

Building a theoretical framework for ASD screening instruments in Europe

Running Head:

ASD Screening instruments: A theoretical framework

María Magán-Maganto¹, Sigrídur Lóa Jónsdóttir², Ana B Sánchez-García³, Patricia García-Primo⁴, Annika Hellendoorn⁵, Tony Charman⁶, Herbert Roeyers⁷, Mieke Dereu⁷, Irma Moilanen⁸, Filippo Muratori⁹, Manuel Posada de la Paz⁴, Bernadette Rogé¹⁰, Iris J Oosterling¹¹, Anneli Yliherva¹² and Ricardo Canal-Bedia¹.

¹ Institute of Community Integration (*INICO*), Faculty of Education, University of Salamanca, Spain

² State Diagnostic and Counselling Centre, Iceland

³ Department of Didactic, Organization and Research Methods, Faculty of Education, University of Salamanca, Spain

⁴ Institute of Rare Diseases Research, ISCIII, Madrid, Spain

⁵ Department of Educational Sciences, Utrecht University, The Netherlands

⁶ Institute of Psychiatry, Psychology & Neuroscience, King's College London, UK

⁷ Department of Experimental Clinical and Health Psychology, Ghent University, Belgium

⁸ University and University Hospital of Oulu, Finland

⁹ IRCCS Stella Maris Foundation, Department of Developmental Neuroscience, Pisa, Italy

¹⁰ Octogone/CERPP, Université Toulouse 2 Jean-Jaurès and IUF, France

¹¹ Karakter Child and Adolescent Psychiatry University Centre Nijmegen, The Netherlands

¹² Faculty of Humanities, Logopedic Child Language Research Center, University of Oulu, Finland

Corresponding author's contact details:

María Magán-Maganto

e-mail: mmmaria@usal.es

Telephone Number: +34 661902773

Address: C/ Canalejas, 169 Facultad de Educación. Centro de Recursos para el estudio del autismo (INFOAUTISMO). INICO. Universidad de Salamanca. 37008.

Salamanca. España.

Abstract

Background: This study addresses the need for a theoretical base to develop more effective early ASD detection tools. The structure that underlies early ASD detection is explored by evaluating the opinions of experts on ASD screening tools currently used in Europe. **Method:** A process of face and content validity was performed. Firstly, the best constructs were selected from the relevant tests: Checklist for Early Signs of Developmental Disorders (CESDD), Checklist for Autism in Toddlers (CHAT), Early Screening of Autistic Traits Questionnaire (ESAT), Modified Checklist for Autism in Toddlers (M-CHAT), Social Communication Questionnaire (SCQ) and Communication and Symbolic Behavior Scales Developmental Profile (CSBS-DP). The Diagnostic Content Validity Model by Fehring, (1986; 1994) was adapted to make the selection. Afterwards, the items, taken from these tests, were selected to fit into each construct, using the same methodology. **Results:** 12 of the 18 constructs were selected by the experts and 11 items were chosen from a total of 130, reduced to eight after eliminating tautologies. **Conclusions:** Mapping these constructs and items on to the DSM-5 diagnostic criteria for ASD indicated good face and content validity. Results of this research will contribute to efforts to improve early ASD screening instruments.

Keywords

Autism; Screening tools; Early detection; Europe; Instrumental study

Key Practitioner Message:

- The study outlines the main screening instruments used across Europe and the need to improve early detection of ASD
- A novel approach, based on content validity, was used to identify the key behaviours that experts in ASD see as the most relevant for early detection
- Results of this study can help practitioners improve their understanding of ASD in early ages and be more confident in their screening results, while also serving as a theoretical framework to develop more effective screening tests.

Introduction

The emphasis on early detection of Autism Spectrum Disorders (ASD) has steadily increased as earlier detection has been shown to lead to better outcomes for children with ASD (Robins et al., 2016). Universal screening, which is the screening of the whole population, has become common in many countries (García-Primo et al., 2014) as it provides a systematic way of identifying ASD at an early stage. However, its effectiveness has been criticized (AHRQ, 2015) as there is insufficient research to assess the benefits and harms of screening the general population. That said, universal screening remains a valuable resource to continue developing (Robins et al., 2016). It is important to realize that any small improvement in the universal screening process will create a large impact when dealing with the entire population.

To improve the universal screening process, the focus over the last couple of decades has been on the development of the screening instruments. The initiative started in Europe with the Checklist for Autism in Toddlers (CHAT) (Baron-Cohen, Allen, & Gillberg, 1992), and since then more than 20 screening instruments aimed at prospectively identifying children with ASD have been developed and made available internationally (see Charman and Gotham, 2013; Charman, 2014; García-Primo et al., 2014; Yoo, 2016, for reviews). However, the instruments are not yet at an optimal level and efforts to improve them have led to limited benefits.

This study argues that another approach is needed to improve the state of the art from its current level. The aim of this study is to go back to the basics of instrument development and concentrate on developing the tools with validity in mind.

