The Observation Scale for Autism (OSA): A New Screening Method to Detect Autism Spectrum Disorder before Age Three Years

Nils Haglund1,*, SvenOlof Dahlgren2, Karin Källén3, Peik Gustafsson1 and Maria Råstam1

1Department of Clinical Sciences Lund, Child and Adolescent Psychiatry, Lund University, Sweden
2Department of Psychology, University of Gothenburg, Sweden
3Department of Clinical Sciences Lund, Obstetrics and Gynecology, Center of Reproduction Epidemiology, University of Lund, Sweden

Abstract: Background: There is an increasing body of evidence that early interventions for children with Autism Spectrum Disorder (ASD) promote a positive development of social interaction. Thus, tools for early detection of ASD are warranted.

Aim: Development of, and deciding cut-off-levels for, a new screening tool for ASD, the Observation Scale for Autism (OSA). The OSA was designed to be used at the free health check-up at 30 months, offered to all children in Sweden.

Method: The OSA consists of 12 observations and takes less than 10 minutes to use. The performance of the test was investigated by assessing 37 children previously diagnosed ASD, 23 with Down Syndrome (DS) and 26 typically developing children (TD).

Results: Children diagnosed with ASD showed statistically significant higher scores in all 12 items compared to TD children, and significantly higher in 10 items compared to the children with DS. Most of the observations in OSA seemed to cover specific symptoms of ASD, but two of the observations were more related to developmental level. The nine most discriminative items for ASD were identified, and among those, a cut-off limit was chosen (≥3 items). Among children with ASD, 34/37 reached the proposed cut off, compared to 0/26 and 4/23 among children in the TD and DS groups, respectively.

Conclusion: The results suggest that the OSA discriminates children with ASD from TD children and children with DS. Using the suggested cut off, OSA provides high sensitivity for ASD (92%) with a very low false positive rate.

Keywords: Screening, Autism, sensitivity, Intellectual disability, typically developing children.

INTRODUCTION

Early detection of Autism Spectrum Disorders (ASD) has been an important field for research in autism, especially after the publication of studies showing that early intervention and treatment of ASD symptoms are associated with better outcome [1,2]. Different screening instruments have been developed for detection of ASD and there is a discussion of what would be the optimal age for finding ASD symptoms. Convergent data indicate that ASD symptoms emerge in the first two years of life [3-5]. In a recent review [6], it was concluded that both retrospective and prospective studies showed robust evidence that behavioral signs of ASD can be detected early in infancy.

In Sweden, all children are offered a free health check at 30 months of age. If there was a screening tool for ASD which was quick and easy to use for Child Health Center nurses, this would offer an opportunity to detect children with ASD in early childhood. Ideally, the screening tool would be an objective instrument, not dependent on parent’s propensities to recognize their children’s developmental shortcomings. The results from such a screening could then be used to select children for referral to neuro-psychiatric units for detailed examinations to establish (or reject) a diagnosis of autism.

There are few scales available for the assessment of symptoms of ASD in children 3 years and younger. Most of them are based on parent-reported questionnaires, either solely [7-13], or designed to be used in combination with an observation instrument [14,15]. To our knowledge, there are only two published instruments that are entirely based on an observation scheme [16-17]. Both are designed to be used by trained autism nurses, and thus, not suitable for ASD screening performed by Child Health Center nurses.

We conclude that although some appropriate screening tools for early detection of ASD have been developed and shown to have good psychometric properties, there is a need for brief, easy to handle assessment instruments designed for use in the primary health care system. Furthermore, it has been...
reported that existing screening tools depending on
parents’ observation abilities have unsatisfactory value
in discriminating between ASD and non-ASD within the
group of children showing abnormal development [18].

The purpose was to develop a screening observation scale for ASD to be used in primary care
30-months follow-up, to examine the performance of
the instrument on a small sample of children, and to
decide the suitable ASD cut-off limit before testing the
instrument on the population. Based on the small
sample, we also attempted to evaluate if the OSA could
discriminate children with ASD from TD children and
from children with a general developmental delay. The
latter group was represented by children with DS. We
also attempted to identify the most ASD discriminatory
observations (the most autism specific items).

METHODS

Participants

Three groups of children participated, children with
ASD (n=37), children with DS (n=23) and TD children
(n=26) (Table 1).

