11-1-2021

Choosing Performance-Based Outcome Measures of Daily Living Skills and Socialization for Clinical Trials in Autistic Children

Roseann C Schaaf
Thomas Jefferson University

Amy Carroll
Thomas Jefferson University

Elizabeth Conte Waskie
Thomas Jefferson University

Rachel L Dumont
Thomas Jefferson University

Elizabeth Ridgway
Albert Einstein College of Medicine

Choosing Performance-based Outcome Measures for Clinical Trials in Autism

Schaaf RC, Carroll A, Waskie EC, Dumont RL, Ridgway E.

Am J Occup Ther. 2021; 75(6).

Abstract

Background: There is a need for robust and psychometrically sound, performance-based outcome measures of occupational therapy interventions for children with autism.

Objective: To demonstrate a systematic approach for choosing performance-based outcome measures of daily living skills and socialization for children with autism for use in clinical trials of occupational therapy interventions.

Methodology: Performance-based outcome measures of daily living skills and socialization were identified via review of the literature and hand searching. Psychometric properties and other measurement characteristics were rated by experts using a quality indicator scale. A nominal group process was used to achieve consensus on best measures for our planned clinical trial.

Results: Characteristics of each measure are reported as are key considerations for choosing outcome measures including the aims and scope of the planned study, time burden, and transportability from research to clinical practice.

Conclusions: This project demonstrates systematic process for choosing outcome measures for a planned trial.

Key words: activities of daily living, autistic disorder, child, occupational therapy, outcome assessment (health care), socialization

Introduction

Autism Spectrum Disorders (ASD) is characterized by difficulties in social communication and the presence of restrictive and repetitive behaviors. These symptoms affect performance in daily life activities (Schaaf, Toth-Cohen, Johnson, Outten, & Benevides, 2011) and thus, occupational therapy (OT) is frequently a component of a comprehensive program for

ASD. Occupational therapy interventions often address areas that are important to families including functional skills, participation in daily activities, and quality of life (Miller-Kuhaneck & Watling, 2015; Schaaf et al., 2011; Patten, Baranek, Watson, & Schultz, 2013) and as a result are one of the most valued and frequently requested services by parents (Goin-Kochel, Mackintosh, & Myers, 2009; Green et al., 2006, Peacock, 2012). Despite the value and high use of OT for children with ASD, there is a paucity of studies that measure OT outcomes with performance-based, objective measures. Thus, guidance for choosing robust and psychometrically sound measures that assess meaningful outcomes of interventions is needed (Bennett & Bennett, 2000; Lami, Egberts, Ure, Conroy, & Williams, 2017). Equally important is choosing measures consistent with the scope and aims of a particular study; and that are psychometrically strong, precise, and relevant for measuring outcomes reflective of family needs (Askari et al., 2015; Coster & Khetani, 2008; McConachie et al., 2015). Here, we describe a systematic process for identification and rating of performance-based outcome measures that can be used to identify appropriate outcome measures for OT clinical trials.

For this study, we identified two main areas for outcome measurement, Activities of Daily Living (ADLs) and socialization, as these are areas where we identified significant improvements in our pilot trials (Schaaf, Benevides, Kelly & Mailloux, 2012). In this pilot work, we used parent report outcome measures (PROs) to evaluate these constructs. PROs are an important strategy for outcome measurement as they provide the person's perception of change, however, they have limitations including potentially over or under estimating function (Weldring & Smith, 2013). Hence, to increase the rigor of the outcome measurement plan for our future studies, we sought to identify performance-based outcome measures to pair with the PROs.

Performance-based measures are administered by a trained evaluator who observes and rates the child's performance-based on a standard scale (Kazdin, 2013; Lami et al., 2017; Schaaf & Lane, 2015).

Methodology

Design: Mixed methods were used to identify and rate existing performance-based outcome measures of ADLs and socialization for children with ASD ages 6-9 years. A panel of experts reviewed and rated measures and then held a consensus meeting to identify the best measures for the planned trial.

Participants/Reviewers: Four experts in ASD and/or pediatric outcome measurement served as the reviewers. They included a Ph.D. neuropsychologist with expertise in ASD diagnosis, a Ph.D. occupational therapist with expertise in pediatric outcome measurement and instrument development, and two occupational therapists (one Ph.D. and one OTD) with clinical and research experience in ASD and instrument development. All had university academic appointments.

Procedures: The review and consensus process occurred sequentially. First, we conducted a rapid review[1] of the literature to identify performance-based measures of ADLs and socialization appropriate for use with children with ASD ages 6-9 years (the population for our planned clinical trial). Next, we adapted a quality indicator scale (QI) to rate each measure (described below). Ratings were collated and a nominal group process was held to achieve consensus on identifying the best measures for the planned trial. Each step is described below:

[1] A rapid review provides a time efficient strategy to identify, select, and critically appraise data from relevant research on a specific topic. It is a simplified approach to a systematic review where sources are limited due to time constraints (Khangura, Konnyu, Cushman, Grimshaw, & Moher, 2012).

