Child and Adolescent
Mental Health

Child and Adolescent Mental Health 23, No. 4, 2018, pp. 359–367 doi:10.1111/camh.12256

Building a theoretical framework for autism spectrum disorders screening instruments in Europe

María Magán-Maganto[1], Sigrídur Lóa Jónsdóttir[2], Ana B. Sánchez-García[1], Patricia García-Primo[3], Annika Hellendoorn[4], Tony Charman[5], Herbert Roeyers[6], Mieke Dereu[6], Irma Moilanen[7], Filippo Muratori[8], Manuel Posada de la Paz[3], Bernadette Rogé[9], Iris J. Oosterling[10], Anneli Yliherva[11] & Ricardo Canal-Bedia[1]

[1]*Institute of Community Integration (INICO), Faculty of Education, University of Salamanca, Salamanca, Spain*
[2]*State Diagnostic and Counselling Centre, Kópavogur, Iceland*
[3]*Institute of Rare Diseases Research, ISCIII, Madrid, Spain*
[4]*Department of Educational Sciences, Utrecht University, Utrecht, The Netherlands*
[5]*Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, UK*
[6]*Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium*
[7]*University and University Hospital of Oulu, Oulu, Finland*
[8]*Department of Clinical and Experimental Medicine, University of Pisa, IRCCS Stella Maris Foundation, Pisa, Italy*
[9]*CERPPS, Université Toulouse, IUF (Institut Universitaire de France), Toulouse, France*
[10]*Karakter Child and Adolescent Psychiatry, University Centre Nijmegen, Nijmegen, The Netherlands*
[11]*Child Language Research Center, Faculty of Humanities, Logopedics, University of Oulu, Oulu, Finland*

Background: This study addresses the need for a theoretical base to develop more effective early autism spectrum disorders (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. First, 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 Behaviour 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:** Twelve 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 and identify the key behaviours that experts in ASD see as the most relevant for early detection.

Key Practitioner Message

- The study outlines the main screening instruments used across Europe and the need to improve early detection of autism spectrum disorders (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 early ASD and be more confident in their screening results, while also serving as a theoretical framework to develop more effective screening tests.

Keywords: Autism; screening tools; early detection; Europe; face and content validity; instrumental study

Background

The emphasis on early detection of autism spectrum disorders (ASD)[1] 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

Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA 02148, USA

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) and any small improvement made during that 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 & 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 was 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 carried out using a number of different processes. It is important to cover all types of validity to ensure the test is suitable. Cronbach and Meehl (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 a wide range of constructs have been defined to represent ASD among the main screening tests for early ASD. This suggests a difficulty in 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 symptoms similar to ASD may be present in children with other developmental disorders (DD). For example, early signs of ASD like language delays are present in communication disorders and in developmental delays, and some repetitive stereotypic behaviours are also observed in children with developmental delays (Camarata, 2014; Malhi & Singhi, 2014).

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 et al., 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- to 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 somewhat 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 different findings from the researchers' analyses of the items, different samples that have been used to arrive at these differential items or a subjective selection of items from literature reviews. It could also be related to a broader change in the understanding of ASD over time, based on new knowledge generated by research.

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, 2010; Miles, 2011; Ozonoff et al., 2015; Werling & 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.

This 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. Using this network of European experts, this 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 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 to 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.

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 (meaning experts in early detection and diagnosis of ASD working on a daily basis with children with neurodevelopmental disorders) 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 eight experts willing to collaborate.

Instruments

After a review of the screening programmes 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 as follows: 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 Behaviour Scales Developmental Profile, CSBS-DP, was included, even though it is not specific for ASD, as its screening of communication and symbolic behaviours in young children (Wetherby, Allen, Cleary, Kublin, & Goldstein, 2002) taps into important behaviours 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, Carter, Bosson-Heenan, Guyer, & Horwitz, 2006); Child Behaviour CheckList, CBCL (Achenbach & 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).

Procedure

To approach the task of identifying the best constructs and items in the instruments, a process of face and content validity was used. In this study, content validity refers to the constructs that are chosen as most representative of early autism.

A test item has acceptable face validity when it appears to measure the underlying construct (Anastasi & Urbina, 1997). Here, 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. Second, 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. For example, *Eye contact* can be grouped into the construct *Verbal and nonverbal communication* (ESAT), or *Proto-declarative pointing* could be a subcategory within *Joint attention* (CHAT). 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 and aiming to reduce the overlapping and misinterpretation between constructs, a second constructs table (see Appendix C) was drawn up after discussion and agreement with 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 nonverbal communication* were defined as the same construct, excluding from the definition those behaviours that were *Gestures* (meaning nonverbal 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 behaviour other than *Pretend play*. 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 to 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 performed 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 as follows: (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 analysed 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.