The ASD group consisted of 31 children diagnosed
with DSM-IV/ICD-10 autistic disorder/childhood autism
and six children with DSM-IV/ICD-10 PDD-
NOS/atypical autism. They had all been assessed at
age 28-36 months. During the diagnostic process all
children in the ASD group had been evaluated with the
Autism Diagnostic Observation Schedule [19] and the
Autism Diagnostic Interview- Revised [20].

The children with DS were diagnosed at birth. Age
at the assessment with the OSA was 31-55 months.

<table>
<thead>
<tr>
<th>Language group</th>
<th>ASD n=37</th>
<th>DS n=23</th>
<th>TD n=26</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N (%)</td>
<td>n (%)</td>
<td>n (%)</td>
</tr>
<tr>
<td>Boys</td>
<td>34 (92)</td>
<td>14 (61)</td>
<td>8 (31)</td>
</tr>
<tr>
<td>Girls</td>
<td>3 (8)</td>
<td>9 (39)</td>
<td>18 (69)</td>
</tr>
<tr>
<td>Swedish</td>
<td>27 (73,0)</td>
<td>15 (65,2)</td>
<td>23 (88,5)</td>
</tr>
<tr>
<td>Other Nordic languages</td>
<td>1 (2,7)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Other Western European languages</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (3,8)</td>
</tr>
<tr>
<td>Eastern European languages</td>
<td>1 (2,7)</td>
<td>1 (4,3)</td>
<td>1 (3,8)</td>
</tr>
<tr>
<td>Arabic</td>
<td>8 (21,6)</td>
<td>7 (30,4)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>African, Sub-Saharan languages</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Asian languages</td>
<td>0 (0)</td>
<td>0 (0)</td>
<td>1 (3,8)</td>
</tr>
</tbody>
</table>

ASD; Autism Spectrum Disorders, DS; Down Syndrome, TD; Typically Developing children.
contact between nurse and child during the assessment; reciprocal play between nurse and child; the child’s spontaneous use of two word phrases during assessment. The instruction for the observer was to determine whether the child behaved at each observation point as expected for a child with the developmental age of 30 months. If not, the observer would score one point.

Statistical Analysis

Pair-wise group comparisons (ASD vs TD, and ASD vs DS), respectively, were performed using Fisher’s exact tests. The comparisons regarding the age at assessment across groups was made by one-way-ANOVA with Bonferroni post-hoc test. The comparison in gender representation, OSA scores and the with-in group analyses were made with Fisher’s exact tests. The area under the Receiver Operating Characteristic (ROC) curve was used to estimate the overall ability of the OSA to detect autism among children at 2.5 years of age. In this analysis, children with DS were excluded. Statistical analyses were made using SPSS version 22.

Ethics

The study was approved by the Regional Ethics Committee in Lund (Dnr 2010/366, Dnr 2011/299). Parents in all three groups were given oral and written information and gave consent for participation of their child. All data were anonymised before being forwarded to the research team.

RESULTS

The basic characteristics of the three study groups are displayed in Table 1. In the ASD group there was a majority of boys participating (92%), and so was the case in the group of children with DS (64%). In the group of TD children there were a majority of girls (69%) (difference in gender distribution, Chi2(2)=25.4, p<.001). There were statistically significant differences across groups concerning age at assessment (F = 8.392, df = 2, p <.001). The mean age at assessment was 34.4 months in the ASD group, 39.7 months in the DS group, and 32.0 months in the group of TD children (p for difference ASD vs TD, p=.49; ASD vs DS, p=0.12; and DS vs TD, p<.001). The majority of all participating children had Swedish as native language (ASD group 73 %; DS group 65 %; TD group 89 % (difference across groups, Chi2(2)=3.81, p=.15). The second most common language in the ASD and DS groups was the Arabic language (Table 1).

Most of the children in the ASD group (92 %) had more than 4 scores on the 12-item OSA version compared to 17 % in the DS group, and 0 % in the TD group. In the ASD group 46 % of the children scored 8 or more on the OSA while none in the DS group scored more than 6 (Figure 1).

![Figure 1: Number of scores by study group.](image-url)

Overall the children with ASD had statistically significant higher scores on all observations than the typically developing children (Table 2). Compared to the children with DS the children with ASD had statistically significant higher scores on ten of the twelve observations. The only two observations that did not reach statistical significance were “adequate movements” and “two word sentences” (p=1.0 for both items).

