Do parental interviews for ASD converge with clinical diagnoses?

An empirical comparison of the 3di and the DISCO in children with ASD, a clinically-referred group, and typically developing children

Running head: Comparing 3di and DISCO-11

Evers, Kris^{1,2,3}, Debbaut, Edward^{2,3,4}, Maljaars, Jarymke^{1,2,3}, Steyaert, Jean^{2,3,4}, Noens, Ilse^{1,2}

¹ Parenting and Special Education Research Unit, KU Leuven, Belgium

² Leuven Autism Research (LAuRes), KU Leuven, Belgium

³ Child and Adolescent Psychiatry, Z.org UPC KU Leuven, Belgium

⁴ Center for Developmental Psychiatry, KU Leuven, Belgium

Correspondence concerning this article should be addressed to Kris Evers, Parenting and Special Education Research Unit, KU Leuven, Leopold Vanderkelenstraat 32, P.O. Box 3765, 3000 Leuven, Belgium. Tel: +32 16326018, E-mail: kris.evers@kuleuven.be.

ABSTRACT

Two semi-structured parental interviews are available with algorithms developed to measure DSM-5 criteria of ASD, namely the Developmental, Dimensional and Diagnostic Interview (3di) and the Diagnostic Interview for Social and Communication Disorders (DISCO-11). The main aim of this study was to examine the agreement between classification according to both interviews, and their convergence with the clinical diagnosis. Therefore, the 3di and DISCO-11 were administered from three groups of parents of a 4-18 year old. Results showed 75% agreement between both instruments, but in the ASD group only 16% of the children scored above threshold on both instruments. Exploratory analyses suggested that the 3di failed to detect rigid and repetitive behaviors, whereas the DISCO-11 was insufficiently sensitive in detecting socio-communicative problems.

KEYWORDS

Autism Spectrum Disorder; Classification; Diagnostic Assessment; Interview; Instrument

ACKNOWLEDGEMENTS

The authors are extremely grateful to all participating parents and children and to the master students who assisted with data collection. They also want to thank their colleagues at the Department of Child and Adolescent Psychiatry (Z.Org - UPC KU Leuven) for the collaboration. This research was funded by a postdoctoral fellowship awarded to KE (Fonds Wetenschappelijk Onderzoek Vlaanderen, 12L6916N) and the research by JM and ED is funded by the KU Leuven (Leuven Universiteitsfonds 'Opening the Future').

AUTHOR CONTRIBUTIONS

KE, JM, JS, and IN conceived this study and designed the analysis. Data collection was coordinated by KE. KE and ED controlled the clinical files of all participants (from the clinical groups). Data-analysis was performed by KE. The writing of the paper was done by KE, ED, JM, JS, and IN.

CONFLICTS OF INTEREST

Some of the authors (JM, IN) contributed to the development of the DSM-5 algorithm of the DISCO-11. KE, JM, and IN are involved in the DISCO-11 training for researchers.

Do parental interviews for ASD converge with clinical diagnoses?

An empirical comparison of the 3di and the DISCO in children with ASD, a clinically-referred group, and typically developing children

The diagnostic assessment of Autism Spectrum Disorder (ASD) in children is a complex information collection process that should entail first-hand observations by an experienced clinician, but also requires information about the developmental history and current functioning of the individual (Constantino & Charman, 2016; Falkmer, Anderson, Falkmer, & Horlin, 2013; Gardner, Erkfritz-Gay, Campbell, Bradley, & Murphy, 2016). Semi-structured parental interviews are commonly used to gather information about development and current daily-life functioning in a standardized manner (Ashwood, Buitelaar, Murphy, Spooren, & Charman, 2015).

In 2013, the latest version of the DSM has been published, with significant changes related to the diagnostic criteria for ASD (American Psychiatric Association, 2013). In addition to the disappearance of the different sub-classifications in favor of an overarching classification of ASD, DSM-5 now requires individuals to meet criteria on two instead of three core domains of impairment, namely (1) deficits in social interaction and social communication (criterion A), and (2) repetitive and restricted patterns of activity, behaviors and interests (RRBIs, criterion B). In order to classify for a DSM-5 diagnosis of ASD, the individual's behavior should meet all three sub-criteria under criterion A, and two out of four sub-criteria under criterion B. Because of this rule, problems in social interaction and social communication are formulated more strictly in DSM-5: non-verbal communication problems (under criterion A) have to be present to get a DSM-5 diagnosis, which was not a requirement to obtain a DSM-IV-TR pervasive developmental disorder diagnosis (American Psychiatric Association, 2000). In addition, RRBIs became more

important in DSM-5 compared to DSM-IV-TR, and sensory problems formed a new sub-criterion under the RRBIs. Language problems were removed from the core ASD characteristics. Related to the onset of symptoms, DSM-5 is less strict compared to DSM-IV-TR: whereas DSM-IV-TR required an onset of problems before the age of three, the DSM-5 criterion C stipulates that symptoms must be present in the ‘early developmental period’ (but also explicitly mentions that symptoms may only become fully manifest later in life, or may be masked by learned strategies later in life).

Authors have suggested that such a change in diagnostic criteria could have a significant impact on the characterization of ASD (Bennett & Goodall, 2016; Kulage, Smaldone, & Cohn, 2014; Sturmey & Dalfern, 2014), and therefore some of the screening instruments and diagnostic assessment instruments have implemented these changes in their algorithms. To the knowledge of the authors, only two parental interviews are available with specially developed algorithms to measure DSM-5 criteria of ASD, namely the Developmental, Dimensional and Diagnostic Interview (3di; Skuse et al., 2004) and the Diagnostic Interview for Social and Communication Disorders (DISCO-11; Wing, 2006).

Both the 3di and the DISCO-11 are parental (or caregiver) interviews that collect information on core ASD symptoms and on other domains of functioning (such as e.g., motor skills, behavior problems or school functioning), and they also provide information on the developmental history. The 3di is a computerized interview that was originally developed to study ASD characteristics in children and adolescents with an (above) average intelligence, but the interview can also be used in individuals with a mild to moderate intellectual disability (Skuse et al., 2004), and recently, a version for adults has been developed (Mandy, Clarke, et al., 2017). The DISCO-11 is semi-structured interview that can be applied for individuals from all ages and levels

of functioning (Wing, 2006). The 3di and the DISCO-11 are intended to measure ASD symptomatology in a dimensional way, and for both, DSM-5 algorithms have been developed that determine diagnostic classification ('ASD' or 'no ASD') based on the instrument. The algorithms of the 3di and the DISCO-11 aim to cover all ASD sub-criteria, but they differ with regards to the selection and the nature of the items included in the algorithms and with regards to how thresholds were set.

The DSM-5 algorithm of the DISCO-11 was developed by including the items with the highest content validity, such that the algorithm covered all exemplars under the different sub-criteria for ASD (Kent et al., 2013). The DSM-5 algorithm of the 3di, however, started from the existing sets of items ('subscales') of the DSM-IV-TR algorithm of the instrument, and then complemented those subscales with additional items (based on consensus among the authors), such that the total set of items included in the newly developed algorithm covered the broad range of criteria mentioned in DSM-5 (Mandy, Wang, Lee, & Skuse, 2017). These item selection processes resulted in a set of 63 items that were included in the 3di algorithm and 85 items that were included in the DISCO-11 algorithm. A recent study suggested that a number of the 3di and DISCO-11 algorithm items could be assigned to different sub-criteria than the ones they were categorized under by the instruments. In that study, an expert panel has compared the items of the DSM-5 algorithms of the 3di and the DISCO-11 to the DSM-5 sub-criteria, and showed that 14-16% of the items should be organized under another sub-criterion compared to the one they were categorized under by the instruments themselves (Evers et al., under review). The 3di and the DISCO-11 included an equal number of items on 'Deficits in social interactions and social communication', but the DISCO-11 included substantially more RRBI items than the 3di (Evers et al., under review). Although both instruments covered the broad range of ASD sub-criteria, the

3di's algorithm tended to focus on some exemplars under specific sub-criteria, whereas the algorithm of the DISCO-11 covered a wider range of exemplars with items that were more broadly formulated, e.g., 3di: 'nods to agree', 'shakes head to disagree', DISCO-11: 'lack of instrumental gestures'.