Test validity is used to qualify the appropriateness of a test for a specific goal. This can be done using a number of different processes. It is important to cover all types of validity to ensure the test is suitable. Cronbach and Meelk (1955) defined four types of validity: content, construct, predictive and concurrent (the last two may be considered as criterion-oriented validity).

In this study, the focus was on content validity, as among the main screening tests in early ASD, there are a wide range of constructs claiming to represent ASD. This reflects the challenge of determining what constructs are essential for discriminating between children who are at risk for ASD and other children. This is especially challenging when the focus is on younger ages, such as toddlers, where ASD can be confused with other developmental disorders (DD).

The importance of content validity derives from a correct use of the assessment instruments (Haynes, Richard & Kubany, 1995). If the test does not have content validity, it can misrepresent a subject's actual risk of ASD, by accentuating the influence of a particular factor underlying ASD, or by undervaluing or completely ignoring another.

A study on the First Year Inventory (FYI), for example, showed that some of the items and constructs were not autism-specific (Watson, Baranek, Crais, Reznick, Dykstra & Perryman, 2007). Children with ASD had higher mean scores than children with DD and typical development (TD) on *Social orienting* and *Receptive communication*, *Social affective engagement*, and *Reactivity*, but children with ASD and other DD performed similarly on most items concerning *Imitation* and *Expressive communication*. Similarly, although the Early Screening of Autistic Traits (ESAT) correctly identified children with ASD in a population screening of 14-15 month-old children, it also identified children with language disorders and intellectual disability (Dietz, Swinkels, Van Daalen, Van Engeland, & Buitelaar, 2006). It is important to note that no false positive cases were found to be with TD in the study, so the constructs may be related to DD in general, instead of ASD specifically.

A concern when looking at content validity is that it is subjective in nature. The content validity of an instrument is largely related to the opinions of the person or people performing the validation. Appendix A provides an overview of the constructs that form the basis of the screening instruments included in this study. These constructs are not identical across instruments, although there is some overlap between them. This could reflect a difference in opinion between the people who developed the instruments, or could be related to a broader change in the understanding of ASD over time.

For example, the CHAT, as the first screening instrument for children as young as 18 months old, was based on findings from experimental psychology and on a concept of autism that is essentially characterized by the lack of typical social competences (Baron-Cohen et al., 1992). In recent years, the development of ASD screening instruments for young children has been guided by a range of findings from retrospective studies on children with ASD, prospective studies on younger siblings of people with ASD who are at high-risk of developing the disorder, as well as clinical experiences (Zwaigenbaum, Bryson, & Garon, 2013). Thus, new instruments also consider the early presence of atypical behaviours such as stereotypies, which were not in the original CHAT. No matter the cause, the differences show that there is no unified theory regarding the constructs that define risk for autism in young children.

This study aims to clearly define what experts from around Europe agree are the main factors in early ASD. Many advances in the field of autism have been made possible through networking and research collaboration that involves joint collection, or sharing of data, most notably in genetic and baby sibling studies (Lajonchere and AGRE Consortium, 2010; Miles, 2011; Ozonoff et al., 2015; Werling and Geschwind, 2015). An interdisciplinary network, Enhancing the Scientific Study of Early Autism (ESSEA), was made possible in Europe by a COST Action (European Cooperation in Science and Technology) funded by the European Science Foundation from 2010 to 2014. This action has brought together more than 80 scientists from 23 European

countries. It was comprised of four working groups, one of which focuses on screening instruments for prospectively identifying autism.

The present study builds upon the European instruments which were identified in García-Primo et al. (2014) which, in itself, was the first result of the collaboration within this working group. By using this network of European experts, the present study is able to counteract the subjectivity of content validity through coming to an agreement between experts. Furthermore, this study strives to understand which, of all the items in all the instruments included in this study, are the best for measuring the chosen constructs. Thus, a theoretical structure of early ASD is created.

The aim of this study was to provide a general vision of what the screening instruments are actually measuring by using face and content validity. The study examined what early signs and symptoms of ASD are measured by the screening instruments used in European studies within an age range from 14-36 months, and what constructs from these instruments best represent early autism.

Methods

Participants

Experts in ASD from nine European countries, members of the COST-ESSEA Action, with completed or ongoing screening studies, were invited to participate in this study. Belgium, Finland, France, Italy, The Netherlands, Spain and the United Kingdom agreed to participate. Eight experts from these seven countries were chosen to collaborate. As all participants were members of the project, no ethical approval was needed for this study.