The least frequent observations scored in the group with ASD were for “adequate eye contact” and “building blocks”. The most frequent item scored were “child pointing”, “pretend play”, “interaction with parents”, “two word sentences” and “adequate response to removal of toy” (Table 2). In the group with DS, observations related to developmental age were most frequently scored, i.e. “adequate movements” and “two word sentences”. The only observations where typically developing children scored were “adequate movements”, “two word sentences” and “adequate response to removal of toy”.

In order to find out if the specific OSA observations were independent of native language, comparisons were made within the study groups. No significant differences were found. In the ASD group, the p-values for difference between language groups for the 12 items varied between .12 and 1.0. The corresponding p-values in the DS and TD groups ranged from p = .13 to 1.0, and p=1.0, respectively (because of the low numbers of scores among children in the TD group only three items could be analysed). Concerning gender, there were no significant differences across groups (p-value ranges: ASD p = .12 to 1.0; DS p = .13 to 1.0; TD p = .30 to .53 (only three observations analysed due to no variation).

In order to increase specificity, the nine most discriminative observations were selected (the 12 original, excluding the observations regarding adequate movements, building blocks, and two-word-sentences).

<table>
<thead>
<tr>
<th>Items</th>
<th>ASD n=37</th>
<th>DS n=23</th>
<th>TD n=26</th>
<th>ASD vs DS</th>
<th>ASD vs TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name recognition</td>
<td>22</td>
<td>1</td>
<td>0</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Adequate response to removal of toy</td>
<td>29</td>
<td>4</td>
<td>1</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Adequate movements</td>
<td>23</td>
<td>14</td>
<td>2</td>
<td>1.0</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Interplay with parents</td>
<td>31</td>
<td>6</td>
<td>0</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Adequate eye contact</td>
<td>15</td>
<td>0</td>
<td>0</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Following pointing direction</td>
<td>27</td>
<td>3</td>
<td>0</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Pretend play</td>
<td>31</td>
<td>2</td>
<td>0</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Child pointing</td>
<td>34</td>
<td>9</td>
<td>0</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Building blocks</td>
<td>18</td>
<td>4</td>
<td>0</td>
<td>.026</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Kicking ball</td>
<td>26</td>
<td>1</td>
<td>0</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Two-word-sentences</td>
<td>29</td>
<td>18</td>
<td>2</td>
<td>1.0</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Waves good-bye</td>
<td>25</td>
<td>2</td>
<td>0</td>
<td>&lt;.001</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

ASD; Autism Spectrum Disorders, DS; Down Syndrome, TD; Typically Developing children.

Table 3 shows the number of scores that each child in the respective study group obtained. Among the 37 children in the ASD group, 34 (92%) scored for three observations or more. The corresponding percentage among children in the DS and TD groups were 17% (3/23) and 0% (0/26), respectively. Thus, using three scores as a cut-off, the sensitivity for the OSA would be 92%, and the false positive rate (1-specificity) among TD children would be 0%. The corresponding area under the ROC-curve, showing the over-all ability of OSA to detect autism, was 0.998 (95%CI: 0.994-1.000).

Sub-analyses were performed, restricted to children who were between 28 and 40 months at assessment (n=31, n=10, n=20 for the ASD, DS, and TD groups, respectively). Within this sub-cohort, there was no age-distribution difference between the study groups (F = 0.361, df = 2, p=.699). In the ASD group, 29/31 children (94%) scored for three observations or more. The corresponding percentage among children in the
DS and TD groups were 30% (3/10) and 0% (0/20), respectively, both significantly different from the rate in the ASD group (p<.001 for both comparisons).

**DISCUSSION**

The OSA was designed for the detection of suspected ASD symptoms in children under the age three years. The scale was specially developed to minimize the influence of different culture and language background, and to be used within the primary child health care program to select children who would benefit from referral to child neuro-psychiatric centers. The children with ASD scored statistically significant higher on ten out of the twelve observation items compared with children with DS, and on all twelve observations compared to the TD group. Using the 9 most discriminative observations, with a cut off of 3 scores, the sensitivity in the ASD group reached 92%, without any false positive test results among TD children. Using the proposed cut off, the false positive rate among children with DS was also low (17%). Although the sample was small, the results indicate that OSA managed to discriminate children with ASD from both TD children and children with DS (p<.001 for a positive test result, ASD group vs TD group, and ASD group vs DS group, respectively). Three children in the ASD group did not reach the proposed cut-off score. Two of these children had originally been diagnosed with PDD-NOS and had few symptoms of ASD according to t ADI-R and ADOS in the diagnostic assessment. The other 5 children diagnosed with PDD-NOS did not, according to our assessment with OSA, differ from children diagnosed with autistic disorder regarding to total score.

