, and a closing in the text messages written by participants	Concurrent multiple baseline design across participants	Efforts were made to generalize the skill with others without the help from the robot	Two participants maintained the learned skill after the intervention stopped but the contribution of the robot to the success is not clear from the study due to other confounding variables
			No inclusion criteria was reported	IQ: 40–63		(2) Description of materials is not sufficient for replication		Multiple sessions Descriptive statistics is used to analyze the results		
[54] 2014	KASPAR: a custom-built humanoid robot with touch sensitive skin	Improving the tolerance to touch and touch-based interactions	24	Previously diagnosed with autism, moderate–severe learning disabilities	4–15	(1) Custom-designed game-based interaction	Playability of the robot and the motivation of the participants to play	Single session	None	Some participants seems to understand the cause–effect relationships while interacting with the robot and exhibit social communication skills with the robot
		Improving the understanding of cause–effect relations	No inclusion criteria was reported			(2) Description of the play scenario is sufficient for replication	Number of tactile interactions with robot	Descriptive analysis of results		

Table 1 continued

References, years	Robots	Goals	Participants #	Diagnosis	Age	Methods	Outcome measures	Research design	Generalization	Findings
[2] 2014	Auti: a custom built pet-like toy robot with furry skin	Improving social skills	18	Autism diagnosis confirmed by GARS-2	4–9	(1) ABA-based intervention	Number of verbal and physical interactions with the robot (with and without prompts)	Group-based design	None	Statistically significant increase in social behaviors in the intervention group
		Reducing interfering behaviors	No inclusion criteria reported			(2) Description of materials is not sufficient for replication		Single session; Generalize linear model and Wald statistics are used to analyze the results		No change in interfering behaviors
[25] 2014	NAO from Aldebaran Robotics, Inc.	Improving imitation skill through robot-mediated prompts	12	Previously diagnosed with autism, no other diagnostic confirmation was provided	7–10	(1) Custom-designed game-based interaction	Success in imitating the robot and the amount or level of prompts required for imitation	Group-based design	None	Adaptive prompting is less-frustrating for the participants than the non-adaptive prompts
			No inclusion criteria reported			(2) Description of materials is not sufficient for replication		Multiple sessions; Descriptive statistics is used to report the findings		

Table 1 continued

References, years	Robots	Goals	Participants			Methods	Outcome measures	Research design	Generalization	Findings
			#	Diagnosis	Age					
[67] 2014	KASPAR: a custom built humanoid robot	Improving social skills through collaborative games played with a robot	6	Previously diagnosed with autism, no other diagnostic confirmation was reported	8–9	(1) Custom-designed game-based interaction	Gaze pattern, words exchanged, and affective responses	Single subject, reversal (A–B–A) design	Outcome measures were evaluated after the intervention was stopped	Significant improvement in gaze pattern and frequency of positive affects
			No inclusion criteria was reported			(2) Description of materials is not sufficient for replication		Multiple sessions		No significant improvement in verbal communication
								Wilcoxon's matched pair signed rank test was used for the data analysis		
[5] 2014	NAO from Aldebaran Robotics, Inc.	Comparing the effectiveness of robot- and human-mediated interventions to improve joint attention skill	12, 6 with ASD and 6 TD	Autism diagnosis confirmed by DSM IV-TR, ADOS, and SCQ	2–5	(1) Custom-designed interaction	Gaze at the robot, prompt level looking at the target, and hit-frequency	Group-base quasi-experimental design	None	Results do not show any statistical significance of the robot over the human with respect to preference in joint attention
			No inclusion criteria reported			(2) Description of materials is not sufficient for replication		Single session		All participants generally performed better in human-mediated sessions as compared to the robot-mediated sessions
								Descriptive statistics is used for analysis of results		

Table 1 continued

References, years	Robots	Goals	Participants #	Diagnosis	Age	Methods	Outcome measures	Research design	Generalization	Findings
[50] 2013	Probo: a custom-built elephant like robot	Comparing the effect of a robot and a human therapist to improve social skills (e.g., greetings goodbye, sharing)	4, Selected based on inclusion criteria	Autism diagnosis confirmed by ADOS-G, ADI-R, and DSM IV	4–9	(1) Social story (Ss) intervention	The level of prompt needed to generate the expected social responses	Single subject, reversal design (ABAC/ACAB)	Outcome measures were evaluated after the intervention stopped	Statistically significant improvement in executing social skills with fewer prompts after the robot-mediated intervention as compared to the human-mediated intervention
						(2) Description of materials is not sufficient for replication		Multiple sessions		
								Mann–Whitney test and visual analysis are used for result analysis		
[29] 2013	NAO from Aldebaran Robotics, Inc.	Comparing the effectiveness of a robot-mediated therapy with a human-mediated therapy to promote the self-initiated question asking ability	6, Selected based on inclusion criteria	Autism diagnosis confirmed by DSM IV-TR and SCQ	8–14	(1) ABA-based intervention	Number of self-initiated questions	Combined crossover multiple baseline SS design	Four follow-up sessions to monitor the progress	Both robot- and human-mediated therapies were effective, no conclusion can be drawn on the differential effectiveness of these two therapies
				IQ >80		(2) Description of materials is not sufficient for replication provided but they are not sufficient for replication		Multiple sessions		
								Visual analysis of results		

