Current status of differential diagnosis for children with autism spectrum disorders

Johnny L. Matson*

Department of Psychology, Louisiana State University, Baton Rouge, LA 70803, United States

Received 11 May 2005; received in revised form 1 July 2005; accepted 19 July 2005

Abstract

Early intervention for autism spectrum disorder (ASD) has proven to be a successful strategy for remediating many difficulties experienced by these children. As a result, accurate diagnoses of children with this range of disorders has become more critical. Additionally, while current training programs are for 3–4 year olds, in efforts to start treatment at younger ages, clinicians are giving these diagnoses at younger and younger ages. A considerable amount of research activity on a technology for making differential diagnoses of ASD has been emerging in recent years. The purpose of this paper is to provide an overview of some of these developments, and to offer opinions on the current status of the area.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Differential diagnosis; ASD; Critical review

Early attempts to diagnose autism and later, a broader group of related developmental disabilities designated autism spectrum disorder (ASD) date to Kanner (1943). The primary focus initially was to describe symptoms that best characterize the condition through observation in a nonsystematic fashion. Typical of this and other early studies were reports of common behaviors observed in a child's repertoire including lack of eye contact, engagement in repetitive behaviors, delayed language development, and social deficiencies. More recently researchers have confirmed many of these early observations by demonstrating that language and social skills are key elements in differential diagnosis from autism and other developmental disabilities such as intellectual disabilities (ID) (Sevin et al., 1995). Nonetheless, the rather informal assessment method led to low reliability in diagnosis and dissatisfaction with this approach. This point is obvious since intensive efforts to develop tests for diagnostic purposes have been a core research area in ASD.
From these humble beginnings, the increased research focus on ASD classification has grown exponentially (Baird, Cass, & Slonims, 2003). Promising studies on interventions have emerged with particular focus on early childhood. It has been argued that an essential element in an underlying component of all phases of study is accurate and early differential diagnosis. After all, how can one study etiology or prescribe an early intervention program until the ASD person is identified. Experts have claimed that autism, the most common and studied of the ASD conditions, can be routinely identified by age 3 years (Howlin & Moore, 1997) and can be identified by 18 months in some severe cases (Baird et al., 2000; Baron-Cohen, Cox, Baird, Swettenham, & Nighingale, 1996).

ASD diagnosis involves clinical observation and questioning of the child. Where applicable, parents and other care givers are interviewed and/or asked to complete checklists. Perhaps more than for most disorders, the use of diagnostic tools for identifying ASD has become the central theme for differential diagnosis, often within the context of a multidisciplinary team (American Academy of Pediatrics, 2001).

A variety of these instruments have been developed. However, four dominant themes have emerged. First, a few measures have been developed extensively. Most notable of these are the Autism Diagnostic Interview-Revised (ADI-R) (LeCouteur et al., 1989; Lord, Rutter, & LeCouteur, 1994), Childhood Autism Rating Scale (CARS) (Schopler, Reichler, & Renner, 1988), and Autism Behavior Checklist (ABC) (Krug, Arick, & Almond, 1980). Second, because of initial successes with early intervention programs for children with ASD, a race has begun to determine the youngest age at which the disorder can be reliably diagnosed, primarily using checklists and to some degree videotape of verbal and nonverbal responses, sometimes in children as young as 6 months. Third, now that a broader range of disorders has been included in the ASD category, scales that measure Asperger’s, PDD and other variants of ASD are beginning to be developed. Psychometrically, the best scales to date in the ASD category however clearly focus almost exclusively on autism. Fourth, adults with ASD have received scant attention. It is unknown if the course of ASDs remain the same into adulthood, and for many reasons, not the least of which is the general lack of national screening for ASD, some people with these disorders go undetected into the teenage years or are misdiagnosed in adulthood.

1. Instruments for diagnosing the ASD spectrum

As noted, the bulk of the diagnostic research on ASD has been with autism which will thus be the primary topic of the review. The initial scales that were developed are still relatively recent in origin and focused on school age children as has traditionally been the case with most childhood problem disorders and the scales which are used to evaluate them (e.g. ADHD, conduct disorder). Pragmatic reasons dictate this situation. At 5–6 years of age, for example, children are required for the first time to sit for extended time periods, follow numerous rules and instructions, and interact with adults and many children with whom they are unfamiliar. These events are often stressful and can produce dysfunctional behavior at increased levels, if they exist in the child’s repertoire at all. Therefore, the most extensively studied scales to date primarily involve the 5–12 year old group.