Step 1: Identification of measures: To identify outcome measures that met the inclusion criteria, the second author conducted a search of the research and gray literature (books and book chapters, on-line assessment resources and websites). The authors also contacted their professional colleagues in the fields of OT and autism research and/or practice to identify outcome measures used in their work. Inclusion criteria: 1) appropriate for children with ASD, age 6-9 years, 2) evaluates ADL skills and/or socialization, 3) administered as a performance measure, and 4) has established psychometric properties. We identified the search terms: children, autism, performance-based measures, assessments, evaluations, evaluate, activities of daily living, daily living skills, daily activities, and/or socialization; and then searched PUBMED, CINAHL, Google Scholar, and OT Search. Measures meeting the inclusion criteria were organized on a secure shared drive including manuals and any relevant studies on the measure that described its psychometric properties.

Step 2: Modify an existing QI rating scale: The QI rating scale used for this project is based on the work by Law and MacDermid (2014). This scale was adapted for our projects needs in relation to clinical group (autism), age group, and focus on performance-based measures. Items on the QI scale address the psychometric properties of each measure on qualities such as reliability, validity, responsiveness to change, and characteristics of the measure such as purpose, scope, scoring, and administration time requirements. The QI scale was reviewed and field-tested for clarity, comprehensiveness, and redundancy by a measurement expert. Based on this review, thirteen items whose content was redundant with inclusion criteria were removed. Two items related to a measure's ability to interpret subscale scores were combined to one item, one item that addressed reliability was expanded into three items to more

clearly define types of reliability (intra-rater, inter-rater, and test-retest reliability), and two items that rated discriminant validity and use of Rasch Analysis were added. The final QI rating scale is detailed in Table 1.

(Insert Table 1 here)

Step 3: Identify expert reviewers: A convenience sampling of eight professionals with expertise in ASD and/or measurement received an email detailing the project and an invitation to serve as reviewers. Four accepted the invitation and received instructions for reviewing the instruments which were placed on a secure shared drive.

Step 4: Review of measures: Each expert independently reviewed and rated each measure. Ratings were submitted via encrypted e-mail.

Step 5: Collation of ratings: The scores for each QI rating from the four reviewers were compiled into an excel spreadsheet. The highest possible score obtainable for each measure by a rater was 23, and the highest possible sum score for a specific measure across all four raters was 92. Higher scores indicated stronger measurement characteristics and psychometric properties.

Step 6: Consensus meeting: The expert reviewers met with the project investigators using an on-line platform. The compiled results of the review were presented to the expert reviewers as the basis for discussion. Experts presented their rationale for ratings and their expert opinion about the measures. Group discussion facilitated the consensus process. Toward the end of the meeting, consensus was reached on the measures that best met the needs of our planned trial.

Results

1. Identification of measures that met inclusion criteria.

Ten performance-based ADL measures and eleven performance-based socialization measures were identified (total = 21 measures). Of these, seven met the inclusion criteria for this project: five ADL measures and two socialization measures. All identified measures are shown in Table 2. It is important to note that the exclusion of a specific measure for this project is not an indication of its potential value, but rather, that it did not fit the identified needs for our future clinical trial.

(Insert Table 2 here)

2. QI Scale ratings for each included measure.

The summed QI ratings for each measure are shown in Table 3 in alphabetical order. The Assessment of Motor and Process Skills (AMPS; Fisher, 2006) received the highest rating on the QI scale for the ADL measures, and the Evaluation of Social Interaction (ESI; Fisher & Griswold, 2010 received the highest rating for the socialization measures. These findings are shown in Table 3 and described.

(Insert Table 3 here)

3. Descriptive Analysis of Included Measures of Daily Living Skills and Socialization

Table 4 shows a description of each included measure and the QI points received. The strengths and limitations of each measure according to the needs of our planned trial are presented.

(Insert Table 4 here)

5. Nominal group process consensus

The experts concluded that the AMPS and the ESI were the best-suited outcome measures for our needs. Group consensus focused on the findings that both measures

have norms for ASD, have strong standardization, are capable of measuring change between known groups, and used Rasch analysis.

Discussion

Recognizing that selection of outcome instruments for a clinical trial is dependent upon the research hypothesis and study intent, this project was designed to guide the selection of performance-based outcome measures of ADLs and socialization for a planned clinical trial of an OT intervention for children with ASD ages 6-9 years. ADLs and socialization were chosen as outcomes based on a prior pilot study that showed that these areas were sensitive to change for the studied intervention (Schaaf et al., 2014). Thus, this is not intended to be a systematic, comprehensive review of outcome measures, but rather, a means to identify and get expert opinion for the future trial. Although there have been at least two systematic reviews of outcome measures for individuals with ASD published in the literature (Askari et al., 2015; McConachie et al., 2015), none were specifically focused on performance-based outcome measures, nor did they target the outcome areas of ADLs and socialization.