Instruments

After a review of the screening programs from the seven participant countries, only the instruments that were specific for ASD screening and applied in the age range of 14-36 months were selected. These were: Checklist for Early Signs of Developmental Disorders, **CESDD** (Dereu et al., 2010); Checklist for Autism in Toddlers, **CHAT** (Baron-Cohen et al., 1992); Early Screening of Autistic Traits Questionnaire, **ESAT**, the 14-item version (Dietz et al., 2006); Modified Checklist for Autism in Toddlers, **M-CHAT**, the 23-item version (Robins, Fein, Barton & Green, 2001), and the Social Communication Questionnaire, **SCQ**, the current version (Rutter, Bailey & Lord, 2003). The Communication and Symbolic Behavior Scales Developmental Profile, **CSBS-DP** was included, even though it is not specific for ASD, as its screening of communication

and symbolic behaviors in young children (Wetherby, Allen, Cleary, Kublin & Goldstein, 2002) taps into important behaviors related to autism. The items from the general categories relating specifically to language, *sounds* and *words*, were not used in this study. The Brief Infant-Toddler Social Emotional Assessment, BITSEA (Briggs-Gowan and Carter, 2006); Child Behavior CheckList, CBCL (Achenbach and Rescorla, 2001); and Infant Characteristics Questionnaire, ICQ (Bates, Freeland & Lounsbury, 1979) were excluded from our study because they were not recognized as specific screening tests for ASD by the literature (García-Primo et al., 2014). The First Year Inventory, FYI (Reznick, Baranek, Reavis, Watson & Crais, 2007) was taken out of the study as it is recommended for earlier ages (11-13 months).

Statistical analyses were performed using IBM SPSS software version 20.

Procedure

Selection of the experts

Following the guidelines from Grant and Davis (1997) and Levin (2001) for the criteria of expertise selection, to be considered an expert, someone has to have: a) a history of publications in refereed journals; b) a number of national presentations; c) relevant research on the phenomenon under study; d) a clinical practice (expertise); and also has to e) be providing direct care to populations who exhibit the phenomenon under study.

Taking into account these criteria, experts were asked to participate in the surveys of the study. At least one expert from each participant country was chosen, and they were given the freedom to propose other experts to participate. The results were a total of 8 experts willing to collaborate.

Face and content validity

To approach the task of identifying the best items in the instruments, a process of face validity was used. A test item has acceptable face validity when it appears to measure the underlying construct (Anastasi and Urbina, 1997).

In this study, content validity refers to the constructs that are chosen as most representative of early autism. Face validity refers to the representativeness of the items in measuring these constructs.

Selection of the constructs

Based on the constructs on which the selected tests were built (see Appendix A), the first constructs table was set up (see Appendix B). The experts were asked to evaluate the representativeness of the constructs in relation to ASD, given the following instructions for filling out the table: a) choose which constructs are most adequate to define early autism, even when they overlap or are very broad; b) for each construct, indicate one category from the representativeness column; c) add other constructs if you consider that important (blank rows were added for this purpose). The criteria for the selection of constructs were derived from the Diagnostic Content Validity Model by

Fehring, (1986; 1994) and the modifications given by Sparks and Lien-Gieschen (1994).

First, each characteristic is rated on a 5-point scale and each rating is then assigned a weight: a) Not Representative = 0; b) Poorly Representative = .25; c) Somewhat Representative = .50; d) Quite Representative = .75; e) Very Representative = 1.

Secondly, the mean score for each characteristic is calculated, truncated to two decimal places. This mean represents the Diagnostic Content Validity (DCV). Lastly, the DCV scores are interpreted: a) discard all constructs with a score of .60 or below; b) major constructs are those with a score between .80 and 1; c) minor constructs are those with a score between .60 and .79.

Once the construct tables were received from all the experts, the DCV scores for each construct were calculated. During the process, the experts raised concerns about the overlap of the constructs. Although the initial instructions explained that the constructs in the table were put together from the selected tests and would naturally have overlap, the results reflected the same concerns that the experts had flagged. For this reason, a second constructs table (see Appendix C) was drawn up after discussion and agreement with the experts (explained in the discussion section), aimed at reducing the overlapping and misinterpretation between constructs. Afterwards, a second round of evaluations was performed with the same experts and the DCV scores were calculated anew.

Due to the high scores, the above criteria were adapted as follows: a) discard all constructs with a score of .74 and below; b) major constructs are those with a score from .85 and 1; c) minor constructs are those with a score from .75 to .84. The cut-off was proposed at .75 because it corresponds to a score of “Quite Representative” and reduces the number of constructs included in the model. This was done to make the model simpler and more clinically useful. If the cut-off used by Sparks and Lien-Gieschen (1994) were applied there would be 10 minor constructs and six major constructs, giving a model with a total of 16 constructs. Only two constructs from the 18 initial constructs would be discarded.

Selection of the items

Continuing from the above process, an “items table” was designed for the study – and can be obtained from the original authors – with all the items from the screening tests (130) in random order on one axis and the selected constructs on the other. The selected experts were asked to assign items to a construct and indicate their representativeness. The instructions for filling in the table were: a) choose for each item one category from the constructs column and also indicate its representativeness of that construct; b) the construct “Other” is a category that can be used if none of the current constructs are a good category to group the item into.

Three variables, "Constructs Mode", "Percentage of Agreement" and “Representativeness”, were calculated using the evaluation of all experts for each item.

“Constructs Mode” was the construct chosen the most for each item and “Percentage Agreement” was the proportion of experts that chose that construct. Representativeness was calculated using the same criteria as in the constructs selection, except a value of zero was given to the scores from experts that disagreed with the construct that was selected the most. The cut-off was set at 100% for agreement and .90 for representativeness, as these values were seen to be the ones that best represented the model, taking into account the average and standard deviation of the analyzed data. Also, it was thought that the higher agreement between experts would produce the best solution.

