Research in Autism Spectrum Disorders 23 (2016) 221–234

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

journal homepage: http://ees.elsevier.com/RASD/default.asp

A critical review of outcome measures used to evaluate the effectiveness of comprehensive, community based treatment for young children with ASD

Michael Stolte[a,*], Sandra Hodgetts[b], Veronica Smith[c]

[a] Centre for Autism Services Alberta, 4752-99 Street, Edmonton, Alberta T6E 5H5, Canada
[b] Department of Occupational Therapy, Faculty of Rehabilitation Medicine, University of Alberta, 8205 114 Street 2-64 Corbett Hall, Edmonton, Alberta T6G 2G4, Canada
[c] Department of Educational Psychology, Faculty of Education, 6-102 Education North, University of Alberta, Edmonton, Alberta T6G 2G5, Canada

ARTICLE INFO

Article history:
Received 15 April 2015
Received in revised form 19 December 2015
Accepted 20 December 2015
Available online 19 January 2016

Keywords:
Autism spectrum disorder
Measurement
Treatment
Preschool
Effectiveness

ABSTRACT

This review critically evaluates reporting and use of standardized measures to assess community based treatments for young children with Autism Spectrum Disorder (ASD). *The Standards for Educational and Psychological Testing* (AERPA, APA & NCME, 1999), a best practice framework for reporting standardized test results, guides the evaluation. Fifty three different outcome measures are identified across 45 studies representing twelve countries. Adaptive behavior, specifically the Vineland Adaptive Behavior Scales and cognitive measures continue to be primary outcome tools, despite a lack of clear fit to core ASD diagnostic constructs. Behavioral, ASD specific, language, social communication, and family wellness tools are under represented. Reporting strengths are use of multiple measures, clear sample descriptions, and use of specialized tools for ASD. Reporting weaknesses are assessment bias, test substitution, and under reporting of test modifications. Clinical and research implications are discussed.

Contents

* Corresponding author.
E-mail address: mstolte@centreforautism.ab.ca (M. Stolte).

http://dx.doi.org/10.1016/j.rasd.2015.12.009
1750-9467/

1. Introduction

Autism Spectrum Disorder (ASD) is diagnosed at younger ages and with increased frequency, with current estimates that approximately 1% of school aged children (Blumberg et al., 2012) meet the diagnostic criteria of qualitative impairment in communication and socialization skills, as well as the presence of repetitive behavioral mannerisms that interfere with daily functioning (American Psychiatric Association, 2013). As the number of children with ASD increases, so has public pressure to provide evidence based treatment. However, though evidence based treatments are established within research settings (e.g., Makrygianni & Reed, 2010) the gap between research and practice is large (Dingfelder & Mandell, 2011; Kasari & Smith, 2013) and despite significant costs associated with treatments in community settings, little is known about how well many treatments developed in research settings generalize into the community. For example, Amendah, Grosse, Peacock and Mandell (2012) estimate costs of $25,099 to $60,000 + per person, per year for behavioral therapies. In Canada, provincial governments are spending up to $40,000 per child, per year on therapies for children with ASD (Madore & Pare, 2006). Lifetime costs are even greater with recent estimates of $2.4 million in the United States and £1.5 million in the United Kingdom (Buescher, Cidav, Knapp, & Mandell, 2014).

The distinction between treatment *efficacy* and *effectiveness* has important implications for bridging research and clinical practice. Treatment *efficacy* is demonstrated through completion of replicable studies in highly controlled research settings, whereas treatment *effectiveness* is the demonstration of the generalizability of efficacious treatments into community settings (Greenberg, 2004). Our understanding of effective treatments for ASD has been consolidated through several extensive and systematic reviews (e.g., National Autism Center, 2009; Wong et al., 2014) yet, the evidence base for ASD treatment *effectiveness* is still an emerging field. In addition, predicting what intervention will work best for an individual child, the specificity of the intervention targets, and the individual responsiveness of ASD symptomology to treatment is still unknown, particularly as knowledge is generalized out of university contexts into the community (Minjarez, Williams, Mercier & Hardan, 2011). Moreover, in the context of implementation science, the lag time between the development of an efficacious practice and its eventual adoption is still estimated to be as high as 20 years (Walker, 2004).

One contributing factor to the slow adoption of efficacious practice is the lack of consensus on measurement tools and wide usage of different instruments (Bolte & Diehl, 2013). Matson (2007) reports that measures of intelligence and adaptive functioning are used most frequently, though a sole focus on these two constructs in the measurement of ASD treatment response can be problematic. For example, regarding cognition, Matson (2007) reports that (1) children often age out of Intelligence Quotient (IQ) measures from pre to post test, forcing substitution of a different IQ instrument normed on an older population, (2) it is not clear whether ASD intervention results in increased scores due to increased compliance, attention, motivation or ability, (3) IQ tests are less reliable at predicting future performance for children at young ages, (4) comorbid psychopathologies may interfere with measurement of the underlying constructs and, (5) IQ tests are not normed on an ASD population. Adaptive measures, though valuable, are normed primarily on the typical developing population (Sattler, 2006), not designed specifically for this and consequently only provide a reference point for identifying delays and strengths in the ASD population.

According to Gould, Dixon, Najdowski, Smith, and Tarbox (2011) measurement outcomes of intensive ASD programs should be: (1) comprehensive, (2) target early childhood development, (3) consider behavior function, (4) directly link assessment items to curricula targets, and (5) be used to track child progress over time. Gould et al. (2011) indicate that a combination of direct observation and indirect assessment (e.g., rating scales and checklists) is an ideal manner to track outcomes. However, after reviewing 27 different tools that may be used to measure ASD intervention progress, they were not able to identify any specific tool that met their five criteria. Four tools identified as being 'of promise' were the Verbal Behavior Milestones Assessment and Placement Program [VB MAPP] (Sundberg, 2008), the Brigance Diagnostic Inventory of Early Development II [Brigance IED II] (Brigance, 2004), the Vineland Adaptive Behavior Scales Second Edition [VABS II] (Sparrow, Cichchetti, & Balla, 2005), and the Brigance Diagnostic Comprehensive Inventory of Basic Skills Revised [CIBS R] (Brigance, 1999). The VABS was described as "by far the most popular assessment" (p. 998). To strengthen these tools, Gould et al. (2011) recommended simplified administration of the VB MAPP, increased psychometric evaluation of the VB MAPP and Brigance IED II, and content linking of the VABS and CIBS R more clearly to a curriculum.

Bolte and Diehl, 2013 reflected the lack of consensus on measurement tool selection to evaluate treatment response in their review of 195 prospective ASD treatment trials from 2001 to 2010. They identified 289 unique measurement tools, of which the vast majority (61.6%) were only used once. The top five utilized tools reported in this review were the Aberrant Behavior Checklist [5%] (Aman, Singh, Stewart, & Field, 1985), Clinical Global Impressions [4.6%] (Guy 1976), VABS [3.9%] (Sparrow, Balla, & Cicchetti, 1984), investigator designed video observations (1.9%), and the Bayley Scales of Infant Development [1.7%] (Bayley, 1993). Bolte and Diehl, 2013 concluded that "greater consistency in the use of measurement tools in ASD clinical trials is a worthwhile and achievable goal" (p. 2499) as the sheer number of tools and tool symptom