One value of this project is in demonstrating a process for choosing performance-based outcome measures that are consistent with the aims and scope of a specific, planned clinical trial. Selection of appropriate outcome measures that are sensitive to assessing change in dependent variables is an important aspect of study design. Thus, the procedures used in this project may be useful to guide researchers and clinicians through the process of outcome selection for clinical trials. Here, we highlight five important considerations when choosing outcome measures for OT clinical trials.

1. Careful assessment of the psychometric properties. Psychometrically robust outcome instruments that provide meaningful information are essential for clinical trials and intervention research, and for building evidence in support of OT (Coster & Khetani, 2008; Mulcahey et al., 2010). While there are several types of reliability and validity to consider, sensitivity of an instrument to detect change in response to treatment is arguably the most important psychometric property of an outcome measure used in clinical trials (Fok & Henry, 2015; McConachie et al., 2015). Accordingly, sensitivity to detect change was an important aspect of the modified QI scale and these measures scored highly on our QI scale.
2. Measurement burden on the family and child. While scientific integrity of outcome measurement in ASD research is a primary consideration, the impact of time burden on the child and family must also be considered (Ebesutani, Bernstein, Chorpita, & Weisz, 2012; Hinshaw et al., 2004). Excessive time required for participation in and completion of assessments/outcome measures may dissuade families from participation in research due to their daily life responsibilities. Further, measures that are time-intensive may hamper optimal performance by children with ASD who may have shortened attention or focus during the assessment process (Hinshaw et al., 2004). Recognizing the impact of time burden, the American Academy of Child and Adolescent Psychiatry (Hinshaw et al., 2004) recommended that protocols be streamlined to reduce burden. Their recommendations include automation of assessment protocols or the use of Item Response Theory (Coster, 2008).

3. Transportability of the measurement plan to clinical practice: This is an important consideration when the application of outcome measurement shifts from research to practice (Ebesutani et al., 2012). Often, the assessment phase in research is "time-intensive, resource heavy and may be too costly for real-world implementation" (Ebesutani et al., 2012, p. 141). To address this issue, Ebesutani et al. (2012) developed an algorithm-based assessment protocol that reduced administration and interpretation burden but maintained accuracy in identification and classification of participants to target appropriate intervention. Following this recommendation, algorithm-based decision making may be a potential strategy for OT. Development and testing of algorithms that can provide adequate information for characterization of subjects, and provide reliable, valid, and sensitive outcome measurement are important next steps to enhance research participation and translation to practice.
4. Norm-based versus criterion-references measurement: When considering an outcome measurement plan for a study, the approach to outcome measurement including the decision to use criterion-referenced and/or norm-referenced measures is important. For this study, we focused on performance-based, norm-referenced outcome measures because our aim is to compare ADL and socialization performance of our study groups to a normative sample. Our QI scale rated norm-referenced measures higher (better) than criterion-referenced outcome measures. The differences between criterion and norm-referenced tests have important implications for study design and outcome measurement plan.

Because our focus was on norm-based outcome measures, solid criterion-referenced measures of ADLs and social skills rated lower than norm-referenced ones. Case in point is the criterion-referenced GOAL measure (Miller et al., 2013). Although the GOAL scored highly on the QI items that evaluated standardization and responsiveness, it scored low on the items that rated normative reference because it is a criterion-referenced tool. One advantage of the GOAL is that it is sensitive to various degrees of change as it measures the magnitude of longitudinal change in ADLs allowing comparison of the child's performance in relation to their prior scores rather than comparing the score to a standard as in normed references testing. While this is a useful characteristic of an outcome measure, it did not meet the needs for our future trial and thus, its rating may not adequately represent its many strengths as a criterion-referenced outcome measure. This was also the case for the Social Profile (Donohue, 2013). Thus, this project highlights the importance of defining and describing the research question and expected outcomes clearly to determine whether norm-referenced or criterion-referenced measures or some combination of each, are best suited to the study objectives and design and choose outcome measures accordingly.

5. Measurement of outcomes that are relevant and meaningful to families: In OT outcome research, an important area of interest is the participant's ability to participate in meaningful life activities (Coster & Khetani, 2008). Thus, to accurately capture the diversity of performance skills and participation opportunities, it may be necessary to evaluate performance in context (i.e.: real-life environments) and include activities that are meaningful to the child/family. To accomplish this, an outcome measurement plan

that combines PROs may be needed (Askari et al., 2015; Coster, 2008; McConachie et al., 2015). PROs can provide a perspective on the value of a given outcome in the context of the individual's daily life. One value of the top-rated measures in this project, the AMPS and the ESI, is that they are performance-based outcome measures that appreciate context and meaning by utilizing tasks that are important to the individual in the most natural context possible. For these reasons, these measures were well suited for our OT clinical trials that focused on measuring functional outcomes.