After selection of the items, those that had the same meaning as another item, and had lower values of agreement and representativeness, were discarded.

Results

Constructs selection

Twelve constructs selected by the experts scored .75 or above (see Table I) from a total of 18. The six constructs discarded with a score of .74 or below were: *Emotion* and *Eye gaze*; *Object use* (symbolic and constructive play with objects) *other than Pretend play*; *Gestures*; *Understanding*; *Emotional reaction*; and *Motor abnormalities*. A score between .85 and 1 was considered as a Major construct (*Social interaction/ Social interchange*; *Interest in others*; *Joint-attention – other than Proto-declarative pointing*). The Minor constructs were those with a score between .75 and .84 (see table).

[Insert Table I here]

Item selection

Eleven items were selected with 100% agreement and .90 representativeness (see Table II) from a total of 130 items. Three items, ESAT10, MCHAT18 and SCQ15, were eliminated because of duplicate meaning, leaving a total of eight items.

[Insert Table II here]

With these items a theoretical model was built based on the DSM-5 diagnostic criteria for ASD (American Psychiatric Association, 2013) (See Figure 1). Seven of the 12 constructs rated as best representing early autism mapped onto domain A, social communication and social interaction, and five onto domain B, restricted, repetitive patterns of behavior, interests, or activities. Our stringent cut-off criteria for agreement and representativeness resulted in no item assigned to four constructs (*social interaction, communication, abnormal language, social play*). The model, however, includes screening test items contributing to all three criteria in domain A, and all but one criterion in domain B, *insistence on sameness*, indicating good face and content validity of the selected items.

[Place Figure 1 here]

Discussion

Many ASD screening instruments have been developed to help prospectively identify children with autism at an early age. This study included instruments used in ASD screening studies in Europe. The aim was to provide a general view of what these instruments are measuring, and which of their constructs best represent risk for early autism. Six instruments were selected for the purpose of this study. They were based on 18 constructs and contained a total of 130 items.

Eight items and 12 constructs were identified as best describing early signs of autism. The 12 constructs were chosen by the eight experts following an adaption of the DCV model (Fehring, 1986), discarding all constructs with a score of 0.74 or below. The 12 items were selected with an agreement of 100% that the item belonged to the same construct and a value of representativeness greater or equal to 0.90.

It is not surprising that the constructs receiving the highest scores, and considered as Major constructs, are *Social interaction*, *Interest in others* and *Joint attention*. Studies on early signs of ASD indicate that deficits in these behaviors are among the first symptoms to appear in young children who are later diagnosed with ASD, along with atypical eye contact (see Mitchell, Cardy & Zwaigenbaum, 2011; Paul, Loomis, & Chawarska, 2014; Zwaigenbaum et al., 2013, for reviews) that was rated as one of the minor constructs in this study. Some of the discarded constructs,

such as motor abnormalities, have not been found to be specific for ASD, but are also seen in children with other DD. A review of studies that have compared children with ASD and other DD showed that no motor behaviors were found to specifically discriminate between them in the first year of life, and findings are inconclusive when comparing these groups at 2 years of age (Mitchell et al., 2011). The evaluation made by the experts in this study is based on current knowledge of early signs of ASD, and even though some constructs were discarded, this does not mean that they are not present in young children with that condition, but only that they were not judged as most adequately defining the early signs.

Screening for ASD only identifies risks or behaviors indicative of the condition that should lead to further assessment. Thus, a direct comparison between screening test items and diagnostic criteria has to be done with caution, especially with regards to a young population. The fact that the screening instruments in this study (with the exception of the SCQ) focus on identifying early autism may have resulted in differences to the DSM-5 diagnostic constructs. For example, when assessing many young children who are later diagnosed with ASD, the construct related to speech and peer relationships may not yet be relevant. The same applies to adherence to routines and ritualized behavior that are often ambiguous in younger children or may emerge later (Guthrie, Swineford, Nottke, & Wetherby, 2013; Mitchell et al., 2011).

One of the requirements for an ASD diagnosis according to DSM-5 is persistent deficits across multiple contexts in all three social-communication symptom categories, and two of four restricted and repetitive categories (American Psychiatric Association, 2013). Research on the diagnostic validity of the DSM-5 for toddlers has shown that it has resulted in fewer children diagnosed with ASD compared with DSM-IV-TR (Kulage, Smaldone, & Cohn, 2014; Worley and Matson 2012). As young children with ASD may not yet present the full pattern of behaviors, or their symptoms may not be very clear, a more relaxed diagnostic threshold for toddlers on the DSM-5 has been suggested to ensure that most children with ASD are correctly identified (Barton, Robins, Jashar, Brennan, & Fein, 2013; Worley and Matson 2012).

