Research in Autism Spectrum Disorders 62 (2019) 55–65

Contents lists available at ScienceDirect

Research in Autism Spectrum Disorders

journal homepage: www.elsevier.com/locate/rasd

Translation and cultural adaptation of parent-report developmental assessments: Improving rigor in methodology

Michaela DuBay*, Linda R. Watson

University of North Carolina at Chapel Hill, Allied Health Sciences, 321 S Columbia St., 27516, Chapel Hill, NC, United States

ARTICLE INFO

No. of reviews completed is 3

Keywords:
Translation
Cultural adaptation
Screening
Assessment
Parent-report

ABSTRACT

Most parent-report autism screening and assessment tools are developed and validated in English. Tools developed in one language with a specific culture require translation and cultural adaptation for use with a culturally or linguistically different population. Traditional translation methods are not sufficient to maintain the validity of these tools in the new population. A more rigorous methodology that maintains linguistic, construct, and technical equivalence is recommended. This article summarizes the basic stages in this translation and cultural adaptation process along with descriptions and rationales for various possible strategies used to complete each stage. A quality appraisal checklist is provided to allow readers to systematically analyze the quality of a translation or to guide instrument developers in the translation process.

1. Introduction

Parent-report screening and assessment tools are designed to identify children who are showing symptoms associated with various developmental conditions, including autism spectrum disorder (ASD). Diagnosis of these developmental conditions requires knowledge of the way the children function in their everyday life. Because observations in a clinical setting do not always provide a valid representation of a child's daily functioning, parent report is one of the most common data collection methods for screening and assessment of developmental conditions. Most available tools were created and validated with non-minority English-speaking individuals in western countries such as the United States and England. However, the need to identify children who show signs of developmental difficulties is not unique to these populations. In fact, a resolution of the World Health Organization (World Health Organization, 2014) urged its member states to launch coordinated and comprehensive efforts to build a global capacity to meet the needs of individuals with ASD and other developmental disabilities. Building such capacity requires assessment and screening tools that are valid for use with culturally and linguistically diverse populations.

Because of the resources required to develop new tools, modifying existing tools for use with new populations is often preferred as a more feasible and cost-effective strategy (Guillemin, 1993; Ware, Gandek, & Keller, 1996). Therefore, many tools originally developed in English are being translated into new languages. Traditionally, these self-report tools are translated using a basic forward-back methodology that was originally taken from a wider set of recommendations provided by Brislin (1970). However, simple translation methods are not sufficient to begin using an instrument with a new population (Acquadro, Conway, Hareendran, & Aaronson, 2008; Gjersing, Caplehorn, & Clausen, 2010; Guillemin, 1993). Factors such as translation errors and, more importantly, cultural differences complicate the task of translating an instrument. Traditional methods do not sufficiently address these problems (Beaton, Bombardier, Guillemin, & Ferraz, 2000; Gjersing et al., 2010; Nelson McDermott & Palchanes, 1994). Culture acts as a lens

* Corresponding author.
E-mail addresses: michaela_dubay@med.unc.edu (M. DuBay), linda_watson@med.unc.edu (L.R. Watson).

https://doi.org/10.1016/j.rasd.2019.02.005
Received 1 September 2018; Received in revised form 23 February 2019; Accepted 25 February 2019
Available online 19 March 2019
1750-9467/

through which we view the world, including parenting and child development. Cultures often have distinct perspectives around when children learn certain skills, how children should interact with others, and what is valued as appropriate behavior from children (van Kleeck, 1994). Further, culture influences the way people understand and answer questions (Beaton et al., 2000). Cultural perspectives around child rearing will in turn influence parents' understanding of questions as well as their response patterns. Thus, cultural adaptation should be included in the translation process of parent-report tools. Without appropriate translation and adaptation methods, the translated version of an instrument may differ in over- or under-identification or diagnosis rates as compared to the original instrument (Soto et al., 2015), possibly resulting in a poor quality screener or assessment tool.

Self-report instruments are designed such that a layperson independently responds to a series of questions or statements about themselves or a family member. The Modified Checklist for Autism in Toddlers – Revised (M-CHAT_R/F; Robins et al., 2014), the Infant Toddler Checklist (Wetherby & Prizant, 2002), and the Social Communication Questionnaire (Rutter, Bailey, & Lord, 2003) are examples of commonly used self-report developmental screening tools designed to be completed independently by a caregiver. Self-report instruments present a specific challenge in measurement. The data collected is completely dependent on the wording of the items themselves (along with instructions and answer choices). No professional with training, expertise, or experience in the constructs of interests, is available to clarify the meaning of items as parents are answering the questions. Thus, the reliability and validity of these instruments is based solely on wording and administration method. A quality translation and cultural adaptation process is required in order to appropriately modify a self-report instrument for use with a new population (Acquadro et al., 2008; Beaton et al., 2000). Brislin (1970) first introduced a translation methodology involving back-translation, among other procedures, in the 1970s. Back-translation was then frequently adopted as the sole step necessary in the translation process, although Brislon himself specifically cautioned against this practice (1970). In this traditional methodology, one person translates the instrument into the new target language, one person takes this target translation and back-translates it into the source language, then compares the two source-language versions and resolves discrepancies. Research and clinical teams typically begin using the translation once the two source-language versions appear “equivalent,” though most research teams fail to provide a full description in their publications of the way they conceptualize equivalence.