Limitations

Despite the usefulness of this study for guiding outcome measurement planning for clinical trials of OT interventions for children with ASD, it is important to note that the findings from this project are applicable to specific, pre-determined criteria and may not be generalizable to other intervention trials. Further, prior knowledge of expert reviewers may have influenced the results. While the experts were highly qualified in ASD and outcome measurement, they had various degrees of expertise. We did not provide specific training on how to use the modified QI scale or establish competence in determining quality across reviewers, thus, interpretation and rating of measures may have been impacted by this varying expertise.

In terms of the QI rating scale, the ratings for each item were unweighted for relative importance which may have impacted the final rating score. However, this limitation was somewhat mediated by the reviewers' discussion during the consensus meeting where the needs of the planned clinical trial were considered in the final ratings.

For one instrument, the AMPS, the reviewers did not have access to the full manual which may have limited their knowledge of the AMPS. Finally, the QI ratings of the AFLS measure were

impacted because it did not have published information related to reliability, validity, responsiveness and time for administration at the time of the project. Thus, the clinicians and researchers seeking to use this criterion-referenced tool may want to check for updated data to evaluate its utility and rigor.

Implications for Occupational Therapy Practice

This project identified essential elements for consideration when choosing outcome measures for clinical trials including the aims and scope of the planned study, time burden, and transportability of measures from research to clinical practice.

- In choosing outcome measures for a clinical trial, consider the psychometric characteristics including validity and reliability, as well as the sensitivity to detecting change in the given construct.
- Time and attention burden for the child and family, as well as the clinicians, is an important consideration when choosing outcome measures.
- Consider the approach to outcome measurement including the decision to use criterion-referenced and/or norm-referenced measures.
- Focus on outcomes that are relevant and meaningful to families.

Conclusions

This project takes an important step forward by disseminating specific characteristics of outcome measures for OT intervention trials in ASD. In this paper, we introduce an approach to the review and identification of outcome measures for consideration in OT outcome studies. We highlight crucial considerations in identifying outcome measures for clinical trials including the aims and scope of the planned study, and point out considerations for transportability of research

to clinical practice. The methodology used in this project may guide other researchers in appropriate outcome measurement selection.

References

Askari, S., Anaby, D., Bergthorson, M., Majnemer, A., Elsabbagh, M., & Zwaigenbaum, L. (2015). Participation of children and youth with autism spectrum disorder: A scoping review. *Review Journal of Autism and Developmental Disorders, 2*(1), 103-114. https://doi.org/10.1007/s40489-014-0040-7

Bennett, S., & Bennett, J. W. (2000). The process of evidence-based practice in occupational therapy: Informing clinical decisions. *Australian Occupational Therapy Journal*, *47*(4), 171-180. https://doi.org/10.1046/j.1440-1630.2000.00237.x

Cohen, I. L., & Sudhalter, V. (2005). *The PDD Behavior Inventory*. Lutz, FL: Psychological Assessment Resources Inc.

Coster, W. J. (2008). Embracing ambiguity: Facing the challenge of measurement. *American Journal of Occupational Therapy*, *62*(6), 743-752. https://doi.org/10.5014/ajot.62.6.743

Coster, W., & Khetani, M. A. (2008). Measuring participation of children with disabilities: Issues and challenges. *Disability and rehabilitation, 30*(8), 639-648. https://doi.org/10.1080/09638280701400375

Coster, W. J., Law, M. C., & Bedell, G. M. (2010). *Participation and Environment Measure for Children and Youth (PEM-CY)*. Boston, MA: Boston University.

Donohue, M. V. (2013). *Social Profile: Assessment of social participation in children, adolescents, and adults*. Bethesda, MD: AOTA Press.

Durand, V. M., & Crimmins, D. B. (1988). *The motivation assessment scale*. Topeka, KS: Monaco & Associates.

Dykes, M. K., & Mruzek, D. W. (2012). *Developmental assessment for individuals with severe disabilities: DASH-3*. Austin, TX: PRO-ED.

Ebesutani, C., Bernstein, A., Chorpita, B. F., & Weisz, J. R. (2012). A transportable assessment protocol for prescribing youth psychosocial treatments in real-world settings. *Psychological assessment, 24*(1), 141. https://doi.org/10.1037/a0025176

Fisher, A. G. (2006). *Assessment of motor and process skills. Vol. 1: Development, standardization, and administration manual* (6th ed.). Fort Collins, CO: Three Star.