Statistical analyses

Statistical analyses (means of the scores from the construct and items tables) were performed using IBM SPSS software version 20.

Results

Constructs selection

Twelve constructs selected by the experts scored .75 or above (see Table 1) from a total of 18. The six constructs discarded with a score of .74 or below were as follows: *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 1).

Item selection

Eleven items were selected with 100% agreement and .90 representativeness (see Table 2) from a total of 130 items. Three items, ESAT10, M-CHAT18 and SCQ15, were eliminated because of duplicate meaning, leaving a total of eight items.

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 behaviour, interests or activities. Our stringent cut-off criteria for agreement and representativeness resulted in no item assigned to four constructs (S*ocial 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.

Table 1. 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 nonverbal communication – other than Gestures (nonverbal 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 2. Items selection

Constructs mode	Percentage of agreement	Representativeness	Item test
Interest in others	100	1.000	CHAT & M-CHAT 2
Proto-declarative pointing	100	1.000	CHAT & M-CHAT 7
Preoccupations	100	.969	SCQ 11
Interest in others	100	.969	ESAT10
Sensory abnormalities	100	.969	M-CHAT 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	M-CHAT 10
Stereotyped behaviour	100	.906	M-CHAT 18

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.

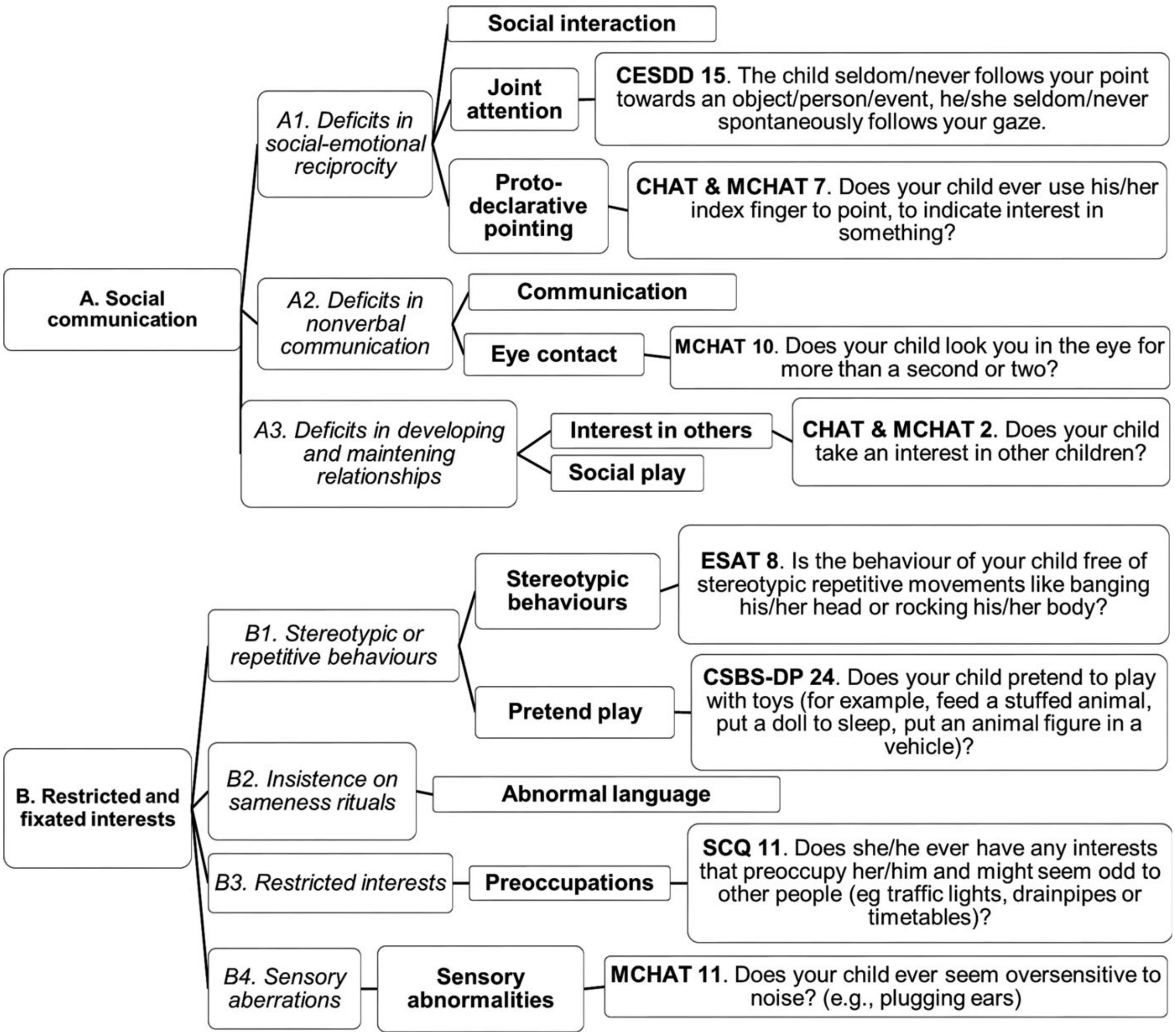

Figure 1. Theoretical model based on the DSM-5 criteria for ASD

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 11 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 behaviours 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 it 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 behaviours 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.

Moreover, the eight selected items to represent the model are the ones given higher scores in representativeness by the experts and with a 100% of agreement, but this does not mean that the behaviours described by these items are the only ones that can represent the model. For example, a child aged between 14 and 36 months is unlikely to have preoccupations with traffic lights, drainpipes or timetables, and thus, SCQ 11 item for preoccupations does not seem relevant for young children on the spectrum. However, the behaviours described by this item could be found more frequently in children with higher language and cognitive capacities (Leekam, Prior, & Uljarevic, 2011) and also tend to emerge or increase frequency from 24 months onwards (Moore & Goodson, 2003), which is in the range of the current study. Furthermore, the examples given by the item are not exclusive, and other examples can be applied, such as a preoccupation with odd objects or uncommon interests, where toddlers may obsess over collecting round objects or he/she could watch a TV show endlessly (Raulston & Machalicek, 2017).