The fields of ASD and developmental disabilities generally continue to use simple translation methods as standard practice. Specifically, minimal quality checking methods are utilized, pre-testing and re-testing of psychometric properties are frequently omitted, and little information about the process is included in peer-reviewed publications. Three recent reviews in the fields of ASD and developmental disabilities illustrate this trend. Soto et al. (2015) and Al Maskari, Melville, and Willis (2018) both reviewed translations and cultural adaptations of ASD screening tools published in peer-reviewed journals. They found limited descriptions of the translation process, frequent omissions of important steps in the process, and overall low rigor in translation and cultural adaptation methodology (Al Maskari et al., 2018; Soto et al., 2015). El-Behadli, Neger, Perrin, and Sheldrick (2015) published an evidence map analyzing the quality of 63 translations of the 9 developmental screening tools recommended by the American Academy of Pediatrics. In this review, there were no supporting peer-reviewed publications for 30% of the translations found. Information on translation methodology for these translations was often collected through personal communications or from instrument manuals. Further, researchers found no information regarding the methods of translation for almost half of the translations. A majority of translations omitted quality checking or pre-testing methods, and 11% specifically stated that culture was not considered during the translation process (El-Behadli et al., 2015). The M-CHAT-R (Robins et al., 2014), is used widely internationally and has at least 60 official translations. The website (www.mchatscreen.com) provides a recommended methodology for translation similar to traditional back-translation methods described earlier.

These findings indicate that a rigorous translation and cultural adaptation process for self-report instruments has not been widely adopted in the fields of ASD and developmental disabilities, although some recent reports suggest that the limitations of simpler, more traditional translation methods are becoming more widely recognized in these fields (Kondolot et al., 2016; Windham et al., 2014). The risk of using an instrument that is poorly adapted to a population’s culture is measurement bias. Specifically, level of risk or impairment in various areas of development, including social communication and repetitive and restrictive behaviors, may be overestimated or underestimated when using a maladapted instrument (Harrison, Long, Tommet, & Jones, 2017).

This paper seeks to present a set of guidelines written from the perspective of early child development and developmental disabilities, including ASD. While there are numerous guidelines in the literature that describe rigorous translation and cultural adaptation methodologies (see Beaton et al., 2000; Gjersing et al., 2010; Nelson McDermott & Palchanes, 1994; Sousa & Rojjanasrirat, 2011; Wild et al., 2005), none highlight the unique experience of parents and other caregivers responding to questions that relate to child behaviors and abilities, a concept that is ripe with cultural meaning. Our field is also in need of specific criteria to assess the quality of the translation and adaptation process for developmental screening and assessment tools. The three reviews described earlier used criteria that were either based only on early works (Soto et al., 2015; El-Behadli et al., 2015), or designed for intervention adaptation rather than assessment adaptation (Al Maskari et al., 2018). We propose that a quality appraisal tool, designed for developmental screening and assessment tools and based on the most up to date literature, will support research and clinical efforts in ASD and developmental disabilities by providing a means to identify the most appropriate tools for use with different culturally and linguistically diverse populations.

There are three main goals of this paper. First, we describe the various dimensions of equivalence that should be considered when adapting a parent-report developmental instrument for a new population. Second, we present clear and concise guidelines in translation and cultural adaptation methodology, synthesized from the array of recommendations that can be found across related fields such as public health and social sciences. Guidelines are organized by the key stages found in the literature (instrument reproduction, pre-testing, and psychometric testing), with descriptions and rationales for the various ways to complete each stage. Fig. A1 presents a visual summary of this process. Third, we propose a basic set of criteria, formulated as a quality appraisal tool

(Table A1), to examine the quality of translated and culturally adapted parent-report developmental screening tools and assessments.