Table 1 continued

References, years	Robots	Goals	Participants #	Diagnosis	Age	Methods	Outcome measures	Research design	Generalization	Findings
[34] 2013	Pleo: a pet-like toy robot	Comparing the effect of a robot with a human and a computer game as a reinforcer in verbal communication	24, Selected based on inclusion criteria	Autism diagnosis confirmed by ADOS-M3 and DAS-II; IQ > 70	9–15	(1) Custom-designed game-based interaction; (2) Description of materials is not sufficient for replication	Number of utterances	Randomized controlled crossover design; Single session; ANOVA and t-test are used for analysis	None	Results show statistical significance of the robot as a reinforcer in verbal communication
[23] 2013	Two custom-built toy robots: GIPY 1 and POL, both small and light-weight	Using robots as a neural orthesis to improve non-verbal cognitive processes	Case I: 4; Case II: 11; No inclusion criteria reported	Autism diagnosis confirmed by DSM IV-TR and CARS; Severe autism	Case I: 7–9; Case II: 7–8	(1) Custom-designed game-based interaction; (2) Description of materials is not sufficient for replication	Game-play skills: eye-contact touching, posture, manipulation, and expressing positive emotion	Group-based design; Single session; Chi-square test and Spearman's rank correlation coefficient are used for analysis	Transition from a dyadic to a triadic interaction was planned to help generalization	Children engaged in free-game play with the robot; Children used the robot as an 'instrument' to communicate with adults
[64] 2013	i-Sobot from Isobot robot	Improving imitation and praxis performance	16, 15 TD and 1 with ASD; No inclusion criteria reported	Autism diagnosis confirmed by ADOS-G	4–7	(1) Custom-designed interaction; (2) Description of materials is not sufficient for replication	Sensory integration and praxis test (SIPT) was used to evaluate the outcome	Quasi experimental pre–post-test design; Multiple sessions; ANOVA and Wilcoxon's signed rank test are used for analysis	Outcome measures were evaluated in a post-test session	The participant with ASD showed improved performance in SIPT score in the post-test session

Table 1 continued

References, years	Robots	Goals	Participants			Methods	Outcome measures	Research design	Generalization	Findings
			#	Diagnosis	Age					
[49] 2013	Robonova from Hitech Robotics	Improving imitative responses and social behaviors	2, Selected based on inclusion criteria	Autism diagnosis confirmed by CARS	5–6	(1) Custom-designed interaction	Frequency of imitation, attention, and physical interaction	Single subject reversal design	Outcome measures were evaluated after the intervention stopped	No improvement in imitation
						(2) Description of materials is not sufficient for replication		Multiple sessions		Some improvements in physical interaction and attention
								Visual analysis of results		
[66] 2012	Probo: a custom-built elephant-like robot	Comparing the effect of a robot and a personal computer to improve four social skills: asking questions, eye gaze, asking for help, and greetings	20, Selected based on inclusion criteria	Autism diagnosis confirmed by ADOS-G and DSM IV-TR	4–9	(1) Social story (Ss) intervention	The level of prompt needed to generate the expected social responses	RCT with three groups: control, intervention 1 (Ss-PC), intervention 2 (Ss-robot)	None	Statistically significant improvement in executing prompt-independent social skills in the intervention 2 group as compared to the control group
						(2) Description of materials is not sufficient for replication		Multiple sessions		
								Mann–Whitney test is used for analysis		

Table 1 continued

References, years	Robots	Goals	Participants #	Diagnosis	Age	Methods	Outcome measures	Research design	Generalization	Findings
[30] 2012	Humanoid robot by Yujin Robot, Co., Ltd.	Improving attention, social skills, and communication through the use of robots	5, 3 with ASD and 2 with medical conditions other than ASD	Previously diagnosed with ASD, no other diagnostic confirmation was reported	19–21	(1) Custom-designed game-based interaction	11 Useful and 5 non-useful behaviors are defined (e.g., talk to, look at, joint attention, etc.)	Single session	None	Varied individual responses to the robot. No firm conclusion can be drawn
			No inclusion criteria reported			(2) Poor documentation of materials, not sufficient for replication		Descriptive presentation of results		
[33] 2012	Pleo: a pet-like toy robot	To find the effect of a robot to improve communication skills (e.g., speech, orient-to-face, etc.)	18 with ASD and 11 TD	Autism diagnosis confirmed by ADOS-M3 and SCQ		(1) Custom-designed game-based interaction	Affective responses, time spent speaking and orienting to the face of the human after the intervention	Group-based quasi-experimental design	Outcome measures were evaluated after the human–robot session	Significant increase in time spent speaking
			No inclusion criteria reported	IQ > 0		(2) Description of materials is not sufficient for replication		Single session		No significant increase in the time of face-to-face orientation
								Paired t-test is used to analyze results		

