

11-1-2021

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

Roseann C Schaaf

Amy Carroll

Elizabeth Conte Waskie

Rachel L Dumont

Elizabeth Ridgway

Follow this and additional works at: <https://jdc.jefferson.edu/otfp>



Part of the [Occupational Therapy Commons](#)

[Let us know how access to this document benefits you](#)

---

This Article is brought to you for free and open access by the Jefferson Digital Commons. The Jefferson Digital Commons is a service of Thomas Jefferson University's [Center for Teaching and Learning \(CTL\)](#). The Commons is a showcase for Jefferson books and journals, peer-reviewed scholarly publications, unique historical collections from the University archives, and teaching tools. The Jefferson Digital Commons allows researchers and interested readers anywhere in the world to learn about and keep up to date with Jefferson scholarship. This article has been accepted for inclusion in Department of Occupational Therapy Faculty Papers by an authorized administrator of the Jefferson Digital Commons. For more information, please contact: [JeffersonDigitalCommons@jefferson.edu](mailto:JeffersonDigitalCommons@jefferson.edu).

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).

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

### 1 Abstract

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

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

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

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

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

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

### 18 Introduction

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

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

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

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

45 Performance-based measures are administered by a trained evaluator who observes and rates the  
46 child's performance-based on a standard scale (Kazdin, 2013; Lami et al., 2017; Schaaf & Lane,  
47 2015).

### 48 Methodology

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

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

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

---

<sup>1</sup> 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).

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

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

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

90 (Insert Table 1 here)

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

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

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

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

107 Results

108 1. Identification of measures that met inclusion criteria.

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

115 (Insert Table 2 here)

116 2. QI Scale ratings for each included measure.

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

122 (Insert Table 3 here)

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

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

127 (Insert Table 4 here)

128 5. Nominal group process consensus

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





## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

- 152 1. Careful assessment of the psychometric properties. Psychometrically robust outcome  
153 instruments that provide meaningful information are essential for clinical trials and  
154 intervention research, and for building evidence in support of OT (Coster & Khetani,  
155 2008; Mulcahey et al., 2010). While there are several types of reliability and validity to  
156 consider, sensitivity of an instrument to detect change in response to treatment is  
157 arguably the most important psychometric property of an outcome measure used in  
158 clinical trials (Fok & Henry, 2015; McConachie et al., 2015). Accordingly, sensitivity to  
159 detect change was an important aspect of the modified QI scale and these measures  
160 scored highly on our QI scale.
- 161 2. Measurement burden on the family and child. While scientific integrity of outcome  
162 measurement in ASD research is a primary consideration, the impact of time burden on  
163 the child and family must also be considered (Ebesutani, Bernstein, Chorpita, & Weisz,  
164 2012; Hinshaw et al., 2004). Excessive time required for participation in and completion  
165 of assessments/outcome measures may dissuade families from participation in research  
166 due to their daily life responsibilities. Further, measures that are time-intensive may  
167 hamper optimal performance by children with ASD who may have shortened attention or  
168 focus during the assessment process (Hinshaw et al., 2004). Recognizing the impact of  
169 time burden, the American Academy of Child and Adolescent Psychiatry (Hinshaw et al.,  
170 2004) recommended that protocols be streamlined to reduce burden. Their  
171 recommendations include automation of assessment protocols or the use of Item  
172 Response Theory (Coster, 2008).

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

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

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

### 223 Limitations

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

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

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

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

### 242 Implications for Occupational Therapy Practice

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

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

### 254 Conclusions

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

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

260 to clinical practice. The methodology used in this project may guide other researchers in

261 appropriate outcome measurement selection.