In addition to variability in how the DSM-5 sub-criteria were covered by both interviews, the decision-making rules of both instruments' algorithms were developed differently, and the way thresholds were set by the 3di and DISCO-11 was also different. To get a diagnostic classification based on the 3di, individuals have to meet threshold on criterion A and criterion B. These thresholds were set based on consensus among the authors of the instrument, and not on statistical analyses (Mandy, Wang, et al., 2017). The DISCO-11 algorithm, however, requires individuals to score above threshold on criterion A and B, but also on criterion C, which stipulates that impairments have to have an early onset. The thresholds of the DISCO's algorithm were based on ROC curve analyses. Hence, the DISCO-11 classification procedure appears to be more conservative, in the sense that individuals have to meet three thresholds (on criterion A, criterion B, and criterion C), compared to two thresholds in the 3di.

The 3di and the DISCO-11 are parental semi-structured interviews with recently developed algorithms that were specially designed to measure DSM-5 characteristics of ASD (Kent et al., 2013; Mandy, Wang, et al., 2017), but classifications based on both interviews have not been empirically compared yet. Therefore, the main goal of this study was to compare the 3di and the DISCO-11, by administering both interviews in three groups of parents of children aged 4-18 years without intellectual disability: (1) an ASD group, consisting of children with a clinical diagnosis of ASD, (2) a clinically referred group, consisting of children referred for an ADHD assessment, and (3) a typically developing (TD) group of children. To establish this goal, we determined the

agreement between the classifications based on both diagnostic instruments. To increase insight in the underlying reasons of disagreement, we examined the convergence between both diagnostic instruments and the clinical diagnosis, and we evaluated which thresholds participants did or did not meet.

METHODS

Participants

Parents of three groups of children were interviewed in this study: (1) an ASD group ($n = 25$; 20 boys, 5 girls), (2) a clinically-referred group (CLINREF; $n = 24$; 18 boys, 6 girls), and (3) a typically developing (TD) group ($n = 19$; 13 boys, 6 girls). All parents were Dutch-speaking, and mostly mothers were engaged in the study (69%). Parents in the ASD group and the clinically-referred group did not differ in age, $t(40) = .06$, $p = .95$ ($M_{ASD} = 41.05$ years, $M_{CLINREF} = 40.90$ years), and most of them were highly educated (UNESCO Institute for Statistics, 2012; ASD: 62%; CLINREF: 67%) with parental education not being different between both groups, $\chi^2(1) = .11$, $p = .74$.

All parents had a child without intellectual disability aged 4 to 18 years. The mean age of the children in the three participant groups was not significantly different, $F(2, 65) = .87$, $p = .45$. There was no significant difference in the gender-ratio of the three groups, $\chi^2(2) = .77$, $p = .68$. For more details about our participants, see Table 1.

[INSERT TABLE 1]

The ASD group was recruited via the Autism Expertise Centre of our university hospital. These children were diagnosed by a multidisciplinary team (including a child psychiatrist, a

clinical psychologist, and a speech therapist), according to DSM-criteria (American Psychiatric Association, 2000, 2013). The clinically-referred group was referred for a multidisciplinary diagnostic ADHD assessment by a child psychiatrist and a clinical psychologist, but not all of them received a clinical (ADHD) diagnosis: 15 children (63%) received a clinical diagnosis of ADHD, 8 children (33%) did not get a child psychiatric diagnosis (yet), and 1 child (5%) received a another diagnosis (namely developmental language disorder). None of them was at suspicion of ASD based on the clinical judgement of the multidisciplinary team, and 22 of these children scored below the cutoff of 15 on the Social Communication Questionnaire (Rutter, Bailey, & Lord, Catherine, 2003), which is a screening instrument for ASD ($M = 7.82$, $SD = 5.13$).

To check the clinical diagnosis of participants in the ASD group and the clinically-referred group, the written multidisciplinary clinical reports of these 49 participants were examined by two raters in the context of this research project. These raters compared the clinically reported behaviors to the DSM-5 sub-criteria of ASD, and a research diagnosis was only given when all required symptoms were present (American Psychiatric Association, 2013). Diagnostic formulations were excluded from the anonymized clinical reports. The two raters, a child psychiatrist (ED) and a clinical psychologist (KE), are experienced in the diagnostic assessment of children with ASD, and they both evaluated the same 26% of the clinical reports, resulting in 100% agreement about the research diagnostic status ('ASD' or 'no ASD'). Given this excellent inter-rater agreement, the remaining 36 clinical reports were coded by either of both reviewers. Twenty-one of the 25 participants with a clinical diagnosis of ASD (84%) were also given a research diagnosis. The reports of four remaining children with a clinical diagnosis of ASD did not mention all criteria of ASD: for two participants, problems on criterion A (Deficits in social interaction and social communication) were insufficiently described in the clinical report, and for

another two participants, reports did not mention sufficient problems on criterion B (Restricted, repetitive behaviors, interests and activities). All analyses were repeated after excluding the four participants without a research diagnosis of ASD, in spite of their clinical ASD diagnosis. These results are reported in Supplementary Material, and conclusions remained highly similar. In the Results section, data on the full ASD group ($n = 25$) will be reported. None of the participants of the clinically-referred group were given a research diagnosis of ASD based on their report, but two reports mentioned problems on criterion B.

In the ASD group and in the clinically-referred group, intelligence was assessed during the clinical diagnostic trajectory, using a range of different instruments: the Wechsler Non Verbal Scale of Ability (Wechsler & Naglieri, 2006), the Wechsler Intelligence Scales for Children (Hendriks & Ruiter, 2017; Wechsler, 1992), or a cross-battery combination following the Cattell-Horn-Carroll theory (Schneider & McGrew, 2012). Given this heterogeneous set of measurements, best-estimate intelligence scores will be reported, based on the available total intelligence scores. The best-estimate intelligence scores of the ASD and the clinically referred group were not significantly different, $t(47) = 1.12$, $p = .27$. No intelligence assessments were available for the TD group.

The TD group was recruited via advertisements and the personal social network of the research team and via advertisements in mainstream schools. None of these children had a psychiatric diagnosis or a developmental disorder (based on written parent report, no formal questionnaire was used). Children with a first-degree family member with ASD were not included.

Instruments

Developmental, Dimensional and Diagnostic Interview (3di). The 3di (Skuse et al., 2004) is a computerized interview that is a hybrid in-between a fully structured and semi-structured

interview. Although originally developed to study ASD characteristics in children and adolescents with an (above) average intelligence, the interview can also be used in individuals with a mild to moderate intellectual disability (Skuse et al., 2004), and recently, a version for adults has been developed (Mandy, Clarke, et al., 2017). Prior to the interview, the examiner imputes some basic identifying information, which tailors the wording of questions that are asked during the interview via a laptop. The 3di consists of more than 800 questions, but interviewers can select modules that consist of subsets of questions. In the 3di, the majority of questions concerning atypical behaviors are coded on a 3-point severity scale: 0 (behavior is not present), 1 (minimal evidence of described behavior), and 2 (definite or persistent evidence of described behavior). The 3di assumes the interviewer to rate whether behavior “ever” occurred.

The original version of the 3di demonstrated good psychometric properties (Skuse et al., 2004): inter-rater reliability and test-retest reliability were excellent (intra-class correlation coefficients $> .90$), as were sensitivity (1) and specificity (> 0.97). The abbreviated version of the interview was developed later, and has also shown good psychometric properties (internal consistency: $\alpha > .81$; sensitivity $> .90$; specificity $> .85$; Santosh et al., 2009). More recently, different sets of items were put forward to be included in the DSM-5 algorithm of the 3di (Mandy, Wang, et al., 2017; Slappendel et al., 2016), but only one of those DSM-5 algorithms has been used in a peer-reviewed publication (Mandy, Wang, et al., 2017). The psychometric properties of this newly developed DSM-5 algorithm have not been established yet, but classifications based on a preliminary version of the algorithm were compared to ADOS-2 classifications, and yielded a sensitivity of .84 and a specificity of .54 (Slappendel et al., 2016).