Fisher, A. G., & Griswold, L. A. (2010). *Evaluation of social interaction* (2nd ed.). Fort Collins, CO: Three Star Press.

Fok, C. C. T., & Henry, D. (2015). Increasing the sensitivity of measures to change. *Prevention Science, 16*(7), 978-986. https://doi.org/10.1007/s11121-015-0545-z

Gardner, M.F. (1994). *GS(M)DS: Gardner Social (Maturity) Developmental Scale: manual*. Burlingame, California: Psychological and Educational Publications.

Goin-Kochel, R. P., Mackintosh, V. H., & Myers, B. J. (2009). Parental reports on the efficacy of treatments and therapies for their children with autism spectrum disorders. *Research in Autism Spectrum Disorders*, *3*(2), 528-537. https://doi.org/10.1016/j.rasd.2008.11.001

Green, V. A., Pituch, K. A., Itchon, J., Choi, A., O'Reilly, M., & Sigafoos, J. (2006). Internet survey of treatments used by parents of children with autism. *Research in Developmental Disabilities*, *27*(1), 70-84. https://doi.org/10.1016/j.ridd.2004.12.002

Haley, S. M., Coster, W. J., Dumas, H. M., Fragala-Pinkham, M. A., & Moed, R. (2012). The pediatric evaluation of disability inventory-computer adaptive test. Boston, MA: CRE Care, LLC.

Harrison, P. L., & Oakland, T. (2003). *Adaptive Behavior Assessment System— Second Edition*. San Antonio, TX: Psychological Corporation.

Hinshaw, S. P., Hoagwood, K., Jensen, P. S., Kratochvil, C., Bickman, L., Clarke, G., ... & Vitiello, B. (2004). AACAP 2001 research forum: Challenges and recommendations regarding recruitment and retention of participants in research investigations. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(8), 1037-1045. https://doi.org/10.1097/01.chi.0000129222.89433.3d

Kazdin, A. E. (2013). *Behavior modification in applied settings* (7TH ed.). Belmont, CA: Wadsworth Thomson Learning.

Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., & Moher, D. (2012). Evidence summaries: the evolution of a rapid review approach. *Systematic reviews*, *1*(1), 10. https://doi.org/10.1186/2046-4053-1-10

King, G., Law, M., King, S., Hurley, P., Hanna, S., Kertoy, M., Rosenbaum, P., Young, N. (2004). *Children's Assessment of Participation and Enjoyment (CAPE) and Preferences for Activities of Children (PAC)*. San Antonio: Harcourt Assessment, Inc.

Lami, F., Egberts, K., Ure, A., Conroy, R., & Williams, K. (2017). Measurement properties of instruments that assess participation in young people with autism spectrum disorder: A systematic review. *Developmental Medicine & Child Neurology*, *60*(3), 230-243. https://doi.org/10.1111/dmcn.13631

Law, M. C., & MacDermid, J. (2014). *Evidence-based rehabilitation: A guide to practice*. Thorofare, NJ: Slack Incorporated.

McConachie, H. Parr, J.R., Glod, M., Hanratty, J…….Williams, K. (2015). Systematic review of tools to measure outcomes for young children with autism spectrum disorder. *Health Technology Assessment*, *19*(41). https://doi.org/10.3310/hta19410

Miller, L., Oakland, T., & Herzberg, D. (2013). *Goal oriented assessment of lifeskills*. San Antonio, TX: Pearson Assessments.

Miller-Kuhaneck, H., & Watling, R. (2015). Occupational therapy: Meeting the needs of families of people with autism spectrum disorder. *American Journal of Occupational Therapy*, *69*(5). https://doi.org/10.5014/ajot.2015.019562.

Mueller, M. M. & Partington, J. W. (2015). *The Assessment of Functional Living Skills*. Pleasant Hill, CA: Behavior Analysts, Inc.

Mulcahey, M. J., DiGiovanni, N., Calhoun, C., Homko, E., Riley, A., & Haley, S. M. (2010). Children's and parents' perspectives about activity performance and participation after spinal cord injury: Initial development of a patient-reported outcome measure. *American Journal of Occupational Therapy*, *64*(4), 605-613. https://doi.org/ 10.5014/ajot. 2010.08148

Partington, J. W. (2010). *The assessment of basic language and learning skills-revised (the ABLLS-R)*. Pleasant Hill, CA: Behavior Analysts.

Patten, E., Baranek, G. T., Watson, L. R., & Schultz, B. (2013). Child and family characteristics influencing intervention choices in autism spectrum disorders. *Focus on Autism and Other Developmental Disabilities*, *28*(3), 138-146. https://doi.org/10.1177/1088357612468028

Peacock, E. (2012). Top 8 Autism Therapies – Reported by Parents | Autism Speaks. Retrieved from https://www.autismspeaks.org/blog/top-8-autism-therapies-reported-parents

Portney, L.G., & Watkins, M.P. (2009). *Foundations of Clinical Research: Applications to Practice (3rd ed.).* Upper Saddle, NJ: Pearson Education, Inc.