262

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

### 263 References

- 264 Askari, S., Anaby, D., Bergthorson, M., Majnemer, A., Elsabbagh, M., & Zwaigenbaum, L.  
265 (2015). Participation of children and youth with autism spectrum disorder: A scoping  
266 review. *Review Journal of Autism and Developmental Disorders*, 2(1), 103-114.  
267 <https://doi.org/10.1007/s40489-014-0040-7>
- 268 Bennett, S., & Bennett, J. W. (2000). The process of evidence-based practice in occupational  
269 therapy: Informing clinical decisions. *Australian Occupational Therapy Journal*, 47(4),  
270 171-180. <https://doi.org/10.1046/j.1440-1630.2000.00237.x>
- 271 Cohen, I. L., & Sudhalter, V. (2005). *The PDD Behavior Inventory*. Lutz, FL: Psychological  
272 Assessment Resources Inc.
- 273 Coster, W. J. (2008). Embracing ambiguity: Facing the challenge of measurement. *American*  
274 *Journal of Occupational Therapy*, 62(6), 743-752. <https://doi.org/10.5014/ajot.62.6.743>
- 275 Coster, W., & Khetani, M. A. (2008). Measuring participation of children with disabilities:  
276 Issues and challenges. *Disability and rehabilitation*, 30(8), 639-648.  
277 <https://doi.org/10.1080/09638280701400375>
- 278 Coster, W. J., Law, M. C., & Bedell, G. M. (2010). *Participation and Environment Measure for*  
279 *Children and Youth (PEM-CY)*. Boston, MA: Boston University.
- 280 Donohue, M. V. (2013). *Social Profile: Assessment of social participation in children,*  
281 *adolescents, and adults*. Bethesda, MD: AOTA Press.
- 282 Durand, V. M., & Crimmins, D. B. (1988). *The motivation assessment scale*. Topeka, KS:  
283 Monaco & Associates.



## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

- 284 Dykes, M. K., & Mruzek, D. W. (2012). *Developmental assessment for individuals with severe*  
285 *disabilities: DASH-3*. Austin, TX: PRO-ED.
- 286 Ebesutani, C., Bernstein, A., Chorpita, B. F., & Weisz, J. R. (2012). A transportable assessment  
287 protocol for prescribing youth psychosocial treatments in real-world  
288 settings. *Psychological assessment*, 24(1), 141. <https://doi.org/10.1037/a0025176>
- 289 Fisher, A. G. (2006). *Assessment of motor and process skills. Vol. 1: Development,*  
290 *standardization, and administration manual* (6th ed.). Fort Collins, CO: Three Star.
- 291 Fisher, A. G., & Griswold, L. A. (2010). *Evaluation of social interaction* (2nd ed.). Fort Collins,  
292 CO: Three Star Press.
- 293 Fok, C. C. T., & Henry, D. (2015). Increasing the sensitivity of measures to change. *Prevention*  
294 *Science*, 16(7), 978-986. <https://doi.org/10.1007/s11121-015-0545-z>
- 295 Gardner, M.F. (1994). *GS(M)DS: Gardner Social (Maturity) Developmental Scale: manual*.  
296 Burlingame, California: Psychological and Educational Publications.
- 297 Goin-Kochel, R. P., Mackintosh, V. H., & Myers, B. J. (2009). Parental reports on the efficacy of  
298 treatments and therapies for their children with autism spectrum disorders. *Research in*  
299 *Autism Spectrum Disorders*, 3(2), 528-537. <https://doi.org/10.1016/j.rasd.2008.11.001>
- 300 Green, V. A., Pituch, K. A., Itchon, J., Choi, A., O'Reilly, M., & Sigafos, J. (2006). Internet  
301 survey of treatments used by parents of children with autism. *Research in Developmental*  
302 *Disabilities*, 27(1), 70-84. <https://doi.org/10.1016/j.ridd.2004.12.002>
- 303 Haley, S. M., Coster, W. J., Dumas, H. M., Fragala-Pinkham, M. A., & Moed, R. (2012). The  
304 pediatric evaluation of disability inventory-computer adaptive test. Boston, MA: CRE  
305 Care, LLC.