Table 1 continued

References, years	Robots	Goals	Participants #	Diagnosis	Age	Methods	Outcome measures	Research design	Generalization	Findings
[65] 2012	NAO from Aldebaran Robotics, Inc.	Comparing the effect of a robot and a human to improve social skills (e.g., turn-taking, initiation) during a motor imitation game	4, Selected based on inclusion criteria	Autism diagnosis confirmed by DSM IV	2–6	(1) Custom-designed imitation-based interaction (2) Description of materials is not sufficient for replication	Number of initiation with and without prompts, duration of eye gaze and smile/laughter, and the number of gaze shift	Single subject A–B–A–C design Multiple sessions Mann–Whitney test and visual analysis are used for the analysis of results	Outcome measures were evaluated after the intervention stopped	The robot does not have more significant effect in improving the target behaviors than the human
[24] 2012	Troy: a custom-built humanoid robot	Improving turn-taking skill during social communications	2 No inclusion criteria reported	Previously diagnosed with ASD, no other diagnostic confirmation was reported	3, 8	(1) Custom designed turn-taking interaction (2) Description of materials is not sufficient for replication	Number of engagement initiated Number of engagement responded to	Single subject, pre–post-test design Multiple session Pre–post-test results were compared numerically	Transition from a triadic to a dyadic interaction is planned to help generalization	Numerically noticeable improvement in the target skill

Table 1 continued

References, years	Robots	Goals	Participants #	Diagnosis	Age	Methods	Outcome measures	Research design	Generalization	Findings
[63] 2012	NAO from Aldebaran Robotics, Inc.	Reducing interfering behaviors	1	Autism diagnosis confirmed by ADOS and SB5	10	(1) Custom-designed interaction	Stereotyped behaviors, communications, and social interactions were scored according to the GARS-2 scale	Single session	None	Improvements in interaction and communication
			No inclusion criteria reported	IQ = 107		(2) Description of materials is sufficient for replication		Results are analyzed using descriptive statistics		Lesser number of stereotyped behaviors
[13] 2011	Lego Mind storm NXT	Improving social communication through the use of robots	1	Autism diagnosis confirmed by CARS	11	(1) Custom-designed interaction based on an assessment of the participant	Affective responses to the robot	Multiple sessions	Multiple sessions are designed to generalize the learned skill with different persons and in different settings	Positive attitudes toward the robot
			No inclusion criteria was reported	Non-verbal		(2) Description of materials is not sufficient for replication	Frequency of interactions	Descriptive statistics is used to analyze the results from the only one participant		Increase in performance
										The participant was able to generalize the learned skill to new people

Table 1 continued

References, years	Robots	Goals	Participants			Methods	Outcome measures	Research design	Generalization	Findings
			#	Diagnosis	Age					
[19] 2008	Tito: a custom-built humanoid robot (like soft-toy)	Improving social skills (imitation) through a robot	4	Autism diagnosis confirmed by ADOS-G	4–5	(1) Custom-designed interaction	Eye-gaze, facial expression, action and words, and gesture directed toward the robot	Single subject reversal design	Outcome measures were evaluated after the intervention stopped	Statistically some participants executed fewer repetitive behavior with the robot
			No inclusion criteria reported			(2) Description of materials is not sufficient for replication	Imitation of the robot's behavior	Multiple sessions		Overall, the imitation performance is better with the human therapist than the robot therapist
								Regression was used to analyze the results		
[48] 2007	FACE: a custom-built android face	Improving social communication through understanding facial emotions	4	IQ 52–105	7–20	(1) Custom-designed interaction	Responses of the participants were evaluated using CARS scale	Single session	None	Participants showed some improvements according to the CARS scale
			No inclusion criteria reported	Autism diagnosis confirmed by ADOS-G and ADI-R		(2) Description of materials is not enough for replication		Results are analyzed using descriptive statistics		

Table 1 continued

References, Robots years	Goals	Participants #	Diagnosis	Age	Methods	Outcome measures	Research design	Generalization Outcome measures	Findings
[55] 2005 Robota: a custom-built humanoid robot doll	Improving imitation and social interaction skill	4	Previously diagnosed with autism, no other diagnostic confirmation was reported	5–10	(1) Custom designed interaction	Eye-gaze, touch, imitation, and proxemics	Multiple sessions	were evaluated after the intervention stoppedSome participants showed improved performance in interacting with humans	
		No inclusion criteria reported			(2) Description of materials is not sufficient for replication		Qualitative analysis of results		

based on established approaches to behavioral intervention (e.g., ABA, social story, etc.) [2,29,50,66]. For example, the RMI described in [2] used a toy robot (Auti) to promote physical and verbal interaction abilities in a group of participants with ASD using the principle of ABA. Playful movements of the robot were considered a reward for the participants (reward is a core component in an ABA-based intervention). The reward was offered to reinforce positive behaviors such as gentle speaking and touching. Challenging behaviors, such as screaming or hitting, were discouraged through removal of the reinforcing movements of the robot.