A concern in this study was the overlapping and lack of definition of the constructs. It became apparent that several screening tests did not have clear definitions of the constructs they measure. Moreover, many of the items within these screening tests measure behaviors without further empirical analyses (e.g. factor analyses) that link them to a theoretical framework. It is difficult to know whether constructs such as *Social Interaction* used in the SCQ, or *Social Interchange* used in the M-CHAT, are the same concept or define different behaviors. The same problem exists with constructs like *Communication* (from the CSBS-DP and the SCQ), *Early language and Communication* (from the M-CHAT) or *Verbal and non-verbal communication* (from the ESAT).

Another drawback was the fact that some of the smaller constructs within the instruments could be grouped into bigger ones. For example, *Eye contact* can be grouped into the construct *Verbal and non-verbal communication* (ESAT), or *Proto-declarative pointing* could be a subcategory within *Joint-attention* (CHAT).

To solve this confusion, it was agreed among all the experts that the constructs *Social Interaction* and *Social Interchange* were the same, as well as *Sensory Abnormalities* and *Reaction to sensory stimuli* (see Appendix A). *Communication* and *Verbal and non-verbal communication* were defined as the same construct, excluding from the definition those behaviors that were *Gestures* (meaning non-verbal communication using conventional and symbolic gestures). *Joint attention* was defined excluding *Proto-declarative pointing*; *Object use* (symbolic and constructive play with objects) was defined as behavior other than *Pretend Play*. This process helped to clarify the similarities and differences between tests for the experts in this study, and it could be beneficial for the practitioners across Europe applying these tools with the objective of identifying risk of ASD at an early age. Currently, practitioners and researchers cannot effectively compare their results with other instruments because it is difficult to know if the different screening tools are measuring the same behaviors.

This study of operational definitions of constructs was important, and it is important to continue performing this exercise. The indices of content validity can be expected to change over time, and thus it has been recommended that content validity of

instruments be periodically examined to reflect revision in the targeted constructs (Haynes et al., 1995).

The number of experts selected for development and validation of instruments does not need to be a fixed size (Grant and Davis, 1997). For the purpose of the present study, 8 experts participated. There are examples of recent studies that have used the DCV model and have involved from as few as 4 experts (Schulz, Lopes, Herdman, Lopes, & Barros, 2013) to over 200 experts (Paloma-Castro et al., 2013). The decision on the number of experts for content validation depends on the desired expertise and range of representation among them (Grant and Davis, 1997). A limitation of our study is that no quantitative analysis was made on the background of the experts participating. However, their selection was based on qualitative analysis taking into account the main criteria defined by Grant and Davis (1977), such as their leading role in screening studies in their respective countries and participation in the COST-ESSEA Action (see Expert Selection in the Method section).

The use of content experts is recognized in instrument development and validation studies (Grant and Davis, 1997; Kassam-Adams, Marsac, Kohser, Keady, March & Winston, 2015; Levin, 2001; Shek and Yu, 2014), but the approach taken in this study, where a DCV model was applied for selection of constructs (Fehring, 1986; Sparks and Lien-Gieschen, 1994), has not been used before in studies on ASD screening instruments. Bringing this more structured technique to the ASD research community

should facilitate the instrument development process and allow for more collaborative efforts towards building more effective screening tools.

To conclude, 12 constructs and eight items were identified as best representing signs of early autism. The resulting model, and the processes used to create it, should be seen as an important step in creating a more effective early ASD screening instrument. This study hopes to contribute to a better understanding of the validity process and improve the theoretical base of ASD screening instruments.

Acknowledgments

We wish to thank all ESSEA COST members and experts that have collaborated to make this study possible.

Study Funding

This work was supported by a Short Term Scientific Mission (STSM) grant from COST Action BM1004; and the State Program of Research, Development and Innovation Oriented towards Social Challenges from The Ministry of Economy and Competition [grant number PSI2013-47840-R].

Conflicts of interest

None.

Data from this study has been presented in the XVII National Congress AETAPI (Spanish Association for Autism Professionals), Barcelona, Spain, in November 2014 and the COST ESSEA Conference, Toulouse, France in September, 2014.

References

- Achenbach, T. M., & Rescorla, L. A. (2001). The manual for the ASEBA school-age forms & profiles. Burlington, VT: University of Vermont, Research Center for Children, Youth, and Families.
- Agency for Healthcare Research and Quality (AHRQ), Vanderbilt Evidence-based Practice Center, McPheeters, M. L., Weitlauf, A., Vehorn, A., Taylor, C., Sathe, N. A., Krishnaswami, S., Fonnesebeck, C. & Warren, Z. E. (2015). Screening for autism spectrum disorder in young children: A systematic evidence review for the U.S. Preventive Services Task Force. AHRQ Publication No. 13-05185-EF-1. <http://www.uspreventiveservicestaskforce.org/Page/Document/draft-evidence-review106/autism-spectrum-disorder-in-young-children-screening>. Accessed October 8 2016.
- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5)*. Washington, DC: American Psychiatric Publishing.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Upper Saddle River, NJ: Prentice Hall.
- Baron-Cohen, S., Allen, J., & Gillberg, C. (1992). Can autism be detected at 18 months? The needle, the haystack, and the CHAT. *The British Journal of Psychiatry: The*

Journal of Mental Science, 161, 839–843.