2. Considerations in equivalence

Language plays a central role in the vast majority of self-report instruments, where a respondent answers a series of questions (written or oral). Language itself is a representation of a group's worldview, where word meanings are dependent on the speaker's understanding of each concept (Castro, Barrera, & Holleran Steiker, 2010; McKenna & Doward, 2005). Caregiver views of child development are heavily driven by culture, and the language used to probe caregiver observations of children must be interpretable within the context of their specific cultural lens and experiences. Translation must take into account this cultural lens rather than translating each word into the target language. A literal translation of words may fail to produce an instrument that has meaning to speakers of the target language (Behr, 2018; Chapman & Carter, 1979; Colina, Marrone, Ingram, & Sanchez, 2017; Nelson McDermott & Palchanes, 1994). The goal in reproducing an instrument in the target language is to maintain equivalence to the foundational concepts within the source instrument, while changing only the linguistic code, such that the same concepts are fully represented in the new cultural context (Acquadro, Jambon, Ellis, & Marquis, 1996; Gjersing et al., 2010). Thus, the term *cultural adaptation* as opposed or in addition to *translation* is used to represent the process of maintaining complete equivalence between versions (McKenna & Doward, 2005; Ware et al., 1996). Equivalence should be maintained on a variety of key dimensions to sufficiently adapt the new version to the target culture. Most descriptions of these dimensions come from either Flaherty et al. (1988) or Guillemin (1993), though the specific descriptions and terminology used by these authors differed. A synthesis of their equivalence dimensions, with the addition of descriptions from later authors, is presented here.

Linguistic equivalence consists of two levels: semantic and idiomatic. Semantic equivalence requires that the sentence structure and words of two instruments have the same meaning (Guillemin, 1993), including the conceptual definitions of the terms used (Beaton et al., 2000). Translators must consider multiple meanings of words and language differences where grammatical structures have no equivalent (Beaton et al., 2000; Guillemin, 1993). In traditional forward-back translation methodology, semantic equivalence is the primary, and sometimes only, dimension of equivalence that is considered. Idiomatic equivalence refers to the translations of colloquialisms, idioms, and expressions (Beaton et al., 2000; Flaherty et al., 1988; Sousa & Rojjanasrirat, 2011). For example, if translated literally, expressions such as "in his own world," or "reading between the lines" may become meaningless in another language. These expressions need to be adapted to maintain the underlying meaning within the target language and culture (not responding to external stimuli or interpreting unstated information, respectively). Considerations of idiomatic equivalence may be more common in items describing social or emotional states (Guillemin, 1993; Ware et al., 1996), which are likely to be included in measures of social and communication development.

There are three closely related levels of *construct equivalence*, including conceptual, experiential, and content equivalence. For conceptual equivalence, the concept that the instrument measures must exist within the target culture (Gjersing et al., 2010; McKenna & Doward, 2005; Stewart & Napoles-Springer, 2000). To maintain experiential equivalence, Guillemin (1993) states that items describing experiences, situations, or contexts should fit the target cultural context. One example is a concept included in many screening tools for ASD: "peek-a-boo," a common infant game in western, English-speaking cultures. An equivalent experience should be substituted (a common infant social communication game, ideally with turns) rather than attempting to translate the term "peek-a-boo" to the target language. Maintaining experiential equivalence may require significant changes in items describing activities or experiences (Guillemin, 1993; Ware et al., 1996). Finally, content equivalence signifies that the concept itself has the same relevance and acceptance within the target culture and that items within the instrument appropriately represent the concept without significant omissions (Gjersing et al., 2010; Squires et al., 2013; Stewart & Napoles-Springer, 2000).

For *technical*, or operational, *equivalence*, the methods of data collection must yield the same data in the target population as with the source instrument (Squires et al., 2013). This often refers to the format of administration, such as pencil/paper, telephone, or face-to-face interview (Gjersing et al., 2010; Sartorius & Kuyken, 1994). Cultural norms dictate what type of information is shared with whom, where, and how (Carter et al., 2005; Flaherty et al., 1988). The mode of administration used in the source instrument may be inadequate to collect the intended data in the target culture and may need to be adapted (Flaherty et al., 1988; Guillemin, 1993). For example, if one cultural group deems it inappropriate for strangers to ask questions about a child's development, administration strategies may use an interview format, incorporating time to familiarize the caregiver with the interviewer. Technical equivalence also refers to the most appropriate format (e.g., grid vs list), and level of language (i.e., formality, complexity) for the target population (Stewart & Napoles-Springer, 2000; Ware et al., 1996). For parent report tools that assist in screening or diagnosing children with ASD or other developmental disabilities, it is imperative for the tool to have wide applicability across the population. Consequently, the variability in caregiver education and literacy within the general population will influence the types of adaptations required to achieve technical equivalence.