The fact that no item was assigned to four constructs (S*ocial interaction, Communication, Abnormal language, Social play*) which represent key areas, especially within the early signs of ASD, reflects a shortcoming of the item composition of the screening instruments used in the study. Some could argue that lower values of agreement and representativeness should be chosen to add an item to the other constructs. However, to add an item to the *Communication* construct would lower the cut-off to 88% agreement and .59 representativeness for the item 'Does she/he ever use gestures, other than pointing or pulling your hand, to let you know what she/he wants?' (SCQ 23). As the representativeness value shows, the *Communication* construct is poorly represented by current screening items and this scenario would lead to the inclusion of 42 items in the model. For reference sake, the item with highest representativeness for *Abnormal language* was 'Does she/he ever get her/his pronouns the wrong way round (i.e. saying "you" or "she/he" for "I"?)' (SCQ 5) with 100% agreement and .75 representativeness. For the *Social play* construct, it was 'The child seldom/never enjoys playing games (like peek-a-boo, being swung, being tossed in the air. . .).' (CESDD 11) with 88% agreement and .78 representativeness. For the construct *Social interaction*, the item chosen would be 'Does she/he smile back if someone smiles at her/him?' (SCQ 27) with 100% agreement and .84 representativeness.

Further research with clinical samples is needed to test the model and different combinations, which is out of the scope of this study.

Screening for ASD only identifies risks or behaviours indicative of the condition that should lead to further assessment. Thus, a direct comparison between screening test items and diagnostic criteria has to be carried out with caution, especially with regards to a young population. The mapping onto DSM-5 was not performed to justify either DSM-5 or the model, but for practical reasons concerning the implementation of the model in the screening process. The fact that the screening instruments in this study (with the exception of the SCQ) focus on identifying autism in early ages 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 behaviour 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 & Matson, 2012). As young children with ASD may not yet present the full pattern of behaviours, 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 & 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 behaviours 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 behaviours. The same problem exists with constructs such as *Communication* (from the CSBS-DP and the SCQ), *Early language and Communication* (from the M-CHAT) or *Verbal and nonverbal communication* (from the ESAT).

Another drawback was the fact that some of the smaller constructs within the instruments could be grouped into bigger ones. To solve this confusion, it was agreed among all the experts what constructs overlapped or were the same, for example *Social interaction* and *Social interchange* (see methods section). 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 behaviours.

This study of operational definitions of constructs was important, and it is essential to continue performing this exercise. Specific behaviours described by the selected items should be defined more clearly with the possibility to add new behaviours that better describe the different constructs. 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 results of this research will also contribute to efforts towards a unified conceptualization of ASD screening constructs for younger ages. Currently, in other psychology areas, authors like Morrison and Grammer (2016) and Nigg (2017) agree on the importance of conceptualizing, measuring and integrating constructs through disciplinary approaches. Reviewing validity and unifying concepts should be seen as necessary steps in advancing the field of knowledge.