Ricketts, C. (2009). A plea for the proper use of criterion-referenced tests in medical assessment. *Medical Education, 43*(12), 1141-1146. https://doi.org/10.1111/j.1365-2923.2009.03541.x

Roll, K. & Roll, W. (2013). The REAL: The Roll evaluation of activities of life User's guide. San Antonio, TX: Pearson Assessment.

Schaaf, R. C., Benevides, T. W., Mailloux, Z., Faller, P., Hunt, J., van Hooydonk, E., Freeman, R., Lieby, B., Sendecki, J., Kelly, D. (2014). An intervention for sensory difficulties in children with autism: A randomized control trial. *Journal of Autism and Developmental Disorders, 44,* 1493-1506. https://doi.org/10.1007%2Fs10803-013-1983-8

Schaaf, R. C., Benevides, T. W., Kelly, D., & Mailloux-Maggio, Z. (2012). Occupational Therapy and sensory integration for children with autism: A feasibility, safety, acceptability and fidelity study. *Autism*, *16*(3), 321-327. https://doi.org/10.1177/1362361311435157

Schaaf, R. C., Toth-Cohen, S., Johnson, S. L., Outten, G., & Benevides, T. W. (2011). The everyday routines of families of children with autism: Examining the impact of sensory processing difficulties on the family. *Autism*, *15*(3), 373-389. https://doi.org/10.1177/1362361310386505

Schaaf, R. C., & Lane, A. E. (2015). Toward a best-practice protocol for assessment of sensory features in ASD. *Journal of Autism and Developmental Disorders, 45*(5), 1380-1395. https://doi.org/10.1007/s10803-014-2299-z

Sparrow, S. S., & Cicchetti, D. V., & Balla, D. (2005). *Vineland adaptive behavior scales* (2nd ed.). Circle Pines, MN: American Guidance Service.

Uniform Data System for Medical Rehabilitation. (2006). The WeeFIM II clinical guide, v.6.0. Buffalo, NY: Elsevier

Weldring, T., & Smith, S. M. S. (2013). Patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). *Health Services Insights, 6*, 61–68. https://doi.org/10.4137/HSI.S11093

Table 1

Quality Indicator Rating Scale

Quality Indicators for Outcome Measures			
Instructions: Please use the quality indicators below to rate each of the 7 assessments using this Quality Indicators for Outcome Measures scale. We have placed the materials needed to perform the review and complete the quality rating scale on line at *Google Docs* for your ease (we will give you access to this). **Background:** Step one of this study consisted of a review of currently available performance-based assessments that measure participation – related outcomes. Seven instruments met the following inclusion criteria: • objective, performance-based outcome measures • address the relevant domains of interest (daily living skills and socialization) • valid for use with children with ASD ages 6-9 years • adequate reliability and validity The next step is to rate each of the 7 instruments on these quality indicators. After rating is complete, we will participate in an on-line meeting to discuss and come to consensus. Thank you for your participation!			
Evaluator:			**Date:**
Tool Title and Author			
Indicator	**Scale**	**Score**	**Comments**
1. What is the scope of the measure?	**1=** Full Comprehensive **0=** Limited or Unknown		
2. Are there sub-scale scores and/or total scores?	**1=** Sub-sales and Total Scores **0=** Only Total Scores		
3. What is the tool's purpose?[a]	**2= To evaluate-** measure the magnitude of longitudinal change in an individual or group on the dimension of interest		

(choose the higher value if more than one purpose)	**1= To predict-** classify individuals into a predetermined category **0= To discriminate**-distinguishes between individuals or groups when no external criterion is available to validate these measures		
4. Time to administer	**3=** 0-30 min. **2=** 30-60 min. **1=** 60-90 min. **0=** >90 min		
Standardization: for adequate standardization, there needs to be a normative study that has a sample size large enough that it is determined to result in an adequate effect size to achieve adequate statistical power. This is something determined statistically and reported in the literature. **Effect size**: a statistical expression of the magnitude of the difference between two treatments or the magnitude of a relationship between two variables **Power:** The ability of a statistical test to find a significant difference that really does exist			
Indicator	**Scale**	**Score**	**Comment**
5. Standardization	**2= Excellent**: available and complete with specific procedures for administration, scoring, and interpretation evidence of reliability and validity **1= Adequate**: available, generally complete but some info is lacking or unclear re: admin, scoring, interpretation or evidence for reliability and validity **0= Poor**: no manual available or manual with unclear administration, scoring and interpretation, no evidence of reliability and validity		