2. CARS

The CARS was developed to assist in screening for autism in North Carolina’s statewide education program (TEACCH) (Schopler et al., 1988). The test, which consists of 15 independent
scales such as verbal communication and imitation involves a Likert rating of 1 = normal to 4 = severely abnormal and produces a range of total scores from 15 to 60. Initial psychometrics of the CARS were reported on 537 children enrolled in TEACCH over a 10 year period. Results on the reliability and validity of the scale were good and rapidly propelled the scale into national and international prominence as a means of detecting autism and establishing gradations of symptom severity among those children who evinced the condition (Matson, Smiroldo, & Hastings, 1998; Matson & Minshawi, 2006; Morgan, 1988; Sturmey & Sevin, 1994). The CARS has proven superior to the ABC and Diagnostic Checklist in distinguishing between autism and children with ID (Teal & Wiebe, 1986). This distinction is quite important because of the substantial percentage of autistic children with ID, and has implications for treatment.

The CARS will most likely need some revisions in the near future if it is to remain in common use. The scale does not conform to DSM-IV criteria nor does it provide for differential diagnosis of autism and other ASDs (Klinger & Renner, 2000). Of course, you could say this of most scaling methods in the area at present. Also, as will be discussed further, early diagnosis is becoming the dominant diagnostic model. Thus, any scale that is to maintain a central role in differential diagnosis of ASDs will need to take these factors into account. Normative data going down to at least 3 years of age appears to be the new standard.

3. ABC

The primary role of the ABC much like the CARS is classroom placement in educational settings. Items on this scale come from criteria described by Creak (1964), Kanner (1943) and Rimland (1964). Five subscales (sensory, social and self-help, language) and 57 Likert items are rated 1–4. The initial psychometrics of the scale were established in 1979 (Krug, Arick, & Almond, 1979) with 14 children and 42 independent raters. These data have been largely replicated (Volkmr et al., 1988; Walden, Bryson, & Rodgers, 1991). The scale, which is filled out by a teacher, has proven useful in discriminating between autism, ID, emotional disturbance and visual impairments. The scale is in need of updating however. The issues noted above as a rational for revising the CARS hold with the ABC as well. Additionally, measurement of more prosocial behaviors and revised cutoff scores may be advisable (Sevin, Matson, Coe, Fee, & Sevin, 1991).

4. ADI and ADI-R

The ADI appeared in the literature at about the same time as the CARS and ABC (LeCouteur et al., 1989). Three content areas, social interaction, communication and repetitive, stereotyped behavior are assessed with an interviewer asking for yes or no responses. The interviewer for this scale must be highly trained since the tester is also required to use probe questions which must be coded from 0 to 2. These scores are then converted into ICD-10 criteria using a scoring template developed for that purpose.

Two clinicians independently assessed 16 autistic and 16 ID children via mother interviews. Psychometrics were generally good. However, unlike the other two measures briefly reviewed, the authors did revise the scale in 1994 based on perceived shortcomings (Lord et al., 1994). The scale now dubbed the ADI-R was designed to differentiate autism from other ASDs including Rett’s Disorder and Childhood Disintegrative Disorder at earlier ages, and the format was shortened to save time. Several items for older persons with higher intellectual abilities including unusual interests were also added. The DSM-IV (APA, 1994) criteria were included. The
methodology for this study was similar to their previous research in that 10 children with autism and 10 with ID or language impairment were evaluated by clinicians via mother interview. The authors report good reliability and validity (Lord et al., 1994).

5. General comments

There are many other measures of autism or ASD, but these three are among the best if not the best at the time of this writing for diagnosing autism. Nonetheless, there are a considerable number of limitations to the scales. (What scale does not have limitations?) The CARS is the only measure where a large sample was used to establish the psychometrics properties of the measure, and even in that instance more independent replication of the scale is needed. However, human nature seems to be development of one’s own scale versus psychometric work on other measures such as the three mentioned here. This phenomenon may not be bad at present, however, since the field is so new and diagnoses are still being debated. Thus, further item and administration refinement are required.