In the context of the study, a module focusing on ASD characteristics was administered (total duration: 60-80 minutes), including the 63 questions that are part of the recently published DSM-5 algorithm (Mandy, Wang, et al., 2017).

Diagnostic Interview for Social and Communication Disorders (DISCO-11). The DISCO-11 is a semi-structured parental (or caregiver) interview that consists of more than 300 questions in total and that can be used in individuals from all ages and levels of functioning (Wing, 2006). For most of the questions, the interviewer has to code atypical behaviors on a three-point severity scale: 0 (marked problem), 1 (minor problem), and 2 (no problem). The DISCO-11 then distinguishes both ‘ever’ and ‘current’ ratings of the individual’s behavior.

Psychometric research showed moderate to excellent inter-rater reliability (82% of the items has a kappa > .75; Leekam, Libby, Wing, Gould, & Taylor, 2002; Wing, Leekam, Libby, Gould, & Larcombe, 2002). Later research confirmed that 90% of the DISCO items has a good to excellent inter-rater reliability, and established a sensitivity of .96 and a specificity of .79 (Maljaars, Noens, Scholte, & van Berckelaer-Onnes, 2012). More recently, the DSM-5 algorithm has been put forward, with a sensitivity ranging from .88 to 1 and a specificity ranging from .72 to .89 (depending on the evaluated participant group; Kent et al., 2013). The abbreviated version of the interviews including 54 items, also demonstrated good psychometric properties (Carrington et al., 2014), and a version of 14 items was recently developed as screening measure (Carrington et al., 2015).

In the context of this study, the full version of DISCO-11 was administered (total duration: 120-180 minutes), and the DSM-5 algorithm was used to determine diagnostic status according to the instrument (Kent et al., 2013).

Procedure

Both interviews (3di and DISCO-11) were administered from all parents, with on average 11.44 months in-between two interviews ($SD = 7.4$). For both clinical groups, the first interview took place during the course of the clinical diagnostic assessment. The order of the interviews was randomized: for 29 parents (44%), the DISCO-11 was administered first, and for 39 parents (57%), the 3di was the first interview. Interviewers were trained clinical psychologists, child psychiatrists or master students clinical psychology. Parents and children older than 16 gave written informed consent, and participants received millage allowance (if applicable) and a small box of chocolates. This study was approved by the KU Leuven Societal and Public Ethics Committee.

Data Analysis

Agreement between both diagnostic instruments was calculated as percentage agreement and κ , based on whether the threshold for ASD was met on a case-by-case basis. To increase insight in the underlying reasons of (dis)agreement, the convergence between both diagnostic instruments and the clinical diagnosis was also calculated with percentage agreement and κ . Post-hoc exploratory analyses determined the number of participants that met threshold for the different (sub)criteria of ASD. All analyses were conducted with SPSS (IBM Corp. Released 2017, 2017). A significance level of .05 was adopted. Sizes of κ were interpreted as follows: ≤ 0 = poor, .01- .20 = slight, .21-.40 = fair, .41-.60 = moderate, .61-.80 = substantial, and .81-1 = almost perfect (Sim & Wright, 2005).

RESULTS

Agreement between Classification According to 3di and DISCO-11

For 75% of the participants, the 3di and the DISCO-11 yielded the same classification. However, taking into account agreement based on chance, both instruments showed a non-significant, fair agreement, $\kappa = .20$, $p = .05$ (see Table 2). In four out of 25 participants the clinical ASD diagnosis was confirmed by both instruments, whereas eight autistic individuals did not to meet threshold on either the 3di and the DISCO-11. Three of the 25 participants with ASD scored positive on the DISCO-11, but not on the 3di, and ten autistic individuals showed the reversed pattern. The DISCO-11 and the 3di showed a fair, significant agreement for both criterion A (Deficits in social interaction and social communication), $\kappa = .47$, $p < .0001$, and for criterion B (Restricted, repetitive behaviors, interests and activities), $\kappa = .41$, $p = .0008$.

[INSERT TABLE 2]

Agreement between Clinical Diagnosis and Classification According to 3di ¹

The 3di classified 14 out of 25 individuals (56%) with a clinical ASD diagnosis as having ASD. Four of the 43 individuals without ASD (9%) met the 3di criteria for ASD, with all of these individuals belonging to the clinically referred group. Taken together, 3di classifications were in line with the clinical diagnosis in 78% of the cases. Taking chance-based agreement into account, the 3di and the clinical status showed a significant, moderate agreement, $\kappa = .50$, $p = <.0001$ (see Table 3).

[INSERT TABLE 3]

To get a better insight into why the 3di failed to pick up a substantial number of individuals with a clinical diagnosis of ASD, exploratory descriptive analyses studied how many cases met

¹ Exploratory post-hoc analyses suggested that there was no significant difference in intelligence, ADOS calibrated severity score, gender or order of interviews between the autistic individuals that scored above threshold versus below threshold on the 3di (all p-values $> .10$).

the 3di threshold of the (sub)criteria of ASD (see Table 4). Note that the 3di requires an individual's behavior to meet threshold on criterion A and B to classify for a diagnosis of ASD. Criterion C (Early onset) is not included in the 3di's algorithm. Twenty-nine percent of the 25 autistic individuals only met threshold for criterion A (Deficits in social interaction and communication), whereas one individual with ASD (4%) only met threshold for criterion B (Restricted, repetitive behaviors, interests and activities). Sixteen percent of the individuals with ASD did not meet the 3di threshold on either of both criteria. In both comparison groups, two individuals scored above threshold on criterion B, and two other individuals of the clinically-referred group met the 3di threshold on criterion A.

[INSERT FIGURE 1]

At the level of the sub-criteria, frequency analyses showed that almost all individuals with a clinical ASD diagnosis met threshold of sub-criterion A1 (Deficits in social-emotional reciprocity) and A2 (Deficits in nonverbal communication), and the majority of them also met threshold on sub-criterion A3 (Deficits in relationships; 80%). Exploratory analyses on the sub-criteria underlying criterion B showed that approximately half of the ASD group scored above the 3di-threshold for sub-criterion B1 (Stereotyped / repetitive behaviors; 48%) and B3 (Restricted, fixated interests; 52%). A slightly lower percentage of individuals with a clinical diagnosis of ASD met threshold on B4 (Hyper- and hyporeactivity; 44%) and on B2 (Insistence on sameness and routines; 36%). Taken together, these results indicated that the overall majority of individuals with ASD met the 3di-threshold on criterion A (80%), whereas 60% of them scored above threshold on criterion B.

[INSERT TABLE 4]

Agreement between Clinical Diagnosis and Classification According to DISCO-11 ²

Following the DISCO-11, 7 out of 25 individuals (28%) with a clinical diagnosis met the DSM-5 criteria for ASD. None of the 43 individuals of the clinically-referred group or the typically developing group scored above the threshold on the DISCO-11. Taken together, DISCO-11 classifications showed an agreement with the participant's clinical diagnosis in 74% of the instances. Taking into account agreement based on chance, the DISCO-11 and the clinical status showed a significant, fair agreement, $\kappa = .33$, $p = .0002$ (see Table 3).

In order to obtain better insight into the low convergence between the DISCO-11 classification and the clinical diagnosis in the ASD group, exploratory descriptive analyses examined how many individuals with ASD met the DISCO-11-threshold on the (sub)criteria of ASD (see Table 5). Remember that someone has to meet threshold on criterion A (Deficits in social interaction and social communication), criterion B (Restricted, repetitive behaviors, interests and activities), and criterion C (Early onset) to obtain a diagnostic classification of ASD according to DISCO-11. Twenty-eight percent of the individuals with ASD only met two of the three criteria, another 28% of the autistic individuals met only one criterion, and 16% of them did not meet the threshold for any of the three ASD-criteria (see Figure 2).