Reproducibility is another aspect that is almost completely ignored in HRI research on RMI. A majority of the articles did not document the intervention materials (robot, sensors, source-code, physical settings, etc.) in such a way that other researchers would be able to re-create the same intervention with the same robot or a robot of similar kind based only on that documentation. According to our review, there was no report of implementing a single RMI in two different sites by two different groups of robotics researchers. The research reported in [54], however, presented a custom-designed human–robot play scenario and documented it with replicable precision (the physical settings, scoring process, actions of both the robot and the human were discussed in detail). The RMI in [54], however, used a custom-built humanoid robot (KASPAR) which is not easy to replicate by other HRI researchers.

5.4 Dependent Variables

The commonly used dependent variables in contemporary HRI research on RMI are *gaze* (the duration or the number of times an IwASD looked at the robot during a RMI session) [4,33,35,60,65,67], *communication* (number of verbal/non-verbal communication with the robot, total number of words exchanged with the robot, etc.) [34,60,67], *affect* (being in an affective state or showing affective responses to the robot) [15,20,33,56], *attention* (focusing on the robot) [33,35], *imitation* (imitating a robot's action or speech) [22,35,56], and *proxemics* (being in a close proximity of the robot) [21]. Although these variables work well to assess the general enthusiasm expressed by an IwASD when (s)he is around a robot, they generally do not hold any direct social significance and often do not provide enough information to gauge the effectiveness of a RMI [3]. For example, how affectionate an IwASD is toward a robot during an intervention does not contribute anything to the core purposes of an ASD intervention as outlined in Sect. 3 (i.e., improvement in independent living, health and well being, and the quality of life). Such variables, however, might still hold social significance if they are placed within the context of achieving a broader, socially important goal. For example, how long a participant stares at a robot during an HRI study (commonly known as 'gaze at the

robot' behavior) might not have any direct importance with respect to improving the quality of life/health/well-being of an IwASD, but this 'gaze at the robot' behavior could be a meaningful dependent variable in an intervention that aims to teach an IwASD how to maintain eye-contact during a social conversation. Such an intervention might start with measuring the 'gaze at robot' behavior of the participants while having a conversation with the robot. When the participant masters the skill of maintaining eye-contact with the robot, the robot gradually could be replaced with other humans (a part of the generalization training).

There is a growing effort in HRI research to carefully operationalize meaningful dependent variables. Some recent studies are defining their dependent variables in such a way that the contribution of the robot in achieving the goal of the intervention was more evident. For example, the RMI discussed in [24] was designed to promote turn-taking behaviors in children with ASD. It used the frequency of self-initiated engagement by a participant as a dependent variable and measured it before and after the intervention. This simple, easy to measure variable provides strong indication about the effectiveness of the RMI. Similarly, the RMI reported in [2] to promote physical and verbal interactions used the number of self-initiated and prompt-dependent physical interactions of IwASDs as dependent variables and measured them throughout the study. The studies reported in [29,50,64,65] also defined dependent variables that were linked to the intended outcome of the study, and thus they hold important information about therapeutic outcomes.

5.5 Research Design

HRI research on RMI is very weak in this domain. Many studies did not incorporate any standard research design. In general, group-based design is more common in HRI research on RMI than the SS design. The number of participants, however, is generally less than 10 ($n < 10$) in a majority of the studies that used a group-based design. Only a few studies with group-based designs made an effort to incorporated a control group and match participants between groups in order to ensure experimental control [5,33,50,66]. For example, the RMI reported in [33] to investigate the engagement behavior of children with ASD used an age-matched control group of 11 participants with an intervention group consisted of 18 participants. The RMI to train join attention, reported in [5], also used an aged-matched control group of six TD children with an intervention group of six children with ASD and showed that the intervention group required significantly more prompt than the control group to accurately direct attention in the experimental settings.

In the case of SS design, there were only few studies which ensured proper experimental control through clear demonstration of treatment effect [19,29,65,67]. For example, the

RMI reported in [65] used a SS ABAC design to compare the effect of a robot with a human in encouraging children with ASD to engage in a motor imitation game. SS reversal design (ABA) was also used to implement the RMI reported in [19,67]. The RMI reported in [29] employed combined crossover multiple baseline design to investigate the role of robots in promoting question-asking behaviors among children with ASD.

Lack of data-analysis standards also was observed in the literature on RMI. A majority of the studies used descriptive statistics or analysis to report their findings. Only a few studies used proper statistical analysis (in case of group-based design) or visual analysis (in case of SS design) while analyzing the results [2,29,34,50,64–67]. Commonly used statistical methods for data analysis (in the cases of group-based design) were ANOVA, t-test, Mann–Whitney test, Chi-square test, Wilcoxons signed rank test, and regression.

A vast majority of the studies reported inter-observer agreement through *Kappa* statistics.