<http://dx.doi.org/10.1192/bjp.161.6.839>

Barton, M. L., Robins, D. L., Jashar, D., Brennan, L., & Fein, D. (2013). Sensitivity and specificity of proposed DSM-5 criteria for autism spectrum disorder in toddlers.

Journal of Autism and Developmental Disorders, 43(5), 1184-1195.

<http://dx.doi.org/10.1007/s10803-013-1817-8>

Bates, J. E., Freeland, C. A., & Lounsbury, M. L. (1979). Measurement of infant difficultness. *Child Development*, 50, 794-803. Retrieved from

<http://www.jstor.org/stable/1128946>

Briggs-Gowan, M. J., Carter, A. S., Bosson-Heenan, J., Guyer, A. E., & Horwitz, S. M. (2006). Are infant-toddler social-emotional and behavioral problems transient?

Journal of the American Academy of Child and Adolescent Psychiatry, 45(7),

849-858. <http://dx.doi.org/10.1097/01.chi.0000220849.48650.59>

Charman, T., & Gotham, K. (2013). Screening and diagnostic instruments for autism spectrum disorders: lessons from research and practice. *Child and Adolescent*

Mental Health, 18, 52-64. <http://dx.doi.org/10.1111/j.1475-3588.2012.00664.x>

Charman, T. (2014). Early identification and intervention in autism spectrum disorders:

Some progress but not as much as we hoped. *International Journal of Speech-Language Pathology*, 16(1), 15–18.

<http://doi.org/10.3109/17549507.2013.859732>

- Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological bulletin*, 52(4), 281. <http://dx.doi.org/10.1037/h0040957>
- Dereu, M., Warreyn, P., Raymaekers, R., Meirsschaut, M., Pattyn, G., Schietecatte, I., & Roeyers, H. (2010). Screening for autism spectrum disorders in Flemish day-care centres with the checklist for early signs of developmental disorders. *Journal of Autism and Developmental Disorders*, 40(10), 1247–1258. <http://dx.doi.org/10.1007/s10803-010-0984-0>
- Dietz, C., Swinkels, S., van Daalen, E., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum disorder in children aged 14-15 months. II: population screening with the Early Screening of Autistic Traits Questionnaire (ESAT). Design and general findings. *Journal of Autism and Developmental Disorders*, 36(6), 713–722. <http://dx.doi.org/10.1007/s10803-006-0114-1>
- Fehring, R. J. (1986). Validation. In M. E. Hurley (Ed.), *Classification of nursing diagnoses: proceedings of the sixth conference* (pp. 183- 190). St.Louis: Mosby.
- Fehring, R. J. (1994). The Fehring model. In R.M.Carroll-Johnson & M. Paquette (Eds.), *Classification of nursing diagnoses: proceedings of the tenth conference* (pp. 55 - 62). Philadelphia: Lippincott.
- García-Primo, P., Hellendoorn, A., Charman, T., Roeyers, H., Dereu, M., Rogé, B., Canal-Bedia, R. (2014). Screening for autism spectrum disorders: state of the art

in Europe. *European Child & Adolescent Psychiatry*, 1–17.

<http://dx.doi.org/10.1007/s00787-014-0555-6>

Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in nursing & health*, 20(3), 269-274.

Doi:10.1002/(SICI)1098-240X(199706)20:3<269::AID-NUR9>3.0.CO;2-G

Guthrie, W., Swineford, L. B., Nottke, C., & Wetherby, A. M. (2013). Early diagnosis of autism spectrum disorder: stability and change in clinical diagnosis and symptom presentation. *Journal of Child Psychology and Psychiatry*, 54(5), 582-590. <http://dx.doi.org/10.1111/jcpp.12008>

Haynes, S. N., Richard, D. C. S. & Kubany, E. S. (1995). Content validity in psychological assessment: a functional approach to concepts and methods. *Psychological Assessment*, 7, 238–247. <http://dx.doi.org/10.1037/1040-3590.7.3.238>

Kassam-Adams, N., Marsac, M. L., Kohser, K. L., Kenardy, J. A., March, S., & Winston, F. K. (2015). A new method for assessing content validity in model-based creation and iteration of eHealth interventions. *Journal of medical Internet research*, 17(4). <http://dx.doi.org/10.2196/jmir.3811>

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.

<http://dx.doi.org/10.1007/s10803-014-2065-2>

Lajonchere, C. M. (2010). Changing the landscape of autism research: the autism genetic resource exchange. *Neuron*, 68(2), 187-191.

<http://dx.doi.org/10.1016/j.neuron.2010.10.009>

Levin, R. (2001). Who are the experts? A commentary on nursing diagnosis validation studies. *Nursing Diagnosis*, 12(1), 29-32. <http://dx.doi.org/10.1111/j.1744-618X.2001.tb00114.x>

Miles, J. H. (2011). Autism spectrum disorders – a genetics review. *Genetics in Medicine*, 13(4), 278-294. <http://dx.doi.org/10.1097/GIM.0b013e3181ff67ba>

Mitchell, S., Cardy, J. O., & Zwaigenbaum, L. (2011). Differentiating autism spectrum disorder from other developmental delays in the first two years of life.