The nine most discriminatory observations were all illustrating skills of social reciprocity. The two observations that did not reach statistical significance between children with ASD and children with DS were both depending on general developmental level (“two word sentences” and “adequate movements”). The two observations that were most frequent scored in the TD group were among the youngest children in the group (24-30 months). These observations were also related to general development (“adequate movements” and “two word sentences” i.e. the same most frequent observations as in the group of children with DS). These findings taken together implies that the nine most discriminative observation situations “kicking ball”, “pretend play”, “adequate response to removal of toy”, “waves goodbye”, “name recognition”, “following pointing direction”, “child pointing”, “interplay with parents” and “adequate eye contact” illustrates ASD symptoms or at least early signs of ASD. These findings are also consistent with findings from Barbaro & Dissanayake [22], evaluating ASD markers at 24 months of age.

### Table 3: Number of Scores among the 9 Most Discriminatory Variables, by Study Group

| Number (i) of scores | Study group | Comparisons
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ASD n=37</td>
<td>DS n=23</td>
</tr>
<tr>
<td></td>
<td>Children with (i) scores</td>
<td>Children with (i) scores or more</td>
</tr>
<tr>
<td>0</td>
<td>n 0</td>
<td>37 100</td>
</tr>
<tr>
<td>1</td>
<td>n 3</td>
<td>37 100</td>
</tr>
<tr>
<td>2</td>
<td>n 0</td>
<td>34 92</td>
</tr>
<tr>
<td>3</td>
<td>n 2</td>
<td>34 92</td>
</tr>
<tr>
<td>4</td>
<td>n 3</td>
<td>32 86</td>
</tr>
<tr>
<td>5</td>
<td>n 2</td>
<td>29 78</td>
</tr>
<tr>
<td>6</td>
<td>n 4</td>
<td>25 73</td>
</tr>
<tr>
<td>7</td>
<td>n 6</td>
<td>19 62</td>
</tr>
<tr>
<td>8</td>
<td>n 10</td>
<td>9 46</td>
</tr>
<tr>
<td>9</td>
<td>n 7</td>
<td>7 19</td>
</tr>
</tbody>
</table>

ASD; Autism Spectrum Disorders, DS; Down Syndrome, TD; Typically Developing children.

Alles' kicking ball, pretend play, adequate response to removal of toy, waves goodbye, name recognition, following pointing direction, child pointing, interplay with parents, adequate eye contact.

Number of children with at least (i) scores versus children with less than (i) scores.
They also found “follow simple commands” and “loss of skills” to be important markers for suspected ASD. To “follow simple commands” is indirectly included in the OSA. In the situation where the children point at different objects they do it on instructions from the observer. However “loss of skills” cannot be observed during the assessment and is therefore not included in the OSA. The children with ASD in the current study had proportionally low score in “eye contact” (40.5%), compared to Barbaro & Dissanayake [22], who found 86% of the children with ASD had deficits in eye contact. This behavior is known to be a core symptom and a key marker for ASD in the literature of autism [23, 24]. However, even if the children with ASD in the current study had a proportionally low score for “eye contact”, they still had a statistically significant higher score than the other two study groups, and “eye contact” was also among the nine most discriminatory observations.

The least discriminatory variables in our study were “building blocks”, “inadequate movements”, “two word sentences”, all variables related to developmental level. Accordingly, these observations did not discriminate children with ASD from children with DS (with the exception of building blocks, \( p = .026 \)), but from group of TD children. This finding is in line with Mitchell et al. [25], who found that early markers including social, communicative and motor deficits distinguished ASD from children with developmental delay. Developmental delay included children with developmental language delay, children with Down syndrome, children with intellectual disability as well as children with other forms of broad delay without any specific diagnostic label. They also found that early signs of ASD overlapped with other types of developmental delay which complicate differential diagnosis at an early age. This could also be found in the present study were children with DS did score on the nine most discriminatory observations even if it was at a low frequency. Since children with ASD scored statistically significant on all items compared to the typically developing children, and on ten of the items compared to the children with Down syndrome, it seems that the OSA contains suitable items to detect children with ASD.