Irrespective of research design, a common issue was that most of the studies consisted of only a single session. Consequently, no matter how sound the research design was or how well the variables were defined/documented, it is extremely difficult to draw any firm conclusion about the impact of an RMI on the participants. Reporting results from a single session RMI can call into question the validity of the consensus (among contemporary HRI literature) that robots are liked by many children with ASD as the 'likability' can be positively biased by the 'novelty' effect.

5.6 Generalization Training

Generalization training is not common in RMIs. Although many studies measured the dependent variables after the RMI was stopped in order to monitor the progress of the participants (e.g., [19,24,29,33,49,55,64,65,67]), they did not include any explicit plan to train the participating IwASDs to practice the target behavior(s) with other humans and in different contexts. Only three RMI reported in Table 1 included a separate generalization phase where the IwASDs were systematically trained to practice a behavior with other humans [13,23,46]. For example, the RMI reported in [23] planned a triadic interaction among an IwASD, a robot, and another human so that the participant can practice a set of cognitive non-verbal behaviors (namely, eye contact, touch, manipulation, and posture) with other humans. Triadic interaction among an IwASD, a robot, and another human was also used in [13] to help the participants generalize the learned skill of social communication with other humans in different environmental settings.

The simplest way to implement generalization training during a RMI is fading the role of the robot gradually. For example [3,24] described RMI where IwASDs were engaged in a triadic interaction with a robot and a human. The robot was gradually removed from the interaction, making it a dyadic interaction between the human and the IwASD. The exact form of generalization training, however, depends on the goal of an intervention and the type of research design.

6 Conclusion

Despite a decade of research, the effectiveness of RMIs to teach new life-skills or eliminate non-functional behaviors is not yet fully understood. This may be due in part to the lack of methodological rigor in robotics research on RMI. For the robotics community, observing appropriate methodological rigor while designing HRI studies on RMI requires awareness of the research standard commonly followed in clinical research on ASD intervention, as well as a detailed sense of how the contemporary robotics research is conforming to such standards. This article suggests that a promising way to prove the effectiveness of RMIs is to establish it as an EBP in autism. Accordingly the article reports a set of guidelines generally observed in clinical research in order to consider an experimental intervention as an EBP in autism and discusses the ways research on robot mediated ASD intervention can follow these guidelines. A review of the contemporary HRI studies on RMI based on these guidelines was then presented. The review has clearly shown a methodological shift in robotics research on RMI, where recent research (e.g., the research published after 2010) is more likely to comply with the clinical standards in research design while assessing the effectiveness of robots in ASD interventions. However, the number of studies that strictly adhered to all guidelines to produce high-quality evidence in favor of the effectiveness of RMIs is still too low. We hope that this article will inspire and help robotics researchers to conduct studies on RMI that meet clinical standards and thereby produce data that will enable RMIs to be considered an EBP in autism.

References

1. American Psychological Association (2001) Policy statement on evidence-based practice in psychology. http://www.apa.org/practice/ebpstatement.pdf. Accessed 3 Feb 2015
2. Andreae H, Andreae P, Low J, Brown D (2014) A study of auti: a socially assistive robotic toy. In: Proceedings of the conference on interaction design and children. ACM, New York, p 245–248
3. Begum M, Serna R, Kontak D, Allspaw J, Kuczynski J, Suarez J, Yanco H (2015) Measuring the efficacy of robots in autism therapy: How informative are standard HRI metrics? In: ACM/IEEE international conference on human–robot interaction

4. Bekele ET, Lahiri U, Swanson AR, Crittendon JA, Warren ZE, Sarkar N (2013) A step towards developing adaptive robot-mediated intervention architecture (ARIA) for children with autism. Neural Syst Rehabil Eng 21:289–299

5. Bekele E, Crittendon JA, Swanson A, Sarkar N, Warren ZE (2014) Pilot clinical application of an adaptive robotic system for young children with autism. Autism 18:598–608

6. Bird G, Leighton J, Press C, Heyes C (2007) Intact automatic imitation of human and robot actions in autism spectrum disorders. Proc Biol Sci 274(1628):3027–3031

7. Boccanfuso L, O'Kane JM (2010) Adaptive robot design with hand and face tracking for use in autism therapy. In: International conference on social robotics, p 265–274

8. Broekens J, Heerink M, Rosendal H (2009) Assistive social robots in elderly care: a review. Gerontechnology 8(2):94–103

9. Cabibihan JJ, Javed H, Ang M Jr, Aljunied SM (2013) Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. Int J Soc Robot 5(4):593–618

10. Chaminade T, Fonseca DD, Rosset D, Lutscher E, Cheng G, Deruelle C (2012) fMRI study of young adults with autism interacting with a humanoid robot. In: International symposium on robot and human interactive communications, p 380–385

11. Constantino JN, Gruber CP (2005) Social responsiveness scale manual. Western Psychological Services, Los Angeles. http://www.wpspublish.com/app/. Accessed 3 Feb 2015