3. Translation and cultural adaptation guidelines

In the growing literature on translation procedures designed to yield valid tools for another language and culture, most authors describe three basic steps in the adaptation process: 1) reproduction of the source instrument into the new target language, 2) pre-testing with members of the target population, and 3) psychometric analysis of the new version. A long and resource-heavy reproduction and pre-testing process is typically recommended, involving many quality assurance measures (Beaton et al., 2000; Sousa & Rojjanasrirat, 2011; Ware et al., 1996); however, this body of literature is largely based on theory rather than empirical data. One older empirical study does suggest that methods using a single forward translation with no further quality checking or pre-testing

procedures is not sufficient (Berkanovic, 1980). However, few studies have compared translation and adaptation methods. From comparison studies that do exist, da Mota Falcao, Ciconelli, and Ferraz (2003) and Perneger, Leplège, and Etter (1999) found few statistical differences between two translated versions of instruments using more or less rigorous methods, and concluded that translations of self-report instruments may be valid without undergoing as many quality assurance measures as once hypothesized. More recently, Acquadro et al. (2008) conducted a literature review of translation and adaptation studies and found some evidence to suggest that more rigorous approaches utilizing multi-step procedures result in higher quality translations. However, they found no empirical evidence to suggest that any specific method is ideal. While these studies provide some insight into the effectiveness of some strategies, more such research is needed to identify the most optimal reproduction and pre-testing methods (Lenderking, 2005). Regardless of how rigorous or complex the reproduction and pre-testing methods are, psychometric properties of an instrument cannot be assumed to be maintained following translation and adaptation (Beaton et al., 2000; McKenna & Doward, 2005). Multiple studies have documented differences in psychometric properties between different language versions of the same instrument, both in traditional and more rigorous translations (Fourie & Feinauer, 2005; Granas, Norgaard, & Sporrong, 2014; Kondolot et al., 2016). Consequently, psychometric properties should be reassessed using the new version with the new target population to ensure that the tool is reliable and valid. Our proposed quality appraisal tool, found in Table A1, presents a minimum set of criteria thought to produce a developmental screening or assessment tool that is sufficiently equivalent to the source instrument for both clinical and research purposes.

3.1. Step 1: reproduction of the instrument

3.1.1. Process of reproduction

The first phase of this process involves reproducing the instrument in the target language. Pre-planning is commonly recommended as a quality assurance method that occurs prior to initiating forward translation. Pre-planning involves examining the source version to determine its translatability in the new cultural context in an effort to identify potential challenges to maintaining construct equivalence, often referred to as a translatability assessment (Acquadro et al., 2018). Prior to initiating translation, teams may annotate the source instrument by offering wording alternatives, clarifying item intent, defining key terms, or replacing culture-specific terminology or concepts with universally understood concepts (Acquadro et al., 1996, 2018; Behr & Scholz, 2011; Chapman & Carter, 1979; Wild et al., 2005). Teams may also perform literature reviews, discuss concepts with experts, or conduct focus groups with parents or other caregivers in the target population (Flaherty et al., 1988; Gjersing et al., 2010; Sartorius & Kuyken, 1994; Stewart & Napoles-Springer, 2000). Such efforts allow researchers to determine how the concepts are defined and understood in the target population and may prevent threats to all dimensions of equivalence.

When the instrument is ready for reproduction, almost all authors recommend using multiple translators trained in dimensions of equivalence. Using a single translator increases the likelihood of including idiosyncrasies or individual language style in the translation (Guillemin, 1993; Wild et al., 2005). Alternatively, creating and comparing multiple translations by multiple translators helps to identify discrepancies in translator interpretation, especially when there are wording ambiguities within the source instrument (Beaton et al., 2000; Sousa & Rojjanasrirat, 2011). Thus, many authors recommend multiple forward translations to be made either independently (Beaton et al., 2000; Sousa & Rojjanasrirat, 2011; Wild et al., 2005) or with a team of translators (Sartorius & Kuyken, 1994; Squires et al., 2013). It should be noted that in the context of the team approach, shared misconceptions may be exacerbated and social norms may make members reluctant to contradict each other (Maneesriwongul & Dixon, 2004). Translators should be native to the target language and culture (Beaton et al., 2000; Guillemin, 1993; Sousa & Rojjanasrirat, 2011), and ideally reside in the target country to ensure the use of local wording (Acquadro et al., 1996; Wild et al., 2005). Translators should have adequate translational competence, including sufficient knowledge in both languages and cultures and training in dimensions of equivalence (Behr, 2018). Multidisciplinary teams can include both individuals who are knowledgeable in the terminology and constructs of the instrument and laypersons who are blind to the concepts of interest, yet knowledgeable in colloquial terminology, such as child care providers or others who regularly interact with parents in the target culture (Acquadro et al., 1996; Beaton et al., 2000; Guillemin, 1993; Sousa & Rojjanasrirat, 2011). This ensures that concepts are appropriately represented in the translated instrument, but that the target population can understand items. Translators should translate items, instructions, and response options. If translations are made independently, forward translations are then collaboratively synthesized into a single version, resolving any discrepancies (Beaton et al., 2000; Sousa & Rojjanasrirat, 2011).