Developmental Disabilities Research Reviews, 17(2), 130-140.

<http://dx.doi.org/10.1002/ddrr.1107>

Ozonoff, S., Young, G. S., Landa, R. J., Brian, J., Bryson, S., Charman, T., Chawarska, K., Macari, S. L., Messinger, D., Stone, W. L., Zwaigenbaum, L & Iosif, A. M. (2015). Diagnostic stability in young children at risk for autism spectrum disorder: a baby siblings research consortium study. *Journal of Child Psychology and Psychiatry*, 56(9), 988-998.

<http://dx.doi.org/10.1111/jcpp.12421>

- Paloma-Castro, O., Romero-Sanchez, J. M., Paramio-Cuevas, J. C., Pastor-Montero, S. M., Castro-Yuste, C., Frandsen, A. J., Albar-Marín, M. J., Bas-Sarmiento, P., & Moreno-Corral, L. J. (2014). Nursing diagnosis of grieving: content validity in perinatal loss situations. *International Journal of Nursing Knowledge*, 25(2), 102-109. <http://dx.doi.org/10.1111/2047-3095.12015>
- Paul, R., Loomis, R., & Chawarska, K. (2014). Adaptive behavior in toddlers under two with autism spectrum disorders. *Journal of autism and developmental disorders*, 44(2), 264-270. <http://dx.doi.org/10.1007/s10803-011-1279-9>
- Reznick J.S., Baranek G.T., Reavis S., Watson L.R., & Crais E.R. (2007). A parent-report instrument for identifying one-year-olds at risk for an eventual diagnosis of autism: the first year inventory. *Journal of Autism and Developmental Disorders*, 37, 1691-1710. <http://dx.doi.org/10.1007/s10803-006-0303-y>
- Robins, D. L., Fein, D., Barton, M. L., & Green, J. A. (2001). The Modified Checklist for Autism in Toddlers: an initial study investigating the early detection of autism and pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 31(2), 131–144. <http://dx.doi.org/10.1023/A:1010738829569>
- Robins, D. L., Adamson, L. B., Barton, M., Connell Jr, J. E., Dumont-Mathieu, T., Dworkin, P. H., Fein, D., Greenstein, M. A., Hsu, H. W., Kerns, C., Newschaffer, C., Plumb, J., Shattuck, P., Turchi, R. & Vivanti, G. (2016).

- Universal Autism Screening for Toddlers: Recommendations at Odds. *Journal of autism and developmental disorders*, 46(5), 1880-1882.
<http://dx.doi.org/10.1007/s10803-016-2697-5>
- Rutter, M., Bailey, A. & Lord, C. (2003) Social Communication Questionnaire (SCQ). *Western Psychological Services*.
- Schulz, C. E., Lopes, C. T., Herdman, T. H., Lopes Jde, L., & de Barros, A. L. (2014). Construction and validation of an instrument for assessment of the nursing diagnosis, risk for infection, in patients following cardiac surgery. *International Journal of Nursing Knowledge*, 25(2), 94-101. <http://dx.doi.org/10.1111/2047-3095.12018>
- Shek, D. T. L., & Yu, L. (2014). Construct Validity of the Chinese Version of the Psycho-Educational Profile-3rd Edition (CPEP-3). *Journal of Autism and Developmental Disorders*, 44(11), 2832–2843.
<http://dx.doi.org/10.1007/s10803-014-2143-5>
- Sparks, S. M., & Lien-Gieschen, T. (1994). Modification of the diagnostic content validity model. *Nursing Diagnosis: ND: The Official Journal of the North American Nursing Diagnosis Association*, 5(1), 31–35.
<http://dx.doi.org/10.1111/j.1744-618X.1994.tb00365.x>
- Swinkels, S. H., Dietz, C., van Daalen, E., Kerkhof, I. H., van Engeland, H., & Buitelaar, J. K. (2006). Screening for autistic spectrum in children aged 14 to 15

- months. I: the development of the Early Screening of Autistic Traits Questionnaire (ESAT). *Journal of Autism and Developmental Disorders*, 36(6), 723-732. <http://dx.doi.org/10.1007/s10803-006-0115-0>
- Watson, L. R., Baranek, G. T., Crais, E. R., Reznick, S. J., Dykstra, J., & Perryman, T. (2007). The first year inventory: retrospective parent responses to a questionnaire designed to identify one-year-olds at risk for autism. *Journal of Autism and Developmental Disorders*, 37(1), 49-61. <http://dx.doi.org/10.1007/s10803-006-0334-4>
- Werling, D. M., & Geschwind, D. H. (2015). Recurrence rates provide evidence for sex-differential, familial genetic liability for autism spectrum disorders in multiplex families and twins. *Molecular autism*, 6(1), 1. <http://dx.doi.org/10.1186/s13229-015-0004-5>
- Wetherby, A. M., Allen, L., Cleary, J., Kublin, K. & Goldstein, H. (2002). Validity and Reliability of the Communication and Symbolic Behavior Scales Developmental Profile With Very Young Children. *Journal of Speech, Language, and Hearing Research*, 45(6), 1202–1218. [http://dx.doi.org/10.1044/1092-4388\(2002/097\)](http://dx.doi.org/10.1044/1092-4388(2002/097))
- Worley, J. A., & Matson, J. L. (2012). Comparing symptoms of autism spectrum disorders using current DSM-IV-TR diagnostic criteria and the proposed DSM-5 diagnostic criteria. *Research in Autism Spectrum Disorders*, 6, 965-970. <http://dx.doi.org/10.1016/j.rasd.2011.12.012>