6. Are norms available for ages 6-9 years?	**1=** Yes; **0=** No		
7. Are there norms for ASD 6-9 years?	**1=** Yes; **0=** No If Yes, please comment on the representation of ASD in the sample.		
Reliability: is the process for determining that the test or measure is measuring something in a reproducible and consistent fashion. **Note: Guidelines for level of the reliability coefficient indicate that it will be rated excellent if the coefficient is greater than .80, adequate if it is from .60 to .79, and poor if it is less than .60**			
Indicator	**Scale**	**Score**	**Comment**
8. Reliability: Intra-observer (Intra-rater): measures variation that occurs within an observer as a result of multiple exposures to the same stimulus/test item	**2= Excellent** if the coefficient is greater than .80, **1= Adequate** if it is from .60 to .79 **0= Poor** if it is less than .60 or Unknown		
9. Reliability: Inter-observer- measures variation between two observers.	**2= Excellent** if the coefficient is greater than .80, **1= Adequate** if it is from .60 to .79 **0= Poor** if it is less than .60 or Unknown		
10. Reliability: Test-retest reliability: measures variations in the test over a period of time	**2= Excellent** if the coefficient is greater than .80, **1= Adequate** if it is from .60 to .79 **0= Poor** if it is less than .60 or Unknown		

Validity: is the degree to which an instrument measures what it is intended to measure.			
Indicator	**Scales**	**Score**	**Comments**
11. Validity: Content - comprehensive and fully represents the domain of the characteristics it claims to measure.	**2=** Negligible ceiling and floor effects (>20%) **1=** Moderate ceiling and floor effects (6-20%) **0=** Unacceptable or unknown celling and floor effects (<5%)		
12. Discriminant Validity: The ability to distinguish between constructs that should not be related to each other (i.e. different dx.)	**1=** Acceptable **0=** Unacceptable of Unknown		
Outcome Measure: To determine the appropriateness of a tool for use as an outcome measure it needs to be determine to have adequate: reliability, validity, and variability. It also needs to have adequate responsiveness (http://www.ncbi.nlm.nih.gov/books/NBK126186/			
Indicator	**Scales**	**Score**	**Comments**
13. Responsiveness: the ability to detect minimally clinically important change over time	**2= Excellent** **1= Adequate** **0= Poor**		
14. Has the measure been exposed to Rasch analysis?	**1=** Yes; **0=** No		

Note: Adapted from quality indicator scale by Law & MacDermind (2014) and Portney & Watkins (2009).

[a] This item's (3. Purpose) rating is not intended to be a Likert scale, however since we were primarily interested in the ability of the measure to evaluate the magnitude of change over time, the purpose "evaluate" was weighted highest. In no way are we negating the importance of the other purposes.

Table 2

Identified Outcome Measures

ADL Assessments

Outcome Measure	Included	Excluded	Reason for Exclusion
Adaptive Behavior Assessment System, 2nd ed. (Harrison & Oakland, 2003)		X	Not performance-based; Parent/caregiver checklist
The Assessment of Basic Language and Learning Skills- Revised (Partington, 2010)	X		
Assessment of Functional Living Skills (Mueller & Partington, 2015)	X		
The Assessment of Motor and Process Skills (Fisher, 2006)	X		
Children Assessment of Participation and Enjoyment and Preferences for Activities of Children (King et al., 2004)		X	Not performance-based; Child completes
Developmental Assessment for Individuals with Severe Disabilities, 3rd edition (Dykes & Mruzek, 2012)	X		

The Functional Independence Measure (WeeFIM; Uniform Data System for Medical Rehabilitation, 2006)		X	Used for children 6 months to 7 years
Goal-Oriented Assessment of Lifeskills (Miller, Oakland, & Herzberg, 2013)	X		
Pediatric Evaluation of Disability Inventory-Computer Adaptive Test (Haley, Coster, Dumas, Fragala-Pinkhal, & Moed, 2012)		X	Not performance-based; Parent report or therapist report based on professional judgment
Roll Evaluation of Activities of Life (Roll & Roll, 2013)		X	Not performance-based; Parent/caregiver rating
Social Assessments			
Outcome Measure	Included	Excluded	Reason for Exclusion
Adaptive Behavior Assessment System, 2nd ed. (Harrison & Oakland, 2003)		X	Not performance-based; Parent/caregiver/teacher behavior rating
Children Assessment of Participation and Enjoyment and Preferences for Activities of Children (King et al., 2004)		X	Not performance-based; Child completes