3.1.2. Checking the quality of the reproduction

Once an instrument has been reproduced in the new language, the quality of the translation should be inspected using quality assurance methods. The purpose of quality assurance methods is to identify and resolve discrepancies in linguistic, construct, or technical equivalence. A variety of methods are discussed in the literature, including but not limited to back-translation, preliminary pre-testing, bilingual equivalence assessment, and expert review. Pre-planning, discussed above, is also considered a quality assurance method. A final consensus meeting with the translation team reviews data collected in these quality-checking phases and makes revisions to improve equivalence across versions.

The vast majority of guidelines within the literature call for back-translation methods to test the quality of the reproduction. Back-translation consists of one or multiple translators, blind to the original instrument, who translate the forward reproduction back into the source language, resulting in one or more back-translations (Beaton et al., 2000; Sousa & Rojjanasrirat, 2011). Some recommend using only informed translators with knowledge of the concepts of interest in order to create a translation that would closely reflect the source instrument (Bracken & Barona, 1991; Sartorius & Kuyken, 1994). Others recommend using only individuals who are blind

to the concept of interest in order to better elicit unexpected meanings and highlight imperfections (Beaton et al., 2000; Guillemin, 1993). This latter suggestion may be especially useful for developmental screening tools and assessments, as caregivers may not have any knowledge of the concept of interest when they answer the questions. The source and back-translated versions are then compared for the various dimensions of equivalence by the entire translation and adaptation team. This methodology, called the forward-back (FB) translation method, is meant to highlight translation errors that may have occurred in the forward translation that would impact the validity of the measure (Chapman & Carter, 1979; Guillemin, 1993). Thus, any errors can be identified and resolved during a final consensus step.

Although the FB methodology is recommended and used by most translation teams across fields, several authors point out that it has no scientific basis (Hagell, Hedin, Meads, Nyberg, & McKenna, 2010; McKenna & Doward, 2005; Swaine-Verdier, Doward, Hagell, Thorsen, & McKenna, 2004). Unfortunately, back-translation can provide misleading information about the forward translation (Acquadro et al., 1996; McKenna & Doward, 2005; Swaine-Verdier et al., 2004). Often times, a high-quality forward translation may significantly alter the wording and structure such that it appears dissimilar to the source version when back-translated. Similarly, a high quality forward translation with a poor quality back-translation would give a false negative impression (Acquadro et al., 1996; McKenna & Doward, 2005; Swaine-Verdier et al., 2004). More damaging is the impression of a false positive. When the forward translation inappropriately retains the source language structure and wording, a back-translator may correct these errors, leaving the impression of a high quality forward translation (Acquadro et al., 1996; Maneesriwongul & Dixon, 2004; Nelson McDermott & Palchanes, 1994; Swaine-Verdier et al., 2004; Wang, Lee, & Fetzer, 2006). In addition, back translation does not easily identify complex or formal language in a forward translation. In this case, an accurate back translation would leave the impression of linguistic equivalence, but the reproduction would lack intelligibility to monolingual caregivers in the target population (Chapman & Carter, 1979; Hagell et al., 2010; Nelson McDermott & Palchanes, 1994; Wang et al., 2006). One study comparing translation methodologies determined that back-translation sometimes identified translation errors that did not actually exist and missed problems that did exist (Behr, 2017). Some authors have recently called for the abandonment of back-translation methods altogether (Behr, 2017; Colina et al., 2017). Instrument developers should use caution if they choose back-translation and may consider adding additional quality checking measures.

Pre-testing is a method used in the second phase (described below), to be completed after reproduction and quality checking. However, some authors recommend using *preliminary pre-testing* methods as a quality check during the reproduction process (Nelson McDermott & Palchanes, 1994; Ware et al., 1995). Preliminary pre-testing can be combined with other quality checking measures, and is sometimes used prior to back-translation so that intelligibility in the target population is checked before back-translation methods check for equivalence (Nelson McDermott & Palchanes, 1994). One comprehensive method, developed as an alternative to the FB process, utilizes preliminary pre-testing in a panel format. This strategy, the Dual Panel approach, has been shown to produce translations that are preferable to monolingual lay audiences (Hagell et al., 2010). In the Dual Panel approach, a panel of five to seven translators makes a consensus translation. Translators are native to the target language, have knowledge of the instrument's underlying model, and have frequent contact with the target population (Hagell et al., 2010; Swaine-Verdier et al., 2004). A second panel of laypersons, monolingual in the target language and similar to the target population, then review the translation with a coordinator to ensure appropriate understanding of items (Hagell et al., 2010; Swaine-Verdier et al., 2004). Lay panels typically find more appropriate and intelligible wording than the bilingual panel, and can highlight unexpected connotations (Swaine-Verdier et al., 2004). Hagell et al. (2010) compared translations using the Dual Panel approach with the FB approach and found similar psychometric properties between the two versions; however, there were more missing data from the FB translation and monolingual laypersons preferred the Dual Panel version.