12. Constantino JN, Davis SA, Todd RD, Schindler MK, Gross MM (2003) Validation of a brief quantitative measure of autistic traits: comparison of the social responsiveness scale with the autism diagnostic interview-revised. J Autism Dev Disord 33(4):427–433

13. Costa S, Soares F, Santos C, Ferreira MJ, Moreira F, Pereira AP, Cunha F (2011) An approach to promote social and communication behaviors in children with autism spectrum disorders: robot based intervention. In: Proceedings of the 2004 IEEE International symposium on robot and human interactive communication. IEEE, Atlanta, p 101–106

14. Costescu CA, Vanderborght B, David DO (2015) Reversal learning task in children with autism spectrum disorder: a robot-based approach. J Autism Dev Disord 45(11):3715–3725

15. Dautehhahn K, Werry I (2007) Towards interactive robotics in autism therapy. Pragmat Cogn 12:1–35

16. David D, Matu S, David OA (2014) Robot-based psychotherapy: concepts development, state of the art, and new directions. Int J Cogn Ther 7(2):192–210

17. Diehl J, Schmitt L, Crowell CR, Villano M (2012) The clinical use of robots for children with autism spectrum disorders: a critical review. Res Autism Spectr Disord 6(1):249–262

18. Diehl J, Crowell CR, Villano M, Wier K, Tang K, Riek L (2014) Clinical applications of robots in autism spectrum disorder diagnosis and treatment. In: A comprehensive guide to autism. Springer, New York

19. Duquette A, Michaud F, Mercier H (2008) Exploring the use of a mobile robot as an imitation agent with children with low-functioning autism. Auton Robots 24(1):147–157

20. Feil-Seifer J, Mataric MJ (2008) B3IA: a control architecture for autonomous robot-assisted behavior intervention for children with autism spectrum disorders. In: International conference on human–robot interaction, p 328–333

21. Feil-Seifer J, Mataric MJ (2010) Using proxemics to evaluate human–robot interaction. In: International conference on human–robot interaction, p 143–144

22. Fujimoto I, Matsumoto T, Silva PRSD, Kobayashi M, Higashi M (2011) Mimicking and evaluating human motion to improve the imitation skill of children with autism through a robot. Int J Soc Robot 3:349–357

23. Giannopulu I (2013) Multimodal cognitive nonverbal and verbal interactions: the neurorehabilitation of autistic children via mobile toy robots. Int J Adv Life Sci 5(3, 4):214–222

24. Goodrich MA, Colton M, Brinton B, Fujiki M, Atherton JA, Ricks D, Maxfield MH, Acerson A (2012) Incorporating a robot into an autism therapy team. IEEE Intell Syst Mag 27(2):52–59

25. Greczek J, Kaszubski E, Atrash A, Matari M (2014) Graded cueing feedback in robot-mediated imitation practice for children with autism spectrum disorders. In: Proceedings of IEEE international symposium on robot and human interactive communication. IEEE, p 561–566

26. Grynszpan O, Weiss PL, Perez-Diaz F, Gal E (2014) Innovative technology-based interventions for autism spectrum disorders: a meta-analysis. Autism 18:346–361

27. Harris SL, Delmolino L (2002) Applied behavior analysis: its application in the treatment of autism and related disorders in young children. Infants Young Child 14:11–18

28. Horner RH, Carr EG, Halle J, McGee G, Odom S, Wolery M (2005) The use of single-subject research to identify evidence-based practice in special education. Except Child 71:165–179

29. Huskens B, Verschuur R, Gillesen J, Didden R, Barakova E (2013) Promoting question-asking in school-aged children with autism spectrum disorders: effectiveness of a robot intervention compared to a human–trainer intervention. Dev Neurorehabil 16:345–356

30. Jordan K, King M, Hellersteth S, Wiren A, Mulligan H (2012) Robotic technology for teaching adolescents with autism spectrum disorder: a feasibility study. In: Proceedings of the 6th international conference on rehabilitation engineering and assistive technology. ACM, Singapore, p 24:1–24:4

31. Kazdin AE (2008) Evidence-based treatment and practice: new opportunities to bridge clinical research and practice, enhance the knowledge base, and improve patient care. Am Psychol 63:146–159

32. Kelley JF (1984) An iterative design methodology for user-friendly natural language office information applications. ACM Trans Inf Syst 2(1):26–41

33. Kim ES, Scassellati B (2012) Bridging the research gap: making HRI useful to individuals with autism. Hum Robot Interact 1(1):26–54

34. Kim ES, Berkovits LD, Bernier EP, Leyzberg D, Shic F, Paul R, Scassellati B (2013) Social robots as embedded reinforcers of social behavior in children with autism. Autism Dev Disord 43:1038–1049

35. Kozima H, Nakagawa C, Yasuda Y (2007) Children–robot interaction: a pilot study in autism therapy. Prog Brain Res 164:385–400

36. Lee J, Takehashi H, Nagai C, Obinata G, Stefanov D (2012) Which robot features can stimulate better responses from children with autism in robot-assisted therapy? Adv Robot Syst 9:1–6