The number of experts selected for development and validation of instruments does not need to be a fixed size (Grant & Davis, 1997). For the purpose of the present study, eight experts participated. There are examples of recent studies that have used the DCV model and have involved from as few as four experts (Schulz, Lopes, Herdman, Lopes Jde, & de Barros, 2014) to over 200 experts (Paloma-Castro et al., 2014). The decision on the number of experts for content validation depends on the desired expertise and range of representation among them (Grant & 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 (1997), 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 & Davis, 1997; Kassam-Adams et al., 2015; Levin, 2001; Shek & Yu, 2014), but the approach taken in this study, where a DCV model was applied for selection of constructs (Fehring, 1986; Sparks & 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.

A predictive validity analysis of the items should be performed with data from a real sample to demonstrate the empirical validity of the proposed model, taking into account also the characteristics that could influence children's development and could be predictive of a future ASD diagnosis. This analysis is beyond the scope of the current study, but is necessary to develop a potential screening tool based on the model.

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.

Acknowledgements

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]. Data from this study have 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. The authors wish to thank all ESSEA COST members and experts that have collaborated to make this study possible. The authors have declared that they have no competing or potential conflict of interests.

Ethical information

As all participants were members of the project, no ethical approval was required for this study.

Correspondence

María Magán-Maganto, Institute of Community Integration (INICO), Faculty of Education, University of Salamanca, Paseo de Canalejas, 169, 37008 Salamanca, Spain. Email: mmmaria@usal.es

Note

[1]Note that ASD, autism, early autism and all related terms to Autism Spectrum Disorders will be used as interchangeable terms to refer to the same condition.

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., ... & 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. Available from http://www.uspreventiveservicestaskforce.org/Page/Document/draft-evidence-review106/autism-spectrum-disorder-in-young-children-screening [last accessed 8 October 2016].

American Psychiatric Association (2013). *Diagnostic and statistical manual of mental disorders, Fifth Edition (DSM-5)*. Washington, DC: Author.

Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th edn). 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.

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*, 1184–1195.

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

Berument, S.K., Rutter, M., Lord, C., Pickles, A., & Bailey, A. (1999). Autism screening questionnaire: Diagnostic validity. *British Journal of Psychiatry*, *175*, 444–451.

Briggs-Gowan, M.J., Carter, A.S., Bosson-Heenan, J., Guyer, A.E., & Horwitz, S.M. (2006). Are infant-toddler social-emotional and behavioural problems transient? *Journal of the American Academy of Child and Adolescent Psychiatry*, *45*, 849–858.

Camarata, S. (2014). Early identification and early intervention in autism spectrum disorders: Accurate and effective? *International Journal of Speech-Language Pathology*, *16*, 1–10.

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*, 15–18.

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.

Cronbach, L.J., & Meehl, P.E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281.

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*, 1247–1258.

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*, 713–722.

Fehring, R.J. (1986). Validation. In M.E. Hurley (Ed.), *Classification of nursing diagnoses: Proceedings of the sixth conference* (pp. 183–190). St.Louis, MO: Mosby.

Fehring, R.J. (1994). The Fehring model. In R.M. Carrol1-Johnson & M. Paquette (Eds.), *Classification of nursing diagnoses:*