Other quality assurance strategies include *bilingual equivalence assessment* or *expert review*. Bilingual individuals can judge the equivalence of items between versions on various dimensions, either in a panel (Sartorius & Kuyken, 1994) or individually, using rating scales (Acquadro et al., 1996; Squires et al., 2013; Stewart & Napoles-Springer, 2000; Ware et al., 1995, 1996). Alternatively, fully bilingual groups can complete the assessment in each language, comparing results using a test-retest analysis (Chapman & Carter, 1979). Another possibility is to review the new tool with experts in the field of ASD or developmental disabilities from the target language and culture (Sousa & Rojjanasrirat, 2011).

Importantly, teams should keep detailed documentation, including difficulties with translation, alternative wording choices, feedback from panels, and rationales for any revisions made throughout the process, especially those resulting in significant deviations from the source version (Acquadro et al., 1996; Swaine-Verdier et al., 2004). Documentation will improve transparency of the process, promote the importance of the various quality checking procedures, provide helpful information for later validation studies, and provide pertinent information to any future research teams who attempt translation efforts on the same instrument (Beaton et al., 2000; Maneesriwongul & Dixon, 2004; Swaine-Verdier et al., 2004).

3.1.3. Reaching a final consensus

A final consensus meeting with the entire adaptation team is deemed one of the most critical steps in reproduction (Beaton et al., 2000; Wild et al., 2005). A multidisciplinary team is most often recommended, including a methodologist, experts in the area of interest, all translators, the source instrument developers (Beaton et al., 2000; Guillemin, 1993), and possibly a layperson monolingual in the target language, such as a monolingual caregiver (Sousa & Rojjanasrirat, 2011). The team reviews all parts of the instrument and all data collected during the reproduction and quality testing process, considering linguistic and construct equivalence (Beaton et al., 2000; Guillemin, 1993; Sousa & Rojjanasrirat, 2011). Re-translation of individual items may be necessary if problems arise (Beaton et al., 2000; Bracken & Barona, 1991; Guillemin, 1993; Sousa & Rojjanasrirat, 2011). Using consensus rather than compromise, the team makes revisions where necessary and develops a consolidated pre-final version that will be used for pre-

testing (Beaton et al., 2000; Guillemin, 1993; Ware et al., 1996; Wild et al., 2005).

3.2. Step 2: pre-testing

Pre-testing is a vital step to evaluate the quality and readability of a translated instrument in the target population following reproduction and quality testing and prior to undergoing full psychometric analyses. Pre-testing involves administration of the adapted instrument to a small sample of target participants to ensure that no obvious, easily reparable problems exist with the instrument's wording or administration mode before expending the considerable resources necessary for full validation studies with larger samples. There are a variety of ways to pre-test the translation. Qualitative data is most often collected, though some authors perform preliminary statistical analyses using quantitative data (Bracken & Barona, 1991; Maneesriwongul & Dixon, 2004).

Pre-testing allows for assessment of face-validity (Acquadro et al., 1996), and examination of linguistic and construct equivalence (Guillemin, 1993; Sousa & Rojjanasrirat, 2011), and identification of any unnecessarily complex or formal language (Acquadro et al., 1996). The sample should adequately represent the diversity within the target population (Acquadro et al., 1996; Bracken & Barona, 1991; Ware et al., 1996; Wild et al., 2005). For developmental screening and assessment tools, the sample should consist of caregivers of children in the target age range, with a variety of education levels, and representing a variety of regions in the target geographic area. Adaptation teams may choose to include some caregivers whose children are known to have developmental difficulties. The literature is not consistent with recommendations for sample size, ranging from 5 (Wild et al., 2005) to 50 or more (Ware et al., 1996). Focus groups may be used, though individual cognitive interviews are recommended (Acquadro et al., 1996). During a cognitive interview, participants complete items as intended and are asked probing questions about each item, including their understanding of the item's meaning, how they come up with answers to the items, and the appropriateness of response options (Beaton et al., 2000; Guillemin, 1993; Stewart & Napoles-Springer, 2000; Wild et al., 2005). Many authors suggest documenting participant reactions such as facial cues, resistance to respond, or nervous laughter, because these reactions could indicate inappropriate wording or administration method (Bracken & Barona, 1991). Researchers should also note possible acquiescence response bias and socially desirable responding in parent-report instruments, particularly if a stigma around child developmental disabilities exists within the culture. Less often, feedback surveys are completed following self-administration of the instrument (Reichenheim & Moraes, 2007; Sousa & Rojjanasrirat, 2011), common with large pre-testing samples (Ware et al., 1996).