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

- 306 Harrison, P. L., & Oakland, T. (2003). *Adaptive Behavior Assessment System— Second Edition*.  
307 San Antonio, TX: Psychological Corporation.
- 308 Hinshaw, S. P., Hoagwood, K., Jensen, P. S., Kratochvil, C., Bickman, L., Clarke, G., ... &  
309 Vitiello, B. (2004). AACAP 2001 research forum: Challenges and recommendations  
310 regarding recruitment and retention of participants in research investigations. *Journal of*  
311 *the American Academy of Child & Adolescent Psychiatry*, 43(8), 1037-1045.  
312 <https://doi.org/10.1097/01.chi.0000129222.89433.3d>
- 313 Kazdin, A. E. (2013). *Behavior modification in applied settings* (7<sup>TH</sup> ed.). Belmont, CA:  
314 Wadsworth Thomson Learning.
- 315 Khangura, S., Konnyu, K., Cushman, R., Grimshaw, J., & Moher, D. (2012). Evidence  
316 summaries: the evolution of a rapid review approach. *Systematic reviews*, 1(1), 10.  
317 <https://doi.org/10.1186/2046-4053-1-10>
- 318 King, G., Law, M., King, S., Hurley, P., Hanna, S., Kertoy, M., Rosenbaum, P., Young,  
319 N. (2004). *Children's Assessment of Participation and Enjoyment (CAPE) and*  
320 *Preferences for Activities of Children (PAC)*. San Antonio: Harcourt Assessment, Inc.
- 321 Lami, F., Egberts, K., Ure, A., Conroy, R., & Williams, K. (2017). Measurement properties of  
322 instruments that assess participation in young people with autism spectrum disorder: A  
323 systematic review. *Developmental Medicine & Child Neurology*, 60(3), 230-243.  
324 <https://doi.org/10.1111/dmcn.13631>
- 325 Law, M. C., & MacDermid, J. (2014). *Evidence-based rehabilitation: A guide to practice*.  
326 Thorofare, NJ: Slack Incorporated.

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

- 327 McConachie, H. Parr, J.R., Glod, M., Hanratty, J. . . . . Williams, K. (2015). Systematic review  
328 of tools to measure outcomes for young children with autism spectrum disorder. *Health*  
329 *Technology Assessment, 19*(41). <https://doi.org/10.3310/hta19410>
- 330 Miller, L., Oakland, T., & Herzberg, D. (2013). *Goal oriented assessment of lifeskills*. San  
331 Antonio, TX: Pearson Assessments.
- 332 Miller-Kuhaneck, H., & Watling, R. (2015). Occupational therapy: Meeting the needs of families  
333 of people with autism spectrum disorder. *American Journal of Occupational*  
334 *Therapy, 69*(5). <https://doi.org/10.5014/ajot.2015.019562>.
- 335 Mueller, M. M. & Partington, J. W. (2015). *The Assessment of Functional Living Skills*. Pleasant  
336 Hill, CA: Behavior Analysts, Inc.
- 337 Mulcahey, M. J., DiGiovanni, N., Calhoun, C., Homko, E., Riley, A., & Haley, S. M. (2010).  
338 Children's and parents' perspectives about activity performance and participation after  
339 spinal cord injury: Initial development of a patient-reported outcome measure. *American*  
340 *Journal of Occupational Therapy, 64*(4), 605-613. [https://doi.org/10.5014/ajot.](https://doi.org/10.5014/ajot.2010.08148)  
341 [2010.08148](https://doi.org/10.5014/ajot.2010.08148)
- 342 Partington, J. W. (2010). *The assessment of basic language and learning skills-revised (the*  
343 *ABLSS-R)*. Pleasant Hill, CA: Behavior Analysts.
- 344 Patten, E., Baranek, G. T., Watson, L. R., & Schultz, B. (2013). Child and family characteristics  
345 influencing intervention choices in autism spectrum disorders. *Focus on Autism and*  
346 *Other Developmental Disabilities, 28*(3), 138-146.  
347 <https://doi.org/10.1177/1088357612468028>

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

- 348 Peacock, E. (2012). Top 8 Autism Therapies – Reported by Parents | Autism Speaks. Retrieved  
349 from <https://www.autismspeaks.org/blog/top-8-autism-therapies-reported-parents>
- 350 Portney, L.G., & Watkins, M.P. (2009). *Foundations of Clinical Research: Applications to*  
351 *Practice (3<sup>rd</sup> ed.)*. Upper Saddle, NJ: Pearson Education, Inc.
- 352 Ricketts, C. (2009). A plea for the proper use of criterion-referenced tests in medical  
353 assessment. *Medical Education*, 43(12), 1141-1146. [https://doi.org/10.1111/j.1365-](https://doi.org/10.1111/j.1365-2923.2009.03541.x)  
354 [2923.2009.03541.x](https://doi.org/10.1111/j.1365-2923.2009.03541.x)
- 355 Roll, K. & Roll, W. (2013). The REAL: The Roll evaluation of activities of life User's  
356 guide. San Antonio, TX: Pearson Assessment.
- 357 Schaaf, R. C., Benevides, T. W., Mailloux, Z., Faller, P., Hunt, J., van Hooydonk, E., Freeman,  
358 R., Lieby, B., Sendeki, J., Kelly, D. (2014). An intervention for sensory difficulties in  
359 children with autism: A randomized control trial. *Journal of Autism and Developmental*  
360 *Disorders*, 44, 1493-1506. <https://doi.org/10.1007/s10803-013-1983-8>
- 361 Schaaf, R. C., Benevides, T. W., Kelly, D., & Mailloux-Maggio, Z. (2012). Occupational  
362 Therapy and sensory integration for children with autism: A feasibility, safety,  
363 acceptability and fidelity study. *Autism*, 16(3), 321-327.  
364 <https://doi.org/10.1177/1362361311435157>
- 365 Schaaf, R. C., Toth-Cohen, S., Johnson, S. L., Outten, G., & Benevides, T. W. (2011). The  
366 everyday routines of families of children with autism: Examining the impact of sensory  
367 processing difficulties on the family. *Autism*, 15(3), 373-389.  
368 <https://doi.org/10.1177/1362361310386505>