Evaluation of Social Interaction, 2nd ed. (Fisher & Griswold, 2010)	X		
The Garden Social Development Scale (Gardner, 1994)		X	Parent rating
The Motivation Assessment Scale (Durand & Crimmins, 1988)		X	Checklist/questionnaire about challenging behaviors
Participation and Environmental Measure for Children and Youth (Coster, Law, & Bedell, 2010)		X	Parent report
PDD Behavior Inventory (Cohen & Sudhalter, 2005)		X	Teacher/parent rating
Pediatric Evaluation of Disability Inventory-Computer Adaptive Test (Haley et al., 2012)		X	Parent/professional performs
Roll Evaluation of Activities of Life (Roll & Roll, 2013)		X	Parent/caregiver rating
Social Profile (Donohue, 2013)	X		
The Vineland Adaptive Behavior Scales (2nd ed.) (Sparrow, Cicchetti, & Balla, 2005)		X	Not performance-based; parent/teacher report or interview

Table 3

Sum of Quality Indicator Ratings for Each Measure

		Measures of ADLs						Measures of Socialization	
Quality Indicator	Scale Range	Total Possible Score	ABLLS-R	AFLS	AMPS	DASH-3	GOAL	ESI	SP
1. The scope of the measure is consistent with the study needs	0-1	4	4	4	4	3	1	2	1
2. Subscales / total scores are available	0-1	4	3	4	4	2	4	1	3
3. Measures change or distinguish between known group	0-2	8	8	8	8	8	8	8	5
4. Feasible for ASD in terms of time to administer	0-3	12	1	0	8	0	6	N/A	6
5.Instrument is standardized	0-2	8	5	3	7	3	7	8	5

6. Norms for 6-9 yrs.	0-1	4	0	1	4	0	2	4	0
7. Norms for ASD 6-9 yrs.	0-1	4	0	1	1	0	1	2	0
8. Intra-observer (rater) reliability	0-2	8	2	0	6	0	1	8	1
Total Points (scored/ total possible points)		**92**	**40/92**	**27/92**	**75/92**	**19/92**	**52/92**	**70/92**	***37/92**

(n = 4 reviewers)

Note. Summed scores for each item for each measure, and total scores for each measure.

*=Not scored by one reviewer

Abbreviations:

ABLLS-R= Assessment of Basic Language and Learning Skills-Revised.

AFLS= Assessment of Functional Living Skills.

AMPS= Assessment of Motor and Process Skills.

DASH-3= Developmental Assessment of Individuals with Severe Disabilities, 3rd edition.

ESI= Evaluation of Social Interaction.

GOAL= Goal-Oriented Assessment of Life skills.

NA= Not Applicable.

SP= Social Profile.

Table 4

Description of Included Measures and Quality Indicator Points Obtained

Analysis of Measures of Daily Living Skills

Measure	QI score of 92 possible points	Strengths (S) and Limitations (L) of each measure for our planned trial
Assessment of Motor and Process Skills (AMPS; Fisher, 2006)	75	**S:** Standardized observation-based measuring performance in ADLs/IADLS for persons with developmental age 2+, natural environment, yields motor and process score **L:** N/A
The Assessment of Basic Language and Learning Skills- Revised (ABLLS-R; Partington, 2010)	40	**S:** 544 skills in 25 skill areas, good inter-rater reliability, scope, magnitude of change **L:** Criterion-referenced, administration time, intra-rater reliability, responsiveness, exposure to Rasch analysis
Assessment of Functional Living Skills (AFLS; Mueller & Partington, 2015)	27	**S:** Observation-based, assessment measuring over 1,900 functional living skills, scope, inclusion of subscale and/or total scores, ability to assess the magnitude of longitudinal change in

		the area of interest **L:** Criterion-referenced, lower on norms for 6-9 year children with and without ASD.
*Developmental Assessment for Individuals with Severe Disabilities, 3rd edition (*DASH-3; Dykes & Mruzek, 2012)	19	**S:** Measures performance in a developmental sequence for children with mild to severe disabilities **L:** Criterion-referenced assessment, lower rating for norms, on some psychometric properties including discriminant validity, children functioning at a chronological age of birth to 7 years.
Goal-Oriented Assessment of Lifeskills (GOAL; Miller et al., 2013)	52	**S:** Assesses fine and gross motor skills during ADLs via observation of task performance in children age 7-17 years. **L:** Study requires minimal of 6 years.
<u>Measures of Socialization</u>		
Evaluation of Social Interaction, 2nd edition (ESI; Fisher & Griswold,	70	**S:** Quality of social interaction via observation of an individual (2 years to adulthood) in two

2010)		social interactions with typical social partners **L:** Lower on norms for ASD (small sample), subscales or totals scores, discriminative validity, and responsiveness.
The Social Profile (SP; Donohue, 2013)	37	**S:** Assesses individual or group behaviors on three subscales: activity participation, social interaction, and group membership, test-retest reliability, and inter-observer reliability **L:** Criterion-referenced. One rater did not rate the SP.