Bilingual pre-testing samples can provide unique information regarding the equivalence of instruments, though fully bilingual laypersons may be difficult to recruit (Guillemin, 1993; Ware et al., 1996). Methods include equivalence ratings (Guillemin, 1993; Squires et al., 2013) or test-retest or split-half analyses (Acquadro et al., 1996; Sartorius & Kuyken, 1994; Sousa & Rojjanasrirat, 2011; Wang et al., 2006). Items with low correlations on such equivalence tests, or low ranking items on equivalence ratings, are revised.

Data collected during the pre-testing phase should be used to revise problem items, instructions, response options, or administration methods so that linguistic, construct, technical, and in turn, psychometric equivalences are maintained (Soto et al., 2015; Sousa & Rojjanasrirat, 2011). Similar to the reproduction phase, the pre-testing phase should be well documented, including rationales for any changes made (Beaton et al., 2000; Maneesriwongul & Dixon, 2004). Finally, individual items may be sent back into the reproduction phase if significant issues are identified.

3.3. Step 3: psychometric testing

Although ignored in many guidelines in the literature (Squires et al., 2013; Wild et al., 2005), the re-testing of psychometric properties of an instrument may be the most important step in the process of translating and adapting a screening tool. Many consider it a gross misconception to assume that the psychometric properties of an instrument are retained when translated and adapted for a new population (Beaton et al., 2000; McKenna & Doward, 2005). Without psychometric data, the reliability and validity of a given adapted screening measure are unknown.

Parent-report tools are used in developmental screening tools as well as more comprehensive assessments. Developmental screening tools are used clinically to identify children who exhibit concerning behaviors so they may be referred for a full evaluation and developmental support. Comprehensive assessments are used to provide information for evaluation teams in order to determine a differential diagnosis. Consequently, assessment tools should have better diagnostic validity than screening tools. When examining the properties of a new screening or assessment tool, researchers should prioritize analyses that test whether the tool consistently collects appropriate data and differentiates between those who do or do not have a given developmental disorder.

3.3.1. Recruitment

There are multiple recruitment strategies used to test psychometric properties of screening and diagnostic instruments. In case control studies, two separate populations are recruited: those more likely to have the disorder and those less likely to have the disorder (Campbell et al., 2015). Although this strategy will likely be fruitful in identifying participants with the disorder, diagnostic accuracy will be somewhat exaggerated compared to clinical practice, as statistical differences between participant groups are likely to be pronounced (Campbell et al., 2015). Cross-sectional studies, however, recruit only from populations likely to have the disorder, providing a more valid measure of diagnostic accuracy for assessment tools in a clinical setting, where all patients present with developmental concerns (Campbell et al., 2015). Alternatively, participants may be recruited from the general population. This may provide a more accurate representation of a screening tool intended for general population screening. A less robust method involves a known-groups analysis, where participants are recruited who are already known to either have the developmental condition of interest or be typically developing (Ware et al., 1995, 1996). While few strict guidelines are provided, sample sizes should be large

enough for the analyses of interest, which will depend on the number of items in the tool, the number of response options, and the variability of responses (Beaton et al., 2000; Sousa & Rojjanasrirat, 2011). In general population samples, the required sample size will also depend on the prevalence of the condition. For ASD, the estimated prevalence is 1–2% of the population (Baio, Wiggins, & Christensen, 2018), whereas communication disorders affect approximately 10% of children (NCBI, 2016). Consequently, studies of ASD-specific tools would require larger samples from the general population, as compared to communication disorder-specific tools, in order to identify enough cases for psychometric analysis.

3.3.2. Methodology

Methodology for the assessment of psychometric properties is dependent on the intended statistical analyses. For most tests, a large sample of participants will complete the adapted tool and analyses will follow. Participants may complete an additional assessment or provide additional information in validity studies. Tests of diagnostic accuracy involve the administration of a gold standard index test to be used as a definitive factor that categorizes participants as having the condition in question or not having it (Campbell et al., 2015). Diagnostic accuracy of the instrument is determined by comparing those correctly and incorrectly identified by the instrument (Campbell et al., 2015). To reduce bias, researchers can alternate the order of administration, use appropriate time delay between examinations, blind examiners, recruit a diverse, yet representative population compared to the target population (Campbell et al., 2015; Kim, Lee, Choi, Huh, & Park, 2015), and use an index test that is valid and reliable in the target culture and language (Habbema, Eijkemans, Krijnen, & Knottnerus, 2009).