[INSERT FIGURE 2 HERE]

At the level of the criteria and sub-criteria, the DISCO-11 failed to detect 66% of the autistic individuals on criterion A. At the level of the sub-criteria, descriptive analyses suggested that fewer individuals with ASD scored above threshold on sub-criterion A1 (Deficits in social-

² Exploratory post-hoc analyses suggested that there was no significant difference in intelligence, ADOS calibrated severity score, gender or order of interviews between the autistic individuals that scored above threshold versus below threshold on the DISCO-11 (all p-values > .10).

emotional reciprocity; 60%) or A2 (Deficits in nonverbal communication; 64%), than on sub-criterion A3 (Deficits in relationships; 84%). Seventy-two percent of the ASD group scored above threshold for criterion B. Sub-criterion analyses indicated that the majority of individuals with ASD scored above the DISCO-11 threshold for B2 (Insistence on sameness and routines; 80%), and a significant number of individuals with ASD also met threshold of sub-criterion B1 and B4 (68% and 64%, *resp.*). In addition, only 52% of the individuals with a clinical diagnosis of ASD scored above threshold on the DISCO-11 for criterion C. In none of the comparison groups, individuals met DISCO-11 threshold for criterion A, but four individuals met threshold on criterion C (one TD individual and three individuals from the clinically-referred group), five individuals from the clinically-referred group met threshold for criterion B, and four other individuals from the clinically-referred group met threshold for both criteria B and C.

[INSERT TABLE 5]

DISCUSSION

Two semi-structured interviews that are specially developed to measure ASD symptoms have recently published new algorithms to examine DSM-5 characteristics: the 3di (Mandy, Wang, et al., 2017; Skuse et al., 2004), and the DISCO-11 (Kent et al., 2013; Wing, 2006). The present study was the first to compare the diagnostic classification according to these DSM-5 algorithms, and to examine the convergence with the clinical diagnostic status of three participant groups: (1) a group of children with ASD; (2) a clinically-referred group consisting of approximately 60% children with ADHD; and (3) a TD group. We will first discuss the low agreement between both instruments, and in an attempt to provide insight into this poor convergence, we will elaborate on the fact that the 3di was better in detecting socio-communicative problems than RRBIs in our

study, while the opposite pattern was found for the DISCO-11. We will finally provide alternative explanations why both instruments might have failed to predict clinical diagnoses in a substantial part of the ASD group in our study, and give some suggestions for future work.

Poor Agreement between Classification Based on 3di and DISCO-11

In 75% of the cases, the classification based on the 3di and the DISCO-11 was the same. However, taking the agreement based on chance into account (related to the greater number of participants with a clinical ASD diagnosis compared to those without an ASD diagnosis), this number significantly dropped ($\kappa = .20$). Focusing on the ASD group, only four of the 25 individuals scored above threshold of both the 3di and the DISCO-11, thirteen children were picked up by only one of the two instruments, and eight children with a clinical ASD diagnosis were not classified according to either instrument.

Although both instruments have not been compared before, these numbers are lower than expected based on previous studies that have compared the classifications according to the 3di or DISCO-11 with another instrument. The DSM-IV-TR algorithms – and not the DSM-5 algorithms – of the 3di and the DISCO were previously compared to classifications according to ADI-R or ADOS, and they have demonstrated a moderate to substantial agreement. The 3di agreed with the ADI-R classification in 65% of the 29 participants, of which 13 with a clinical PDD diagnosis (Skuse et al., 2004). Including the same participant sample, the short version of the 3di DSM-IV-TR algorithm has demonstrated a moderate-substantial agreement with the ADI-R, with κ -values between .48-.78 (Santosh et al., 2009). This short version of the DSM-IV-TR algorithm has also demonstrated a moderate agreement ($\kappa = .55$) with the ADOS-2 classifications in a group of 146 children at suspicion of having ASD (Slappendel et al., 2016). DISCO classifications showed a substantial agreement ($\kappa = .69$) with ADOS classifications in a group of children with and without

ASD and co-occurring intellectual disability (Maljaars et al., 2012), and a moderate agreement ($\kappa = .48$, calculated based on the numbers reported in Table 6, Maljaars et al., 2012) with ADI-R classifications in a group of 30 children and 27 adults with and without ASD (Nygren et al., 2009). Given that the agreement we have found is lower than what could be expected based on the DSM-IV-TR algorithms of both instruments, we examined the convergence of both instruments' classifications and the participants' clinical diagnoses.

The 3di is Better in Detecting Socio-Communicative Problems than RRBIs

The 3di picked up 56% of the individuals with a clinical ASD diagnosis, and 28% of the autistic individuals scored above threshold on one of the two criteria (criterion A 'Socio-communicative problems', criterion B 'RRBIs'). The remaining 16% of the ASD group did not meet threshold on either criteria A or B. In addition, the 3di identified 17% of false positives, all part of the clinically-referred group. These numbers – and especially the poor sensitivity – are significantly lower than expected based on previous research. As mentioned earlier, the psychometric properties of the 3di's DSM-5 algorithm are not available yet, but a previous study has shown that classifications based on the DSM-5 algorithm are strongly in line with 3di's classifications based on the DSM-IV-TR algorithm (83% agreement; Mandy, Wang, et al., 2017). It is important to note that the strong agreement between the DSM-IV-TR and DSM-5 algorithm does not necessarily imply an equally strong agreement of the DSM-5 algorithm's classification with the clinical diagnosis. However, the psychometric properties of the DSM-IV-TR algorithm have been studied. Skuse et al. (2004) compared 3di classifications with clinical judgement in a group of 27 children with ASD, 43 children without ASD, and 60 TD children aged 2-12 years. They established good agreements for the different PDD sub-classifications (κ -values between .47 and .86), and this number increased after combining these PDD sub-classifications into one

“autism spectrum” classification ($\kappa = .95$), which significantly exceeds the agreement we have found ($\kappa = .50$).

To study why the classifications according to the DSM-5 algorithm diverge from the previously published DSM-IV-TR algorithm and to provide insight into the low sensitivity of the 3di in our study, exploratory analyses were conducted, suggesting that the 3di’s poor ability to detect autistic individuals was mostly related to criterion B. Whereas 80% of the autistic children meet threshold on criterion A (‘Deficits in social interaction and social communication’), only 60% of them meet the criterion B (‘RRBIs’) threshold. The poor ability of the DSM-5 algorithm to detect RRBI sub-criteria might be associated with how the algorithm was constructed. The item selection procedure started from the DSM-IV-TR version of the 3di algorithm. This strong link between the DSM-5 and DSM-IV-TR version of the algorithm is confirmed by the high agreement between classifications based on both algorithms (Mandy, Wang, et al., 2017), and might unintentionally also imply that RRBIs – which became more important in DSM-5 – are not fully represented in the new algorithm. Two recent studies in which 3di algorithm items were assigned to DSM-5 sub-criteria (Evers et al., under review; Slappendel et al., 2016) confirm that RRBIs are currently underrepresented in the 3di DSM-5 algorithm. The item mapping on the published DSM-5 algorithm of the 3di (Evers et al., under review) showed that the majority of items focus on socio-communicative problems (43 items), whereas fewer items assessed RRBIs (28 items). Previously, Slappendel and colleagues (2016) already demonstrated that some exemplars were lacking in the (preliminary version of) the algorithm’s items. This appeared especially the case for exemplars under sub-criterion B2 (‘Insistence on sameness and routines’) and sub-criterion B4 (‘Hyper- or hyporeactivity to sensory input’), which were exactly the sub-criteria that picked up fewest of our autistic participants.