37. Lord C, Rutter M, Goode S, Heemsbergen J, Jordan H, Mawhood L, Schopler E (1989) Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. J Autism Dev Disord 19:185–212

38. Lord C, Rutter M, DiLavore P, Risi S (2001) Autism diagnostic observation schedule (ADOS) manual. Western Psychological Services, Los Angeles. http://www.wpspublish.com/app/. Accessed 3 Feb 2015

39. MAG Initiative (2012) Autism spectrum disorders: guide to evidence based interventions. http://autismguidelines.dmh.mo.gov/documents/Interventions.pdf. Accessed 3 Feb 2015

40. Mesibov G, Shea V (2011) Evidence-based practices and autism. Autism 15:114–133

41. Mesibov G, Shea V, Schopler E (2005) The TEACCH approach to autism spectrum disorders. Kluwer Academic/Plenum Publishers, New York

42. Nathan PE, Gorman JM (2007) A guide to treatments that work, 3rd edn. Oxford University Press, New York

43. Odom SL (2013) Technology-aided instruction and intervention (TAII) fact sheet. The University of North Carolina, Frank Porter Graham Child Development Institute, Autism Evidence-Based Practice Review Group, Chapel Hill. http://autismpdc.fpg.unc.edu/sites/autismpdc.fpg.unc.edu/files/2014-EBP-Report.pdf. Accessed 3 Feb 2015

44. Odom SL, Brantlinger E, Gersten R, Horner RD, Thompson B, Harris K (2004) Quality indicators for research in special education and guidelines for evidence-based practices: executive summary. Council for Exceptional Children Division for Research, Arlington

45. Odom S, Collet-Klingenberg L, Rogers S, Hatton D (2010) Evidence-based practices in interventions for children and youth with autism spectrum disorders. Prev School Fail 54:275–282

46. Pennington R, Saadatzi MN, Welch KC, Scott R (2014) Using robot-assisted instruction to teach students with intellectual disabilities to use personal narrative in text messages. J Spec Educ Technol 29(4):4958

47. Pierno A, Mari M, Lusher D, Castiello U (2008) Robotic movement elicits visuomotor priming in children with autism. Neuropsychologia 31:448–454

48. Pioggia G, Sica ML, Ferro M, Igliozzi R, Muratori F, Ahluwalia A, Rossi DD (2007) Human–robot interaction in autism: face, an android-based social therapy. In: Proceedings of IEEE international conference on robot and human interactive communication. IEEE, Jeju, p 605–612

49. Pop CA, Petrule A, Pintea S, Peca A, Simut R, Vanderborght B, David DO (2013a) Imitation and social behaviors of children with ASD in interaction with robonova. A series of single case experiments. Transylv J Psychol 14(1):71–93

50. Pop CA, Simut RE, Pintea S, Saldien J, Rusu AS, Vanderfaeillie J, David DO, Lefeber D, Vanderborght B (2013b) Social robots vs. computer display: Does the way social stories are delivered make a difference for their effectiveness on ASD children? Educ Comput Res 49(3):381–401

51. Reed FD, Hyman SR, Hirst JM (2011) Applications of technology to teach social skills to children with autism. Dev Neurorehabil 5:1003–1010

52. Reichow B, Volkmar FR, Cicchetti DV (2008) Method for evaluating and determining evidence-based practices in autism. Autism Dev Disord 38:1311–1319

53. Riek LD (2012) Wizard of Oz studies in HRI: a systematic review and new reporting guidelines. Int J Hum Robot Interact 1(1):119–136

54. Robins B, Dautenhahn K (2014) Tactile interactions with a humanoid robot: novel play scenario implementations with children with autism. Int J Soc Robot 6:397–415

55. Robins B, Dautenhahn K, te Boekhorst R, Billard A (2005) Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills. Univers Access Inf Soc 4:105–120

56. Robins B, Dautenhahn K, Dubowski J (2006) Does appearance matter in the interaction of children with autism with a humanoid robot? Interact Stud 7:479–512

57. Rogers SJ, Dawson G (2012) Play and engagement in early autism: the early start Denver model. www.ucdmc.ucdavis.edu/edsl/esdm/training.html. Accessed 3 Feb 2015

58. Rogers SJ, Vismara LA (2008) Evidence-based comprehensive treatments for early autism. J Clin Child Adolesc Psychol 37:8–38

59. Rutter M, Bailey A, Lord C, Berument SK (2003) Social communication questionnaire. Western Psychological Services, Los Angeles. http://www.wpspublish.com/app/. Accessed 3 Feb 2015

60. Scassellati B (2005) Quantitative metrics of social response for autism diagnosis. In: International workshop on robots and human interactive communication, p 585–590

61. Scassellati B, Admoni H, Mataric M (2012) Robots for use in autism research. Annu Rev Biomed Eng 14:275–294

62. Schopler E, Reichler RJ, Renner BR (1988) The childhood autism rating scale (CARS) manual. Western Psychological Services, Los Angeles. http://www.wpspublish.com/app/. Accessed 3 Feb 2015