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

- 369 Schaaf, R. C., & Lane, A. E. (2015). Toward a best-practice protocol for assessment of sensory  
370 features in ASD. *Journal of Autism and Developmental Disorders*, 45(5), 1380-1395.  
371 <https://doi.org/10.1007/s10803-014-2299-z>
- 372 Sparrow, S. S., & Cicchetti, D. V., & Balla, D. (2005). *Vineland adaptive behavior scales* (2nd  
373 ed.). Circle Pines, MN: American Guidance Service.
- 374 Uniform Data System for Medical Rehabilitation. (2006). *The WeeFIM II clinical guide*,  
375 v.6.0. Buffalo, NY: Elsevier
- 376 Weldring, T., & Smith, S. M. S. (2013). Patient-reported outcomes (PROs) and patient-reported  
377 outcome measures (PROMs). *Health Services Insights*, 6, 61–68. [https://doi.org/  
378 10.4137/HSLS11093](https://doi.org/10.4137/HSLS11093)
- 379

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

Table 1

### *Quality Indicator Rating Scale*

Quality Indicators for Outcome Measures			
<p><b>Instructions:</b> 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 <i>Google Docs</i> for your ease (we will give you access to this).</p> <p><b>Background:</b> 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:</p> <ul style="list-style-type: none"> <li>• objective, performance-based outcome measures</li> <li>• address the relevant domains of interest (daily living skills and socialization)</li> <li>• valid for use with children with ASD ages 6-9 years</li> <li>• adequate reliability and validity</li> </ul> <p>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!</p>			
<b>Evaluator:</b>			<b>Date:</b>
<b>Tool Title and Author</b>			
Indicator	Scale	Score	Comments
<b>1. What is the scope of the measure?</b>	<b>1=</b> Full Comprehensive <b>0=</b> Limited or Unknown		
<b>2. Are there sub-scale scores and/or total scores?</b>	<b>1=</b> Sub-sales and Total Scores <b>0=</b> Only Total Scores		
<b>3. What is the tool's purpose?<sup>a</sup></b>	<b>2= To evaluate-</b> measure the magnitude of longitudinal change in an individual or group on the dimension of interest		

CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

(choose the higher value if more than one purpose)	<p><b>1= To predict-</b> classify individuals into a predetermined category</p> <p><b>0= To discriminate-</b>distinguishes between individuals or groups when no external criterion is available to validate these measures</p>		
<b>4. Time to administer</b>	<b>3=</b> 0-30 min. <b>2=</b> 30-60 min. <b>1=</b> 60-90 min. <b>0=</b> >90 min		
<p><b>Standardization: for adequate standardization</b>, 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.</p> <p><b>Effect size:</b> a statistical expression of the magnitude of the difference between two treatments or the magnitude of a relationship between two variables</p> <p><b>Power:</b> The ability of a statistical test to find a significant difference that really does exist</p>			
<b>Indicator</b>	<b>Scale</b>	<b>Score</b>	<b>Comment</b>
<b>5. Standardization</b>	<p><b>2= Excellent:</b> available and complete with specific procedures for administration, scoring, and interpretation evidence of reliability and validity</p> <p><b>1= Adequate:</b> available, generally complete but some info is lacking or unclear re: admin, scoring, interpretation or evidence for reliability and validity</p> <p><b>0= Poor:</b> no manual available or manual with unclear administration, scoring and interpretation, no evidence of reliability and validity</p>		

CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

<p><b>6. Are norms available for ages 6-9 years?</b></p>	<p><b>1= Yes; 0= No</b></p>		
<p><b>7. Are there norms for ASD 6-9 years?</b></p>	<p><b>1= Yes; 0= No</b>  If Yes, please comment on the representation of ASD in the sample.</p>		
<p><b>Reliability: is the process for determining that the test or measure is measuring something in a reproducible and consistent fashion.</b></p> <p><b>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</b></p>			
<p><b>Indicator</b></p>	<p><b>Scale</b></p>	<p><b>Score</b></p>	<p><b>Comment</b></p>
<p><b>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</b></p>	<p><b>2= Excellent</b> if the coefficient is greater than .80,  <b>1= Adequate</b> if it is from .60 to .79  <b>0= Poor</b> if it is less than .60 or Unknown</p>		
<p><b>9. Reliability: Inter-observer- measures variation between two observers.</b></p>	<p><b>2= Excellent</b> if the coefficient is greater than .80,  <b>1= Adequate</b> if it is from .60 to .79  <b>0= Poor</b> if it is less than .60 or Unknown</p>		
<p><b>10. Reliability: Test-retest reliability: measures variations in the test over a period of time</b></p>	<p><b>2= Excellent</b> if the coefficient is greater than .80,  <b>1= Adequate</b> if it is from .60 to .79  <b>0= Poor</b> if it is less than .60 or Unknown</p>		



CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

<sup>a</sup> 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.

CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

Table 2

*Identified Outcome Measures*

**ADL Assessments**

<u>Outcome Measure</u>	<u>Included</u>	<u>Excluded</u>	<u>Reason for Exclusion</u>
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		

CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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
<b>Social Assessments</b>			
<u>Outcome Measure</u>	<u>Included</u>	<u>Excluded</u>	<u>Reason for Exclusion</u>
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

CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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 (2 <sup>nd</sup> ed.) (Sparrow, Cicchetti, & Balla, 2005)		X	Not performance-based; parent/teacher report or interview

CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

--

CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

Table 3  
*Sum of Quality Indicator Ratings for Each Measure*

<u>Quality Indicator</u>	<u>Scale Range</u>	<u>Total Possible Score</u>	<u>Measures of ADLs</u>					<u>Measures of Socialization</u>	
			<u>ABLRS-R</u>	<u>AFLS</u>	<u>AMP S</u>	<u>DASH-3</u>	<u>GOAL</u>	<u>ESI</u>	<u>SP</u>
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

CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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
<b>Total Points (scored/ total possible points)</b>		<b>92</b>	<b>40/92</b>	<b>27/92</b>	<b>75/92</b>	<b>19/92</b>	<b>52/92</b>	<b>70/92</b>	<b>*37/92</b>
(n = 4 reviewers)									
<i>Note.</i> 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, 3 <sup>rd</sup> edition.									
ESI= Evaluation of Social Interaction.									
GOAL= Goal-Oriented Assessment of Life skills.									



## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

NA= Not Applicable.

SP= Social Profile.

CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

Table 4		
<i>Description of Included Measures and Quality Indicator Points Obtained</i>		
<u>Analysis of Measures of Daily Living Skills</u>		
Measure	QI score of 92 possible points	Strengths (S) and Limitations (L) of each measure for our planned trial
<i>Assessment of Motor and Process Skills</i> (AMPS; Fisher, 2006)	75	<b>S:</b> Standardized observation-based measuring performance in ADLs/IADLS for persons with developmental age 2+, natural environment, yields motor and process score <b>L:</b> N/A
<i>The Assessment of Basic Language and Learning Skills- Revised</i> (ABLLS-R; Partington, 2010)	40	<b>S:</b> 544 skills in 25 skill areas, good inter-rater reliability, scope, magnitude of change <b>L:</b> Criterion-referenced, administration time, intra-rater reliability, responsiveness, exposure to Rasch analysis
<i>Assessment of Functional Living Skills</i> (AFLS; Mueller & Partington, 2015)	27	<b>S:</b> 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

CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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

## CHOOSING PERFORMANCE-BASED OUTCOME MEASURES

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