The DISCO-11 is Better in Detecting RRBI than Socio-Communicative Problems

Only a minority of the individuals with ASD (28%) scored above threshold on the instrument, and an additional 12% met threshold for criteria A and B, but not for criterion C ('Early onset'), a criterion that was not included in the algorithm of the 3di. By adding both numbers, the sensitivity of the DISCO-11 DSM-5 algorithm comes closer to what we have found for the 3di, but is still somewhat lower than the 3di. In addition to the 40% of participants with a clinical diagnosis of ASD that met threshold for (at least) criteria A and B, sixteen percent met threshold on criteria B and C and 28% of the autistic participants scored above cut-off on either criterion A, B or C. Sixteen percent of the ASD group did not meet threshold on any of the three criteria. At the positive side, this low sensitivity coincided with a perfect specificity: No false positives were detected by the DISCO-11. Previous work by Kent and colleagues (2013) has demonstrated that the DSM-5 algorithm has a sensitivity ranging from .85 to 1 and a specificity between .74 and .95, depending upon the age group and the ability level of their participants. The number of autistic children that was detected by the DSM-5 algorithm in the present study was clearly below what could be expected based on these previously established psychometric data (Kent et al., 2013). Participant characteristics might explain these different findings. Whereas the majority of the participants included in Kent et al. (2013) were diagnosed with autistic disorder and had a co-occurring intellectual disability, our participants were 4-18 year old children with an intelligence in the normal range. ASD characteristics might show differently in individuals with intellectual disability compared to individuals with an intelligence in the normal range (Maljaars et al., 2012). Hence, the inclusion of a significant proportion of individuals with an intellectual disability in the algorithm's validation sample may have impacted the algorithm, both with regards to item selection and set thresholds. Even though the DSM-5 algorithm performed comparably well across

different levels of ability and ages (Kent et al., 2013), it has to be noted that sample sizes were rather small for some age groups and levels of ability, which might call for additional research evaluating the need for different algorithms for different subgroups on the autism spectrum.

Additional analyses showed that the low sensitivity of the DSM-5 algorithm in our study seemed to be mostly related to the algorithm's poor ability to detect socio-communicative problems. Whereas 72% of the autistic participants scored above threshold on criterion B ('RRBIs'), only 44% met threshold of criterion A ('Deficits in social interaction and social communication'), with especially 'Deficits in social reciprocity' (sub-criterion A1) and 'Deficits in nonverbal behavior' (sub-criterion A2) performing poorly. Recent findings showed that socio-communicative behaviors might fall apart in more sub-dimensions than RRBIs – at least in a group of verbally fluent children between 3 and 13 years old (Bishop, Havdahl, Huerta, & Lord, 2016). This might suggest that more items are required to cover the socio-communicative domain than the RRBI criteria. However, the DISCO-11 algorithm currently includes slightly more RRBI items compared to items about socio-communicative problems (36 criterion A items; 40 criterion B items), according to a study that assigned algorithm items to DSM-5 sub-criteria (Evers et al., under review). This might potentially cause an under-representation of the wide(r) range of impairments in the socio-communicative domain of impairments (Huerta, Bishop, Duncan, Hus, & Lord, 2012).

Additional Reasons for Poor Convergence, and Study Limitations

In addition to reasons that might explain the poor convergence of both instruments' outcomes and that are specific to either instrument, some alternative mechanisms could also play a role. First, our sample size ($n = 68$) is rather small. To develop the algorithms and determine their psychometric properties, many more participants were included (387 participants in Kent et

al., 2013; 1,081 participants in Mandy, Wang, et al., 2017; 120 participants in Skuse et al., 2004). The lower number of participants that was included in our study might explain the lower convergence with clinical diagnosis found in our study, and therefore these results of our study should be interpreted with caution. However, this sample size is comparable to that of previous validation studies comparing the DSM-IV-TR algorithms of the 3di or DISCO to classifications based on other instruments such as ADI-R or ADOS (Maljaars et al., 2012, $n = 115$; Nygren et al., 2009, $n = 57$; Santosh et al., 2009, $n = 29$; Skuse et al., 2004, $n = 29$; Slappendel et al., 2016, $n = 146$), so our limited sample size is quite unlikely to be the only explanation for the poor convergence between DISCO-11 and 3di classifications. Second, different interviewers were involved. Although the interviewers all received an elaborate training, the level of experience with the instrument was highly variable, which might have induced a bias towards systematically under-scoring behaviors. However, exploratory analyses do not seem to support this hypothesis, as no significant association was found between the interviewer (master student versus clinician) and the diagnostic classification based on the instrument (for 3di: $\chi^2(1) = .12$, $p = .72$; for DISCO-11: $\chi^2(1) = .32$, $p = .57$; these analyses were not reported in the Results because of their post-hoc and exploratory nature). Third, lower educated parents were un(der)-represented in this study, which could have influenced our results and which limits the generalizability of our findings. Fourth, the DSM-5 was introduced in 2013, but only fully implemented in Flemish child psychiatry around 2015. Some of the autistic individuals were diagnosed around that transition period, which might have had an impact on the diagnostic decision-making. Although the presence of reported DSM-5 criteria was checked in the study protocol, we cannot exclude that DSM-IV-TR criteria were used in the context of clinical assessment for a small number of autistic participants. Finally, a significant number of participants with a clinical ASD diagnosis did not score above threshold on

either instruments. The participants in this study were 4-18 year olds without intellectual disability, while the instruments were developed for a wider age range and a broader range of cognitive abilities. Consequently, this study might have focused on a group for which both instruments may be less sensitive. In addition, the clinical diagnoses resulted from a multidisciplinary assessment at a hospital unit specialized in ASD. Although the procedure of reviewing participants' diagnostic reports confirmed that almost all participants who received a clinical ASD diagnosis fulfilled the diagnostic criteria, we did not administer an independent diagnostic instrument to support this diagnosis. However, in the context of the clinical diagnostic trajectory, an ADOS-2 was used to assess ASD characteristics in 21 out of 25 children, and 81% of them scored above the ADOS-2 cutoff. Interestingly, post-hoc exploratory analyses suggested no significant difference in ADOS calibrated severity scores of the autistic individuals that scored above threshold versus those scoring below threshold on the 3di and the DISCO-11. This suggests that the 3di and the DISCO-11 might be equally sensitive for individuals with greater versus fewer autistic symptoms.

Future Directions

The poor ability of both instruments to predict clinical ASD diagnoses in this group of children without intellectual disability might illustrate that the change in diagnostic criteria has altered the characterization of ASD (Bennett & Goodall, 2016; Kulage et al., 2014; Sturmey & Dalfern, 2014). It therefore seems crucial to validate the DSM-5 algorithms on a large DSM-5 diagnosed group of individuals with ASD, instead of DSM-IV-TR diagnosed individuals (Kent et al., 2013; Mandy, Wang, et al., 2017). The next step is therefore to replicate our design using an independent group of individuals with a DSM-5 diagnosis. When our data are confirmed by another sample, the DSM-5 algorithms need to be improved, by either including different items in

the algorithm, based on the suggestions provided in previous studies (Evers et al., under review; Slappendel et al., 2016). The algorithms could also be refined by changing the thresholds. As both the 3di and the DISCO-11 algorithm have an excellent specificity (based on our results, and on previous data by Kent et al., 2013; Skuse et al., 2004), there is still much (statistical) room left to shift the balance to an improved sensitivity. Further research is also needed to determine the need for developing different algorithms for different subgroups. Some studies have demonstrated that the DSM-IV-TR algorithms perform comparable across different ages, genders and levels of intellectual ability, but the sample size for some subgroups was rather limited (Kent et al., 2013; Maljaars et al., 2012; Mandy, Wang, et al., 2017; Slappendel et al., 2016).

Conclusions

We compared the classifications based on two semi-structured parental interviews that measure ASD characteristics and have newly developed DSM-5 algorithms, by administering the 3di and the DISCO-11 in three groups of participants between 4 and 18 years old: a group of children with ASD, a clinically-referred group, and a TD group. Although our study has demonstrated a poor agreement between both instruments, and a limited ability of both instruments' DSM-5 algorithms to predict clinical diagnoses in this group, we are still convinced of the role semi-structured parental interviews should play in the diagnostic assessment of ASD (Ashwood et al., 2015; Constantino & Charman, 2016; Falkmer et al., 2013; Gardner et al., 2016; Penner, Zwaigenbaum, & Roberts, 2015). Clinicians should always combine evidence across different sources of information and should never exclusively focus on the quantitative outcome of a diagnostic instrument. These results could also be regarded as a call towards a more dimensional approach of the algorithms of these interviews. At this moment, the quantitative outcome of the algorithms is a categorical decision ("above" versus "below" threshold), whereas

the algorithms might benefit from a more dimensional quantitative approach (as done for other instruments), indicating the degree of behaviors characteristic of ASD. Having said that, our results have also shown that there is work to be done in the fine-tuning of the different DSM-5 algorithms, such that they perform better in detecting individuals clinically diagnosed on the basis of DSM-5 criteria.