63. Shamsuddin S, Yussof H, Ismail L, Hanapiah FA, Mohamed S, Piah HA, Zahair N (2012) Initial response in HRI—a case study on evaluation of child with autism spectrum disorders interacting with a humanoid robot NAO. In: Proceedings of the international symposium on robotics and intelligent sensors. Elsevier, Kuching, Sarawak, Malaysia, p 1448–1455

64. Srinivasan S, Lynch KA, Bubela DJ (2013) Effect of interactions between a child and a robot on the imitation and praxis performance of typically developing children and a child with autism: a preliminary study. Percept Motor Skills Phys Dev Meas 116(3):885–904

65. Tapus A, Peca A, Aly A, Pop C, Jisa L, Pintea S, Rusu AS, David DO (2012) Children with autism social engagement in interaction with NAO, an imitative robot. Interact Stud 13(3):315–347

66. Vanderborght B, Simut R, Saldien J, Pop C, Rusu AS, Pintea S, Lefeber D, David DO (2012) Using the social robot probo as a social story telling agent for children with ASD. Interact Stud 13:348–372

67. Wainer J, Robins B, Amirabdollahian F, Dautenhahn K (2014) Using the humanoid robot KASPAR to autonomously play triadic games and facilitate collaborative play among children with autism. Auton Ment Dev 6(3):183–199

68. Wilczynski S, Green G, Ricciardi J, Boyd B, Hume A, Ladd M, Odom S, Rue H (2009) National standards report: The national standards project addressing the need for evidence-based practice guidelines for autism spectrum disorders. National Autism Center. http://www.nationalautismcenter.org/affiliates/reports.php. Accessed 3 Feb 2015

69. Wong C, Odom S, Hume K, Cox AW, Fettig A, Kucharczyk S, Brock ME, Plavnick JB, Fleury VP, Schultz TR (2014) Evidence-based practices for children, youth, and young adults with autism spectrum disorder. The University of North Carolina, Frank Porter Graham Child Development Institute, Autism Evidence-Based Practice Review Group, Chapel Hill. http://autismpdc.fpg.unc.edu/sites/autismpdc.fpg.unc.edu/files/2014-EBP-Report.pdf. Accessed 3 Feb 2015

Momotaz Begum is a Research Assistant Professor in the Computer Science Department at the University of Massachusetts Lowell. Her research interests include robot perception and learning, human–robot interaction, assistive robotics and SLAM. The current focus of her research is to use robotics, computer vision and AI in novel ways to develop effective quality of life technologies for people with physical and cognitive disabilities. Begum received her PhD from the University of Waterloo, Canada and MEng from Memorial University, Canada, both in Electrical and Computer Engineering with a major in intelligent mobile robotics.

Richard W. Serna received a BA in Psychology from the University of New Mexico in 1980 and a PhD in Psychology from Utah State University in 1987. He spent 21 years at the University of Massachusetts Medical Schools Eunice Kennedy Shriver Center as a Faculty Member in the Behavioral Sciences Division. There he conducted federally funded translational stimulus-control research in auditory and visual stimulus discrimination with children with developmental disabilities. In 2011, he joined the Faculty of the Psychology Department at the

University of Massachusetts Lowell (UML) as an Associate Professor, where he teaches courses in learning, autism spectrum disorder (ASD) and behavioral intervention. His current diverse research work includes the study of methods for teaching symbol-referent relations and meaning to children with neuro-developmental disabilities and the development of virtual-reality methods for training practitioners who treat individuals with ASD. He currently serves on the Board of Editors for the Journal of the Experimental Analysis of Behavior, the Experimental Analysis of Human Behavior Bulletin, for which he is a former Editor, and the Brazilian Journal of Behavior Analysis. He is currently the Graduate Director of the Master of Science in Autism Studies Program at UML.

Holly A. Yanco is a Professor of Computer Science at the University of Massachusetts Lowell and the Director of the New England Robotics Validation and Experimentation (NERVE) Center. In 2015, she was awarded the designation of University Professor. Her research interests include human–robot interaction, multi-touch computing, interface design, robot autonomy, fostering trust of autonomous systems, evaluation methods for human–robot interaction, and the use of robots in K-12 education to broaden participation in computer science. Yanco's research has been funded by the National Science Foundation, including a CAREER Award, the Army Research Office, DARPA, NIST, Microsoft, and Google. Yanco was the General Chair of the 2012 ACM/IEEE International Conference on Human–Robot Interaction and is now the Co-chair of the conference's steering committee. She served on the Executive Council of the Association for the Advancement of Artificial Intelligence (AAAI) from 2006 to 2009 and was the Symposium Chair for AAAI from 2002 to 2005. She is a Senior Member of AAAI. Yanco has a PhD and MS in Computer Science from the Massachusetts Institute of Technology (MIT) and a BA in Computer Science and Philosophy from Wellesley College.