An additional problem, which makes for considerable difficulty, is the rapid evolution of intervention strategies and attitudes about treatment. A broader range of ASDs are now being discussed and identified, and treatment of a very intensive sort is being proposed at very young ages, particularly for children with autism and PDD-NOS. Obviously, it is difficult to treat an afflicted group if reliable and valid means of identifying that group does not exist. The latest trend in the ASD assessment literature is to address changes in criteria, particular to assessing autism at very young ages and further developing Asperger’s diagnosis and assessment.

6. Early childhood diagnosis

With an eye toward early screening of children with ASD, biomarkers are being tested for differences in ASD and normal developing children on proteins, metabolites and the immune system such as elevated immune system B cells. Similarly, researchers hope that by mapping the genome, specific genetic patterns may be used to identify types of ASD. Some developmental psychologists are looking at voice differences, abnormalities in eye contact, body or limb posturing and atypical sounds and words in children as young as 3 months. In the long term, these efforts are likely to yield some useful results.

Major problems exist however, with these approaches. First, detection of genetic causes as with Down’s Syndrome, where such identification has occurred, does not factor out environment and the overall complexity of the Genome. Thus, it is possible to have a person with Down’s Syndrome with normal IQ or profound ID, although the mean is moderate ID. Ultimately, behavior based conditions must rely greatly on assessments of observable behavior and must be based on the person’s developmental course. Symptom patterns are not static and may become less or more serious over time.

Second, the studies at present use paper and pencil measures to classify the child somewhere between 4 and 6 years of age. For early screening, the child might be classified at 2 and then followed up 1–3 years later to determine the predictive validity of the diagnosis. Similarly, genetic and metabolic tests at present would assess children classified in the 3–4 year old range with a standardized measure (Cox et al., 1999; Lord, 1995). The problem in this methodology comes when no good anchor measure exists such as with Rett’s Disorder, or the definition has been changed as with autism. Thus, for example, the CARS and ADI-R use somewhat different criteria which would result in somewhat different group classification. Thus, even greater
attention to multiple psychometric studies with bigger samples and more replications across more laboratories are urgently needed if these early screening methods are to be adequately tested.

To date, most early screening in clinical practice does not occur prior to age 2 years and often is not conducted until age 6 years (Howlin & Moore, 1997). Having said that, researchers are attempting to push the envelope by developing scales aimed at identifying autistic children at younger and younger ages. Pragmatic/administrative issues which provide access for early assessment also need to be worked out.

7. CHAT

The Checklist for Autism in Toddlers (CHAT) differs from the more established scales previously discussed in that it is specifically aimed at the screening of very young children for autism. The measure was designed for parents and pediatricians, who are to administer the measure to 18-month olds. Primary areas of evaluation are categorized into pretent/imaginative play, protodeclarative pointing and gaze monitoring (Baron-Cohen, Allen, & Gillberg, 1992). Two sections of the CHAT have yes/no questions for the parent while the third section is for in-office observations of the pediatrician. The initial 1992 study was filled out on 50 children and 41 of their siblings. Four were identified as autistic. Diagnoses were confirmed at 30 months. The research group hoped to develop the CHAT as a screening measure, and the ensuing studies reflect that goal. Baron-Cohen et al. (1996) screened 16,235 English children at 18 months of age, identifying 10 possible autism cases. Baird et al. (2000) followed the sample for 6 years, finding a poor rate of predicting validity. Scambler, Rogers, and Wehner (2001) in an apparent recognition of this interpretation assessed 44 children with the CHAT, but waited until they were 2–3 years of age for the initial evaluation. Their predictive validity was similar to the ADOS-G (Matson, Nebel-Schwalm, & Matson, in press), ADI-R and DSM-IV for identifying autism in young children. However, this does not address the point of predicting autism in school age children. Additionally, the necessary reliability and validity research is yet to be done. Thus, the idea has promise, and the CHAT and other scales may have applicability in this regard. But, these are empirical questions requiring considerable additional investigation such as the optimal age for first assessment for establishing good predictive validity (e.g. one and a half, two and two and a half or 3 years of age).