BIBLIOGRAPHY

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th edition text revision)*. Washington, D.C.: Author.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th edition)*. Washington, D.C.: Author.
- Ashwood, K. L., Buitelaar, J., Murphy, D., Spooren, W., & Charman, T. (2015). European clinical network: Autism spectrum disorder assessments and patient characterisation. *European Child & Adolescent Psychiatry*, 24(8), 985–995.
<https://doi.org/10.1007/s00787-014-0648-2>
- Bennett, M., & Goodall, E. (2016). A Meta-analysis of DSM-5 Autism Diagnoses in Relation to DSM-IV and DSM-IV-TR. *Review Journal of Autism and Developmental Disorders*, 3(2), 119–124. <https://doi.org/10.1007/s40489-016-0070-4>
- Bishop, S. L., Havdahl, K. A., Huerta, M., & Lord, C. (2016). Sub-dimensions of social-communication impairment in autism spectrum disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 57(8), 909–916.
<https://doi.org/10.1111/jcpp.12510>
- Carrington, S., Kent, R., Maljaars, J., Le Couteur, A., Gould, J., Wing, L., ... Leekam, S. R. (2014). DSM-5 Autism Spectrum Disorder: In search of essential behaviours for diagnosis. *Research in Autism Spectrum Disorders*, 8(6), 701–715.
<https://doi.org/10.1016/j.rasd.2014.03.017>
- Carrington, S., Leekam, S., Kent, R., Maljaars, J., Gould, J., Wing, L., ... Noens, I. (2015). Signposting for diagnosis of Autism Spectrum Disorder using the Diagnostic Interview

- for Social and Communication Disorders (DISCO). *Research in Autism Spectrum Disorders*, 9, 45–52. <https://doi.org/10.1016/j.rasd.2014.10.003>
- Constantino, J. N., & Charman, T. (2016). Diagnosis of autism spectrum disorder: Reconciling the syndrome, its diverse origins, and variation in expression. *The Lancet Neurology*, 15(3), 279–291. [https://doi.org/10.1016/S1474-4422\(15\)00151-9](https://doi.org/10.1016/S1474-4422(15)00151-9)
- Evers, K., Maljaars, J., Carrington, S., Carter, A., Happé, F., Leekam, Susan, & Noens, I. (under review). *How well are DSM-5 diagnostic criteria for ASD represented in standardized diagnostic instruments?*
- Falkmer, T., Anderson, K., Falkmer, M., & Horlin, C. (2013). Diagnostic procedures in autism spectrum disorders: A systematic literature review. *European Child & Adolescent Psychiatry*, 22(6), 329–340. <https://doi.org/10.1007/s00787-013-0375-0>
- Gardner, L., Erkfritz-Gay, K., Campbell, J. M., Bradley, T., & Murphy, L. (2016). Purposes of Assessment. In J. L. Matson (Ed.), *Handbook of Assessment and Diagnosis of Autism Spectrum Disorder* (pp. 27–43). https://doi.org/10.1007/978-3-319-27171-2_3
- Hendriks, M. P. H., & Ruiter, S. (2017). *Wechsler Intelligence Scales for Children 5th edition*. Amsterdam, the Netherlands: Pearson.
- Huerta, M., Bishop, S. L., Duncan, A., Hus, V., & Lord, C. (2012). Application of DSM-5 criteria for autism spectrum disorder to three samples of children with DSM-IV diagnoses of pervasive developmental disorders. *The American Journal of Psychiatry*, 169(10), 1056–1064. <https://doi.org/10.1176/appi.ajp.2012.12020276>
- IBM Corp. Released 2017. (2017). IBM SPSS Statistics for Windows (Version Version 25.0.). Armonk, NY: IBM Corp.

- Kent, R. G., Carrington, S. J., Le Couteur, A., Gould, J., Wing, L., Maljaars, J., ... Leekam, S. R. (2013). Diagnosing autism spectrum disorder: Who will get a DSM-5 diagnosis? *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 54(11), 1242–1250. <https://doi.org/10.1111/jcpp.12085>
- Kulage, K. M., Smaldone, A. M., & Cohn, E. G. (2014). How will DSM-5 affect autism diagnosis? A systematic literature review and meta-analysis. *Journal of Autism and Developmental Disorders*, 44(8), 1918–1932. <https://doi.org/10.1007/s10803-014-2065-2>
- Leekam, S. R., Libby, S. J., Wing, L., Gould, J., & Taylor, C. (2002). The Diagnostic Interview for Social and Communication Disorders: Algorithms for ICD-10 childhood autism and Wing and Gould autistic spectrum disorder. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 43(3), 327–342.
- Maljaars, J., Noens, I., Scholte, E., & van Berckelaer-Onnes, I. (2012). Evaluation of the criterion and convergent validity of the Diagnostic Interview for Social and Communication Disorders in young and low-functioning children. *Autism*, 16(5), 487–497. <https://doi.org/10.1177/1362361311402857>
- Mandy, W., Clarke, K., McKenner, M., Strydom, A., Crabtree, J., Lai, M.-C., ... Skuse, D. (2017). Assessing Autism in Adults: An Evaluation of the Developmental, Dimensional and Diagnostic Interview—Adult Version (3Di-Adult). *Journal of Autism and Developmental Disorders*, 1–12. <https://doi.org/10.1007/s10803-017-3321-z>
- Mandy, W., Wang, A., Lee, I., & Skuse, D. (2017). Evaluating social (pragmatic) communication disorder. *Journal of Child Psychology and Psychiatry*, 58(10), 1166–1175. <https://doi.org/10.1111/jcpp.12785>

- Nygren, G., Hagberg, B., Billstedt, E., Skoglund, Å., Gillberg, C., & Johansson, M. (2009). The Swedish Version of the Diagnostic Interview for Social and Communication Disorders (DISCO-10). Psychometric Properties. *Journal of Autism and Developmental Disorders*, 39(5), 730–741. <https://doi.org/10.1007/s10803-008-0678-z>
- Penner, M., Zwaigenbaum, L., & Roberts, W. (2015). Diagnosis: Screening, surveillance, assessment, and formulation. In E. Anagnostou & J. Brian (Eds.), *Clinician's Manual on Autism Spectrum Disorder* (pp. 5–17). https://doi.org/10.1007/978-3-319-03056-2_2
- Rutter, M., Bailey, A., & Lord, Catherine, C. (2003). *The Social Communication Questionnaire*. Los Angeles, CA: Western Psychological Services.
- Santosh, P. J., Mandy, W. P. L., Puura, K., Kaartinen, M., Warrington, R., & Skuse, D. H. (2009). The construction and validation of a short form of the developmental, diagnostic and dimensional interview. *European Child & Adolescent Psychiatry*, 18(8), 521–524. <https://doi.org/10.1007/s00787-009-0004-0>
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell-Horn-Carroll model of intelligence. In *Contemporary intellectual assessment: Theories, tests, and issues*, 3rd ed (pp. 99–144). New York, NY, US: The Guilford Press.
- Sim, J., & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, 85(3), 257–268. <https://doi.org/10.1093/ptj/85.3.257>
- Skuse, D., Warrington, R., Bishop, D., Chowdhury, U., Lau, J., Mandy, W., & Place, M. (2004). *The Developmental, Dimensional and Diagnostic Interview (3di): A Novel Computerized Assessment for Autism Spectrum Disorders*. Retrieved from

<http://www.sciencedirect.com/science/article/B987N-4XKH91R-8/2/67972fcb712fc0208a344e6367e901ea>

Slappendel, G., Mandy, W., Ende, J. van der, Verhulst, F. C., Sijde, A. van der, Duvekot, J., ...

Greaves-Lord, K. (2016). Utility of the 3Di Short Version for the Diagnostic Assessment of Autism Spectrum Disorder and Compatibility with DSM-5. *Journal of Autism and Developmental Disorders*, 46(5), 1834–1846. <https://doi.org/10.1007/s10803-016-2713-9>

Sturmey, P., & Dalfern, S. (2014). The Effects of DSM5 Autism Diagnostic Criteria on Number of Individuals Diagnosed with Autism Spectrum Disorders: A Systematic Review.