It was also found that the assessment with OSA did not differ depending on gender or native language, the last not least important since it was constructed to bridge language and cultural differences. Cultural differences regarding the parental awareness of traits of autism in their offspring was indicated in a study by Haglund & Källén [26] in which children of Swedish origin were much more likely to be diagnosed with Asperger syndrome than were children of immigrants.

In a review, Norris & Lecavalier [27] examined five caregiver completed rating scales for screening for ASD. The scales reviewed were: the Social Communication Questionnaire (SCQ) [9], Gilliam Autism Rating Scale/Gilliam Autism Rating Scale-Second Edition (GARS/GARS-2) [14], Social Responsiveness Scale (SRS) [11], Autism Spectrum Screening Questionnaire (ASSQ) [28], and Asperger Syndrome Diagnostic Scale (ASDS) [29]. Only one of these, the SCQ [9], performed well in screening for ASD at age 3 years and older. This review indicated that caregiver- completed rating scales needs much more scientific research to give diagnostic validity.

One of the strengths of the OSA is that the instrument is not depending on the subjective observations by the caregivers, but by trained primary care nurses who perform health check-ups on hundreds of children each year. The OSA should not be mistaken for AOSI (Autism Observation Scale for Infants, 18-items observational measure) [17], or the STAT (Screening Tool for Autism in Toddlers & Young Children, 12-item observations) [1], which both are similar to OSA with the important difference that they are designed to be used by trained autism nurses.

**LIMITATIONS**

One obvious limitation in our study was in the recruitment of the comparison groups. The children in the three study groups were not ideally matched for gender and age. However, since the results remained virtually unchanged when the analyses were restricted to children with similar age, the different age distributions among the study groups do not seem to have introduced any major bias. Another putative major limitation was the fact that the assessments of the participants in the DS group and the ASD group were conducted by psychologists, but by a preschool teacher in the TD group. Even though the preschool teacher had received a short introduction course in autism and OSA, it could not be precluded that the interpretation of the OSA test might differ between the evaluators. Also, the test was not blinded as all examiners were aware of the children’s diagnosis (or no diagnosis). Furthermore, we did not have permission to perform more than one test for each child. Thus, it has not been possible to perform any test – retest- or inter observer agreement analyses. Despite these limitations, the results suggest
that the OSA could be useful in screening for ASD in small children.

Children with DS have a special syndrome that includes cognitive and learning disabilities. This group was used as a proxy for children diagnosed with intellectual disability. Individuals with DS are known to have a general cognitive delay and to be intellectually disabled to some degree [30]. The identification of ASD in children with DS is complicated since the pattern of test scores across subscales among children with DS is not fully explored [31]. One of our aims was to explore if the OSA could discriminate autism from general developmental delay. Children diagnosed with intellectual disability at the age of 2.5 years almost always suffer from multiple disabilities, including autism. In the absence of a suitable control group diagnosed with developmental delay, we chose a group which we know have cognitive impairment, usually without autistic traits. We are, however, aware that children with DS do not represent the total population of children diagnosed with intellectual disabilities.

The results imply that OSA can be used with children with different cultural and language background as well as independent of gender. However, the statistical analyses were performed on very small groups so the results must be regarded with caution.

**CLINICAL IMPLICATIONS**

It is essential to develop screening instruments for the early detection of ASD to make sufficient support, services and treatment available for children with ASD and their families. It is of uttermost importance to note that the proposed screening tool is not intended to be a stand-alone instrument. The OSA could select children suitable for referral to child neuropsychiatric clinics, but the final diagnoses of autism should only be made at diagnostic units using more thorough diagnostic tools, such as e.g. ADOS and/or ADI. The OSA is designed to be a powerful first link in a care chain, aiming at an as early as possible diagnosis of autism so that interventions could be initiated without delay. An ongoing screening study will use the OSA for all 30-months old children coming for their standard check-up at the CHCs in Malmö to further examine the ability of this screening tool to successfully detect ASD traits in children of all cultural and language background.

**FUNDING**

This research was supported by ALF Foundation, Region Skåne and the Lindhaga Foundation.

**CONFLICT OF INTEREST STATEMENT**

All five authors have reported no conflict of interest.

**REFERENCES**