8. STAT

A second attempt at very early identification of autism is described by Stone, Coonrod, and Ousley (2000) who developed the Screening Tool for Autism in Two-year-olds (STAT). Again taking the less ambitious route of older children, they targeted 2–3 year olds versus those 18 months of age or younger. They assessed 40 developmentally delayed children. Seven were diagnosed as autistic during the 20 min interaction between examiner and child based on binary responses to 12 items regarding play, imitation, requesting and attention directing skills. Clinicians were able to accurately categorize most cases, leading to a second study (Stone, Coonrod, Turner, & Pozdol, 2004) Reliability of the scale in both inter-rater and test–retest reliability was found to be high. These researchers also found high agreement for children diagnosed with and without autism compared to the ADOS-G. Thus, this scale appears to have promise as well. But, as with the CHAT, considerably more research is needed. It would appear then that at present, claims of accurately diagnosing very young children with autism not to mention other ASDs is premature.
9. Conclusion

CHAT and STAT data are encouraging but scant at present. However, these are the best options for very young children currently available. Scaling methods with young children assume a trajectory of autistic symptoms that intensify with age. However, more studies that look at stability or intensifying of symptoms with age appear to be needed. It is also possible that symptoms may wax and wane for some young children, or a normal course of development may occur until 2 or 3, followed by rapid onset of autistic symptoms. For cases such as those just noted, the most reliable and valid assessment method will still only result in mediocre predictive validity if given to children when they are very young. Having said this, one must admire Baron-Cohen and associates and Stone and colleagues for their efforts to take on what is a complex and difficult diagnostic issue.

10. Assessing other ASDs

Autism has a number of scales developed specifically for diagnosis. Several of these are reasonably well established psychometrically, and it would appear that additional efforts at scale development with existing and new measures are likely in the near future. However, PDD and Childhood Disintegrative Disorder have not been studied in much detail to date. Asperger’s Syndrome is now receiving a limited amount of attention with respect to scale development and will be reviewed next. Additionally, one scale to diagnose Rett’s Syndrome has been published.

11. Asperger’s Syndrome

The National Autism Society of England (1997), and Wing (1993), present data suggesting that Asperger’s Syndrome may be more common than autism. Despite this, the disorder has been rarely diagnosed until recently and in an excellent review by Howlin (2000), she points out that good diagnostic methods are at best in their infancy. Leekam, Libby, Wing, Gould, and Gillberg (2000) have posited an interesting notion which many clinicians have voiced over the years. Rather than try to differentiate Asperger’s Syndrome from persons who are autistic, it may be more appropriate to distinguish people along the ASD spectrum based on symptom severity. However, the best way to make this determination will be empirical. Mayes and Calhoun (2004) have been among the first to address this issue. They compared children with normal and below normal ID. Based on the data they concluded that DSM-IV’s use of normal cognition as one criteria for establishing separate diagnoses of Asperger’s Syndrome and autism is not valid. They assert that these differences can be explained by age and IQ alone.

Despite the above reservations, attempts have been made to develop assessment scales to evaluate Asperger’s Syndrome. Campbell (2005) reviewed five scales which have been developed for Asperger’s. These include the Asperger’s Syndrome Diagnostic Scale (ASDS) (Myles, Bock, & Simpson, 2001), the Autism Spectrum Screening Questionnaire (ASSQ) (Ehlers, Gillberg, & Wing, 1999), the Childhood Asperger Syndrome Test (CAST) (Scott, Baron-Cohen, Bolton, & Brayne, 2002), the Gilliam Asperger’s Disorder Scale (GADS) and the Krug Asperger’s Disorder Index (KADI).

Campbell’s (2005) conclusions were that “scales demonstrated significant weaknesses, particularly in normative data. He suggested that the KADI was the most empirically sound of the available instruments at this time. It is cautioned however, given the low prevalence of these disorders relative to the general population, that obtaining large scale normative data may be
difficult. What can be concluded at this point is that the diagnosis and thus assessment of Asperger’s Syndrome is in considerable flux at this point. Many debate whether it is a separate ASD or a “higher functioning group of autistic” people. The next few years should be instructive in determining if Asperger’s Syndrome continues to be diagnosed as a separate ASD or is folded into a dimension of ASD based on symptom severity. This point needs to be addressed before the notion of continued, separate Asperger’s scales are to be used.