Review Journal of Autism and Developmental Disorders, 1(4), 249–252.

<https://doi.org/10.1007/s40489-014-0016-7>

UNESCO Institute for Statistics. (2012). *International Classification of Education (ISCED 2011)*. Montreal, Canada: Author.

Wechsler, D. (1992). *Wechsler intelligence scale for children (3rd edition)*. London, United Kingdom: The Psychological Corporation.

Wechsler, D., & Naglieri, J. A. (2006). *Wechsler Nonverbal Scale of Ability (WNV)*. Pearson.

Wing, L. (2006). *Diagnostic Interview for Social and Communication Disorders, 11th edition (DISCO-11)*. Bromley, United Kingdom: Centre for Social and Communication Disorders.

Wing, L., Leekam, S. R., Libby, S. J., Gould, J., & Larcombe, M. (2002). The Diagnostic Interview for Social and Communication Disorders: Background, inter-rater reliability and clinical use. *Journal of Child Psychology and Psychiatry*, 43(3), 307–325.

<https://doi.org/10.1111/1469-7610.00023>

FIGURE CAPTIONS

Figure 1. The Percentage of Individuals Scoring Above Threshold on Two, One or None of the DSM-5 Criteria for ASD on the 3di: Pie Charts for the ASD Group, the Clinically-Referred Group, and the Typically Developing Group

Figure 2. The Percentage of Individuals Scoring Above Threshold on Two, One or None of the DSM-5 Criteria for ASD on the DISCO-11: Pie Charts for the ASD Group, the Clinically-Referred Group, and the Typically Developing Group

Figure 1.

Figure 2.

Table 1.

Characteristics of the Three Participant Groups: Individuals with Autism Spectrum Disorder (ASD), a Clinically Referred Group (CLINREF), and a Typically Developing Group (TD)

Participant group	ASD	CLINREF	TD
Total number	(n = 25)	(n = 24)	(n = 19)
Gender ratio	(20 M : 5 F)	(18 M : 6 F)	(13 M : 6 F)
Mean age (in months; SD between brackets)	117.04 (40.34)	106.21 (28.80)	104.37 (40.57)
Mean IQ (SD between brackets)	102.92 (15.03)	98.82 (8.92)	N.A.
Clinical diagnosis:			
• ASD	14	0	0
• ASD + ADHD	3	0	0
• ASD + other dx.	5	0	0
• ASD + ADHD + other dx.	3	0	0
• ADHD	0	14	0
• ADHD + other dx.	0	1	0
• Other dx.	0	1	0
• No dx. (yet)	0	8	19

Abbreviations. SD = standard deviation; M = male; F = female; dx. = diagnosis.

Note. Intelligence assessment was unavailable for one child in the ASD group, and for two participants in the CLINREF group.

Table 2.

Number of Individuals that Score Above or Below Threshold for the Diagnostic Classification ASD According to the DISCO-11 and the 3di, Presented for the ASD Group, the Clinically-Referred Group (CLINREF) and the Typically Developing Group (TD), and for the Total Participant Group

			3di classification	
			“ASD”	“no ASD”
Above threshold	Participant group	DISCO-11 classification		
Diagnostic classification	ASD group	“ASD”	<u>4</u>	3
		“no ASD”	10	<u>8</u>
	CLINREF group	“ASD”	<u>0</u>	0
		“no ASD”	4	<u>20</u>
	TD group	“ASD”	<u>0</u>	0
		“no ASD”	0	<u>19</u>
	Total group	“ASD”	<u>4</u>	3
		“no ASD”	14	<u>47</u>

Note. The underlined numbers are the instances in which there is an agreement between both instruments.

Table 3.

Agreement Between Clinical Diagnosis, Classification Based on DISCO-11, and Classification Based on 3di: Number of Individuals in the ASD Group, the Clinically-Referred Group (CLINREF), and the Typically Developing Group (TD)

Participant group	Clinical dx.	3di classification		DISCO-11 classification	
		"ASD"	"No ASD"	"ASD"	"No ASD"
ASD	ASD	<u>14</u>	11	<u>7</u>	18
CLINREF	No ASD	4	<u>20</u>	0	<u>24</u>
TD	No ASD	0	<u>19</u>	0	<u>19</u>

Abbreviation. dx. = diagnosis.

Note. The underlined numbers are the instances in which there is an agreement between the clinical diagnosis and the classification based on the diagnostic instrument.

Table 4.

The Percentage of Participants of the ASD Group, the Clinically-Referred Group (CLINREF) and the Typically Developing Group (TD) Meeting Threshold of the 3di for the Criteria and Sub-criteria of ASD (in Percentages)

	Participant group		
	ASD	CLINREF	TD
A1	<u>96</u>	42	16
A2	<u>96</u>	38	21
A3	<u>80</u>	42	0
A	<u>80</u>	25	0
B1	48	21	5
B2	36	42	16
B3	52	25	16
B4	44	33	11
B	<u>60</u>	25	11

A = Deficits in social interaction and social communication; A1 = Deficits in social-emotional reciprocity; A2 = Deficits in nonverbal communication; A3 = Deficits in relationships; B = Restricted, repetitive behavior, interests or activities; B1 = Stereotyped / repetitive behaviors; B2 = Insistence on sameness and routines; B3 = Restricted, fixated interests; B4 = Hyper- or hyporeactivity to sensory input.

Note. The underlined numbers represent the instances that are in line with their clinical ASD diagnosis.

Table 5.

The Percentage of Participants of the ASD Group, the Clinically-Referred Group (CLINREF) and the Typically Developing Group (TD) Meeting Threshold of the DISCO-11 for the Criteria and Sub-criteria of ASD (in Percentages)

	Participant group		
	ASD	CLINREF	TD
A1	<u>60</u>	4	0
A2	<u>64</u>	33	5
A3	<u>84</u>	54	0
A	<u>44</u>	0	0
B1	68	29	0
B2	80	33	0
B3	56	38	0
B4	64	33	5
B	<u>72</u>	38	0
C	<u>52</u>	29	5

A = Deficits in social interaction and social communication; A1 = Deficits in social-emotional reciprocity; A2 = Deficits in nonverbal communication; A3 = Deficits in relationships; B = Restricted, repetitive behavior, interests or activities; B1 = Stereotyped / repetitive behaviors; B2 = Insistence on sameness and routines; B3 = Restricted, fixated interests; B4 = Hyper- or hyporeactivity to sensory input; C = Early onset.

Note. The underlined number represent the instances that are in line with their clinical ASD diagnosis. Percentages might not add up to 1 due to rounding.

Supplementary Material: Additional Analyses

Twenty-one of the 25 participants with a clinical diagnosis of ASD (84%) were also given a research diagnosis based on their multidisciplinary clinical reports. All analyses were repeated after excluding the four participants from the ASD group, without a research diagnosis of ASD. The results of those analyses are reported below, but are highly similar to the results of the full ASD group ($n = 25$).

Agreement between Classification According to 3di and DISCO-11

For 75% of the participants, the 3di and the DISCO-11 yielded the same classification, however, taking into account agreement based on chance, both instruments showed a significant, fair agreement, $\kappa = .22$, $p = .03$ (see Table A1). The clinical ASD diagnosis was confirmed by both instruments for four out of 21 participants.

The DISCO-11 and the 3di showed a fair, significant agreement for both criterium A (Deficits in social interaction and social communication), $\kappa = .35$, $p = .0006$, and for criterium B (Restricted, repetitive behaviors, interests and activities), $\kappa = .40$, $p = .001$.

Table A1.

Number of Individuals that Score Above or Below Threshold for ASD According to the DISCO-11 and the 3di, Presented for the ASD Group, the Clinically-Referred Group (CLINREF) and the Typically Developing Group (TD), and for the Total Participant Group

		3di classification	
		ASD	no ASD
Participant group	DISCO-11 classification		
ASD	ASD	4	2

CLINREF	no ASD	10	<u>5</u>
	ASD	<u>0</u>	0
TD	no ASD	4	<u>20</u>
	ASD	0	0
		0	<u>19</u>
Total	ASD	<u>4</u>	2
	no ASD	14	<u>44</u>

Note. The underlined numbers are the instances in which there is an agreement between both instruments.