Proceedings of the tenth conference (pp. 55–62). Philadelphia, PA: 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 and Adolescent Psychiatry*, *23*, 1005–1021.
Grant, J.S., & Davis, L.L. (1997). Selection and use of content experts for instrument development. *Research in Nursing and Health*, *20*, 269–274.
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*, 582–590.
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.
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*, e95.
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*, 1918–1932.
Lajonchere, C.M. (2010). Changing the landscape of autism research: The autism genetic resource exchange. *Neuron*, *68*, 187–191.
Leekam, S.R., Prior, M.R., & Uljarevic, M. (2011). Restricted and repetitive behaviours in autism spectrum disorders: A review of research in the last decade. *Psychological Bulletin*, *137*, 562.
Levin, R. (2001). Who are the experts? A commentary on nursing diagnosis validation studies. *Nursing Diagnosis*, *12*, 29–32.
Malhi, P., & Singhi, P. (2014). A retrospective study of toddlers with autism spectrum disorder: Clinical and developmental profile. *Annals of Indian Academy of Neurology*, *17*, 25–29.
Miles, J.H. (2011). Autism spectrum disorders – a genetics review. *Genetics in Medicine*, *13*, 278–294.
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*, 130–140.
Moore, V., & Goodson, S. (2003). How well does early diagnosis of autism stand the test of time? Follow-up study of children assessed for autism at age 2 and development of an early diagnostic service. *Autism*, *7*, 47–63.
Morrison, F.J., & Grammer, J.K. (2016). Conceptual clutter and measurement mayhem: Proposals for cross-disciplinary integration in conceptualizing and measuring executive function. In J.A. Griffin, P. McCardle & L.S. Freund (Eds.), *Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research* (pp. 327–348). Washington, DC: American Psychological Association.
Nigg, J.T. (2017). Annual Research Review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk-taking, and inhibition for developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *58*, 361–383.
Ozonoff, S., Young, G.S., Landa, R.J., Brian, J., Bryson, S., Charman, T., ... & 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*, 988–998.
Paloma-Castro, O., Romero-Sanchez, J.M., Paramio-Cuevas, J.C., Pastor-Montero, S.M., Castro-Yuste, C., Frandsen, A.J., ... & Moreno-Corral, L.J. (2014). Nursing diagnosis of grieving: Content validity in perinatal loss situations. *International Journal of Nursing Knowledge*, *25*, 102–109.
Paul, R., Loomis, R., & Chawarska, K. (2014). Adaptive behaviour in toddlers under two with autism spectrum disorders. *Journal of autism and developmental disorders*, *44*, 264–270.
Raulston, T.J., & Machalicek, W. (2017). Early intervention for repetitive behaviour in autism spectrum disorder: A conceptual model. *Journal of Developmental and Physical Disabilities*. Advance online publication. https://doi.org/10.1007/s10882-017-9566-9.
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.
Robins, D.L., Adamson, L.B., Barton, M., Connell Jr, J.E., Dumont-Mathieu, T., Dworkin, P.H., ... & Vivanti, G. (2016). Universal autism screening for toddlers: Recommendations at odds. *Journal of autism and developmental disorders*, *46*, 1880–1882.
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*, 131–144.
Rutter, M., Bailey, A., & Lord, C. (2003). The social communication questionnaire. *Manual. 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*, 94–101.
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*, 2832–2843.
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*, 31–35.
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*, 723–732.
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*, 49–61.
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.
Wetherby, A.M., Allen, L., Cleary, J., Kublin, K., & Goldstein, H. (2002). Validity and reliability of the communication and symbolic behaviour scales developmental profile with very young children. *Journal of Speech, Language, and Hearing Research*, *45*, 1202–1218.
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.
Yoo, H. (2016). Early detection and intervention of autism spectrum disorder. *Hanyang Medical Reviews*, *36*, 4–10.
Zwaigenbaum, L., Bryson, S., & Garon, N. (2013). Early identification of autism spectrum disorders. *Behavioural Brain Research*, *251*, 133–146.

Accepted for publication: 20 November 2017
Published online: 22 December 2017

Appendix A References of constructs

	Authors and year of publication	Constructs	Internal consistency/Concurrent validity
CHAT	Baron-Cohen et al. (1992)	Pretend play, Proto-declarative pointing, Joint attention, Social interest and Social play.	NR/NR
M-CHAT	Robins et al. (2001)	Sensory abnormalities, Motor abnormalities, Social interchange, Early joint attention/Theory of mind and Early language and communication.	Cronbach's alpha .85 (23 items); .83 (critical items)/NR
CESDD	Dereu et al. (2010)	(Target behaviours, not constructs)	NR/NR
CSBS-DP	Wetherby et al. (2002)	Emotion and eye gaze, Communication, Gestures, Understanding and Object use (symbolic and constructive play).	NR, however, test–retest consistency is described/Reported on page 1213
ESAT	Swinkels et al. (2006) Dietz et al. (2006)	Pretend play, Joint attention Interest in others, Eye contact, Verbal and nonverbal communication, Stereotypes, Preoccupations, Reaction to sensory stimuli, Emotional reaction and Social interaction.	NR/NR
SCQ 11/15	Berument, Rutter, Lord, Pickles, and Bailey (1999) Rutter et al. (2003)	Social interaction, Communication, Abnormal language and Stereotyped behaviour.	Cronbach's alpha 0.90/Reported on pages 447–448

NR, Not Reported.

Appendix B Constructs Table one, examples

	Representativeness				
Constructs	Not representative	Poorly representative	Somewhat representative	Quite representative	Very representative
Pretend play					
Sensory abnormalities					
Proto-declarative pointing					
(. . .)					

Appendix C Constructs Table two, examples

	Representativeness				
Constructs	Not representative	Poorly representative	Somewhat representative	Quite representative	Very representative
Social interaction/Social interchange					
Proto-declarative pointing					
Joint attention – other than Proto-declarative pointing					
(. . .)					