12. Rett’s Syndrome

Rett’s Syndrome occurs in approximately 1 in 10,000 to 1 in 15,000 cases, primarily in females (Hagberg & Hagberg, 1997). Thus, it is a very low incidence condition which may explain in part the limited research focus on differential diagnosis. Mount, Charman, Hastings, Reilly, and Cass (2002), however have developed a scale which may be of considerable value in this regard. Many clinicians are likely to be unfamiliar with the disorder, and even if they have heard the name a scale which lays out the symptoms should prove to be a very important educational tool when confronted with such a case. Their study consisted of mailing packets to parents who belonged to a Rett’s support group. One hundred forty-three parents of girls responded to the Rett’s Syndrome Behavior Questionnaire (RSBQ). They compared these data to children with severe to profound mental retardation. The scale proved effective in differentiating between the groups with repetitive hand movements and breathing problems being particularly pronounced in the Rett’s versus mentally retarded group. More research is of course needed, but given the limited number of children to study, progress is likely to be much slower than what is being reported in the autism literature.

13. Multiple ASDs

Cohen, Schmidt-Lackner, Romanczyk, and Sudhalter (2003) have recently developed a scale covering PDD, autism, Asperger’s Syndrome, PDD-NOS and childhood disintegrative disorder. The test, which is called the PDD Behavior Inventory is filled out by caregivers or teachers. It’s primary aim is to track challenging behavior and improvements in social and language skills rather than as a differential diagnosis measure. The authors report good reliability. The test is for children of age 1.5 to 12.5 years. It can be completed in 30–45 min. Standardization of the PDD Behavior Inventory was conducted on 369 parents and 277 teachers of children from a range of racial/ethnic background and geographic areas in the U.S. The validity of this measure against other measures of autism have also been made. The biggest strength of this test is brief administration and larger norms than many other scales in the ASD category. Also, it measures a broader range of ASDs, but autism is still the primary focus. Finally, some challenging behaviors are also included. This latter issue is an important point to address given the many children with ASD who evince these behaviors.

14. Conclusions

Remarkable progress in quantifying diagnostic methods for persons with ASD has occurred in the last 20 years. Relative to many other areas of childhood development where disability, psychopathology or challenging behavior are concerned, the number of researchers involved and studies conducted is much greater. Clearly, a path has been established for using scaling methods based on parent/teacher/paraprofessional report, combined in some instances with structured
observational data provided by clinical psychologists, psychiatrists or pediatricians. Relatively sound scaling methods are available for autistic children 3 years of age and older. Some success in accurately diagnosing younger children with autism has occurred, but considerably more research in this area is needed and is likely to appear in the next decade. It is argued here that great diagnostic specificity for 18–48 month olds may not be as critical as some in the literature suggest for treatment. Being able to label some general symptoms of ASD/PDD may be sufficient at very young ages since the primary goal is early intervention, and the full spectrum of children noted may benefit greatly from such interventions.

Several major issues warrant considerable more attention. First, while ASD is less frequent than ADHD for example, better normative data in particular and more reliability and validity data are needed in established scaling methods. A typical pattern in child psychopathology in general is the publication of a few studies on reliability and validity with limited follow-up research on that scale. Then the next test appears. A few diligent investigators continue research on their scales over many years. These scales are the few that emerge from the many. Second, Asperger’s scales are beginning to appear although development is at a rudimentary stage at present. More is needed and for the other ASDs scale development is virtually nonexistent. Third, data driven research, better establishing if Asperger’s and autism are distinct disorders or the same/similar disorder along an IQ dysfunctional behavior continuum, are needed. Fourth, the general assumption is that the general trajectory of ASD symptoms identified in very young children is steady and progressive. However, idiosyncratic differences are likely. It is possible that some domains such as communication or social skills may be more markedly affected than developmental milestones such as walking and toileting. Problem areas may stabilize at certain ages. Clearly, more research on these and related issues of developmental course are needed. Fifth, co-morbidity is rarely addressed. It is argued that screening measures such as those for persons with intellectual disability which measure a broad range of DSM-IV, ICD-10 disorders exemplified by the ADD (Matson & Bamburg, 1998) or DASH-II (Matson, Coe, Gardner, & Souner, 1991) are needed. Covariation of other disorders (Matson et al., 1998) such as obsessive compulsive disorders, ADHD, anxiety and depression have important implications for assessment and treatment. Specialized measure such as the CARS, and ADI-R could then be used to follow up these screens as could measures such as the Child Depression Inventory where warranted (Helsel & Matson, 1984). Differential diagnosis research in this area is easy to criticize but hard to do. However, researchers in the area have shown considerable enthusiasm. Additional positive changes are likely to appear in short order.

References