Agreement between Clinical Diagnosis and Classification According to 3di

The 3di classified 14 out of 21 individuals (67%) with a clinical ASD diagnosis as having ASD. Four of the 43 individuals without ASD met the 3di criteria for ASD, with all of these individuals belonging to the clinically referred group. Taken together, 3di classifications were in line with the clinical diagnosis in 83% of the cases. Taking chance-based agreement into account, the 3di and the clinical status showed a significant, moderate agreement, $\kappa = .60$, $p = <.0001$ (see Table A2).

Table A2.

Agreement Between Clinical Diagnosis, Classification Based on DISCO-11, and Classification Based on 3di: Number of Individuals in the ASD Group, the Clinically-Referred Group (CLINREF), and the Typically Developing Group (TD)

Participant group	Clinical dx.	3di classification		DISCO-11 classification	
		ASD	No ASD	ASD	No ASD
ASD	ASD	<u>14</u>	7	<u>6</u>	15
CLINREF	No ASD	4	<u>20</u>	0	<u>24</u>
TD	No ASD	0	<u>19</u>	0	<u>19</u>

Abbreviation. dx. = diagnosis.

Note. The underlined numbers are the instances in which there is an agreement between the clinical diagnosis and the classification based on the diagnostic instrument.

To get a better insight in why the 3di failed to pick up a substantial number of individuals with a clinical diagnosis of ASD, exploratory descriptive analyses studied how many individuals met the 3di threshold of the (sub)criteria of ASD (see Table A3). Note that the 3di requires an individual to meet threshold on criterium A and criterium B to classify for a diagnosis of ASD. Criterium C (Early onset), however, is not included in the 3di's algorithm. Four of the 25 autistic individuals (19%) only met threshold for criterium A (Deficits in social interaction and communication), whereas one individual with ASD (5%) only met threshold for criterium B (Restricted, repetitive behaviors, interests and activities). Two individuals with ASD (10%) did not meet the 3di threshold on either of both criteria. In both comparison groups, two individuals scored above threshold on criterium B, and two individuals of the clinically-referred group met the 3di threshold on criterium A.

Table A3.

Participants of the ASD Group, the Clinically-Referred Group (CLINREF) and the Typically Developing Group (TD) Meeting Threshold of the 3di for the Criteria and Subcriteria of ASD (in Percentages)

	Participant group		
	ASD	CLINREF	TD
A and B	<u>67</u>	17	0
only A	19	8	0
only B	5	8	11
None	10	67	89
A1	100	42	13
A2	100	38	17
A3	86	42	0
A	<u>86</u>	25	0

B1	57	21	4
B2	43	42	13
B3	62	25	13
B4	48	33	8
B	<u>71</u>	25	8

A = Deficits in social interaction and social communication; A1 = Deficits in social-emotional reciprocity; A2 = Deficits in nonverbal communication; A3 = Deficits in relationships; B = Restricted, repetitive behavior, interests or activities; B1 = Stereotyped / repetitive behaviors; B2 = Insistence on sameness and routines; B3 = Restricted, fixated interests; B4 = Hyper- or hyporeactivity to sensory input.

Note. The underlined number represent the instances that are in line with their clinical ASD diagnosis. Due to rounding rules, some of the percentages might not add up to 100%.

At the level of the subcriteria, frequency analyses showed that all individuals with clinical ASD diagnosis met threshold of subcriteria A1 (Deficits in social-emotional reciprocity) and A2 (Deficits in nonverbal communication), and the majority of them also met threshold on subcriterion A3 (Deficits in relationships; 86%). Exploratory analyses on the subcriteria underlying criterium B showed that about 60% of the ASD group scored above the 3di-threshold for subcriterion B3 (Restricted, fixated interests; 62%), and subcriterion B1 (Stereotyped / repetitive behaviors; 57%). A lower percentage of individuals with clinical diagnosis of ASD met threshold on B4 (Hyper- and hyporeactivity; 48%) and on B2 (Insistence on sameness and routines; 43%). Taken together, these results indicated that the overall majority of individuals with ASD met the 3di-threshold on criterium A (86%), whereas 71.43% of them scored above threshold on criterium B.

Agreement between Clinical Diagnosis and Classification According to DISCO-11

According to the DISCO-11, 6 out of 21 individuals (29%) with a clinical diagnosis met the criteria for ASD. None of the 43 individuals of the clinically-referred group or the typically

developing group scored above the threshold on the DISCO-11. Taken together, DISCO-11 classifications showed an agreement with the participant's clinical diagnosis ("ASD" versus "no ASD") in 66% of the instances. Taking into account agreement based on chance, the DISCO-11 and the clinical status showed a significant, fair agreement, $\kappa = .35$, $p = .0002$ (see Table A2).

In order to obtain better insight in the low agreement between DISCO-11 classification and clinical diagnosis in the ASD group, exploratory descriptive analyses examined how many individuals with ASD met the DISCO-11-threshold on the (sub)criteria of ASD (see Table A4). Remember that an individual has to meet threshold on criterium A (Deficits in social interaction and social communication), criterium B (Restricted, repetitive behaviors, interests and activities), and criterium C (Early onset), to obtain a diagnostic classification of ASD according to DISCO-11. Six of the 21 individuals with ASD (29%) only met two of the three criteria, five autistic individuals (24%) met only one criterium, and four individuals with ASD (19%) did not meet threshold for any of the three ASD-criteria. Especially on criterium A, the DISCO-11 failed to detect 67% of the autistic individuals. Only 57% of the individuals with a clinical diagnosis of ASD scored above threshold on the DISCO-11 for criterium C, whereas 76% of the ASD group scored above threshold for criterium B. In none of the comparison groups, individuals met DISCO-11 threshold for criterium A, but four individuals met threshold on criterium C (one TD individual and three individuals from the clinically-referred group), five individuals from the clinically-referred group met threshold for criterium B, and four individuals from the clinically-referred group met threshold for both criterium B and C.

Table A4.

Participants of the ASD Group, the Clinically-Referred Group (CLINREF) and the Typically Developing Group (TD) Meeting Threshold of the DISCO-11 for the Criteria and Subcriteria of ASD (in Percentages)

	Participant group		
	ASD	CLINREF	TD
A, B, and C	<u>29</u>	0	0
A and B	10	0	0
B and C	19	17	0
Only A	5	0	0
Only B	19	21	0
Only C	0	13	5
None	19	50	95
A1	62	4	0
A2	67	33	5
A3	90	54	0
A	<u>43</u>	0	0
B1	67	0	0
B2	86	33	0
B3	57	38	0
B4	67	33	5
B	<u>76</u>	38	0
C	<u>57</u>	29	5

A = Deficits in social interaction and social communication; A1 = Deficits in social-emotional reciprocity; A2 = Deficits in nonverbal communication; A3 = Deficits in relationships; B = Restricted, repetitive behavior, interests or activities; B1 = Stereotyped / repetitive behaviors; B2 = Insistence on sameness and routines; B3 = Restricted, fixated interests; B4 = Hyper- or hyporeactivity to sensory input; C = Early onset.

Note. The underlined number represent the instances that are in line with their clinical ASD diagnosis. Due to rounding, some of the percentages might not add up to 100.

At the level of the subcriteria, descriptive analyses suggested that fewer individuals with ASD scored below threshold on subcriteria A1 (Deficits in social-emotional reciprocity; 62%) and A2 (Deficits in nonverbal communication; 67%), than on subcriterion A3 (Deficits in relationships; 90%). Analyses also showed that the majority of individuals with ASD scored above the DISCO-11 threshold for B2 (Insistence on sameness and routines; 86%). A significant number

of individuals with ASD also met threshold of subcriterion B1 and B4 (67%). Fifty-seven percent of the individuals with ASD met threshold on subcriterion B4 (Hyper- and hyposensitivity). Taken together, these results showed that less than half of the children with a clinical diagnosis of ASD met threshold on criterion A (44%), which seems to be related to the low number of autistic individuals passing threshold on criterion A1 and A2. 57% of them met threshold on criterion C, and more than three quarters of the ASD group scored above the DISCO-11 threshold on criterion B (76%).