- Yoo, H. (2016). Early detection and intervention of autism spectrum disorder. *Hanyang Medical Reviews*, 36(1), 4-10. <http://dx.doi.org/10.7599/hmr.2016.36.1.4>
- Zwaigenbaum, L., Bryson, S., & Garon, N. (2013). Early identification of autism spectrum disorders. *Behavioural Brain Research*, 251, 133-146. <http://dx.doi.org/10.1016/j.bbr.2013.04.004>

Table I. Constructs selection

Constructs	N	Mean
1. Social interaction/ Social interchange	8	1.000
2. Interest in others	8	.906
3. Joint-attention – other than Proto-declarative pointing	8	.906
4. Social play	8	.813
5. Proto-declarative pointing	8	.813
6. Stereotyped behaviour	8	.813
7. Communication/ Verbal and non-verbal communication – other than Gestures (non-verbal communication using conventional and symbolic gesture)	8	.781
8. Sensory abnormalities/ Reaction to sensory stimuli	8	.781
9. Preoccupations	8	.750
10. Eye contact	8	.750
11. Abnormal language	8	.750
12. Pretend play	8	.750
13. Emotion and eye gaze	8	.719
14. Object use (symbolic and constructive play with objects) - other than Pretend play	8	.719
15. Gestures	8	.719
16. Understanding	8	.625
17. Emotional reaction	8	.500
18. Motor abnormalities	8	.438
Valid N (listwise)		8

Table II. Items selection

Constructs Mode	Percentage of Agreement	Representativeness	Item Test
Interest in others	100	1.000	CHAT & MCHAT 2
Proto-declarative pointing	100	1.000	CHAT & MCHAT 7
Preoccupations	100	.969	SCQ 11
Interest in others	100	.969	ESAT10
Sensory abnormalities	100	.969	MCHAT 11
Stereotyped behaviour	100	.938	ESAT 8
Joint attention	100	.938	CESDD 15
Stereotyped behaviour	100	.938	SCQ 15
Pretend play	100	.938	CSBS-DP 24
Eye contact	100	.906	MCHAT 10
Stereotyped behaviour	100	.906	MCHAT 18

Figure 1. Theoretical model based on the DSM-5 criteria for ASD

Appendix A. References of constructs.

	AUTHORS AND YEAR OF PUBLICATION	CONSTRUCTS
CHAT	Baron-Cohen, Allen, and Gillberg, (1992)	Pretend play, Proto-declarative pointing, Joint attention, Social interest and Social play.
M-CHAT	Robins, Fein, Barton, and Greene, (2001)	Sensory abnormalities, Motor abnormalities, Social interchange, Early joint attention/ Theory of mind and Early language and communication.
CESDD (Target behaviours, not constructs)	Dereu, Warreyn, Raymaekers, Meirsschaut, Pattyn, Schietecatte, & Roeyers, (2010)	
CSBS-DP	Wetherby, Allen, Cleary, Kublin, and Goldstein, (2002)	Emotion and eye gaze, Communication, Gestures, Sounds, Words , Understanding and Object use (symbolic and constructive play).
ESAT	Swinkels, Dietz, van Daalen, Kerkhof, van Engeland, and Buitelaar, (2006) Dietz, Swinkels, van Daalen, van Engeland, and Buitelaar, (2006)	Pretend play, Joint attention Interest in others, Eye contact, Verbal and non-verbal communication, Stereotypes, Preoccupations, Reaction to sensory stimuli, Emotional reaction and Social interaction.
SCQ 11/15	Berument, Rutter, Lord, Pickles, and Bailey, (1999) Rutter, Bailey, and Lord, (2003)	Social interaction, Communication, Abnormal language and Stereotyped behaviour.

Appendix B. Constructs Table one, examples.

CONSTRUCTS	REPRESENTATIVENESS				
	Not Representative	Poorly Representative	Somewhat Representative	Quite Representative	Very Representative
Pretend play					
Sensory abnormalities					
Proto- declarative pointing					
(...)					

Appendix C. Constructs Table two, examples.

CONSTRUCTS	REPRESENTATIVENESS				
	Not Representative	Poorly Representative	Somewhat Representative	Quite Representative	Very Representative
Social interaction/ Social Interchange					
Proto- declarative pointing					
Joint attention – other than Proto- declarative pointing					
(...)					