3.3.3. Statistical analyses

To examine a tool's performance with the new population, analyses begin with examination of data quality, including data completeness, variability of scores, and response bias. Large portions of missing data on a particular item may indicate a confusing, offensive, or inapplicable item, whereas large portions of missing data on many items in a section may indicate problems with instructions, layout, or administration methods (Ware & Gandek, 1998; Ware et al., 1995, 1996). When comparing findings to results of psychometric testing from the original instrument, participants should select from the same range of response options on each item, and their scale scores should span the same range of scores possible (Stewart & Napoles-Springer, 2000). Similarly, participants should have similar proportions of ceiling or floor level scores as in the source instrument (Ware & Gandek, 1998). In the case of screening tools, there should be large numbers of individuals who score at low risk when the tool is used in a low-risk sample. Measures of acquiescence response bias, social desirable responding, and preferences for extreme response categories (Stewart & Napoles-Springer, 2000) may be particularly relevant for constructs with high stigma, which may be the case with scales related to child developmental difficulties. Finally, scale and item means and standard deviations should be similar, or at least follow a similar pattern of rank order, in both the target and source versions of the instrument (Bracken & Barona, 1991; Ware & Gandek, 1998; Ware et al., 1996). For example, many parent-report tools include items that ask about milestones acquired in typical development as well as behaviors not present in typical development. In both language versions, means from a low-risk sample should indicate frequent endorsement of items describing behaviors present in typical development, especially as children's ages increase. Similarly, means from a low-risk sample should indicate less frequent endorsement of items describing behaviors not present in typical development.

The most frequent type of *reliability analysis* recommended in the literature is internal consistency (Bracken & Barona, 1991; Gjersing et al., 2010). Also commonly recommended are measures of test-retest reliability (Stewart & Napoles-Springer, 2000; Ware & Gandek, 1998; Ware et al., 1996). Inter-rater reliability is less frequently described (Acquadro et al., 1996), yet it may provide especially useful information if the screening tool is intended for use with varying informants, such as parents and childcare providers. Reliability analyses provide information about the tool's consistency.

Arguably, the most important property of a screening or assessment tool is its *diagnostic accuracy*. For any tool, this begins with the scoring process, which requires reexamination when translated and adapted for use with a new population (Acquadro et al., 1996; Stewart & Napoles-Springer, 2000; Ware et al., 1996). Summated rating scale assumptions (Stewart & Napoles-Springer, 2000; Ware & Gandek, 1998; Ware et al., 1995, 1996) or weighting of items should be reexamined (Guillemin, 1993; Stewart & Napoles-Springer, 2000). Similarly, cut-off scores cannot be assumed to be maintained with a new population (Habbema et al., 2009; Stewart & Napoles-Springer, 2000). Cut-off scores, which determine which individuals are considered at high risk according to the instrument, must be chosen strategically, in order to maximize sensitivity and specificity and to minimize errors (i.e. false positives and false negatives; Kim et al., 2015). This is often performed using an area under the curve analysis of the receiver-operating characteristic curve (Campbell et al., 2015; Habbema et al., 2009).

Given appropriate scoring methods and cut-off points, the tool's *sensitivity, specificity, positive predictive values,* and *negative predictive values* should be measured (Campbell et al., 2015; Habbema et al., 2009; Kim et al., 2015; Stewart & Napoles-Springer, 2000). It is helpful for authors to provide frequencies of true and false positives and negatives so that the reader can calculate these and other variables related to diagnostic accuracy (Kim et al., 2015). Tests of convergent and discriminant validity can also provide information about what an instrument is measuring (Beaton et al., 2000; Hagell et al., 2010; Sousa & Rojjanasrirat, 2011). This includes examining correlations with external variables or other measures that are thought be associated or not with the given construct (Anderson, Mcfarlane, Naughton, & Shumaker, 1996; Stewart & Napoles-Springer, 2000; Ware et al., 1996).

Equivalence testing determines if the two versions of the tool, the source and new versions, are measuring the same information. Evidence of equivalence between versions is less important than diagnostic accuracy if the tool is planned for clinical use rather than use in cross-cultural research, where data collected by the different versions of the tool will be pooled or compared. However, if the versions are not found to be equivalent in their measurement, it may be that each is measuring a separate latent construct, which may or may not be equally able to predict the presence of a developmental disability.

Equivalence analyses should show that the translated instrument is equivalent to the original instrument in all observable respects, where the distribution of item and scale scores depends on actual values of latent variables and not on any other characteristic, a concept referred to as measurement invariance (Küçükdeveci, Sahin, Ataman, Griffiths, & Tennant, 2004; Ware & Gandek, 1998). Classical analysis of measurement invariance involves a series of multiple group factor analyses with increasing constraints placed on the parameters until all parameters are fixed. If the final model continues to fit the data from both versions of the instrument, then the two instruments show measurement invariance. An alternative approach, based in item response theory and focused at the item level, is known as differential item functioning (DIF; Anderson et al., 1996; Küçükdeveci et al., 2004). If high DIF is found for a particular item, participants appear to be responding to that item differently across versions. The item should then be reviewed and possibly revised (Anderson et al., 1996; Küçükdeveci et al., 2004). Of note, analyses of measurement invariance and DIF require a considerable variability within the data and a large covariance matrix. Depending on the length and factor structure of the tool, an extremely large sample size may be needed, which may make these analyses logistically difficult.

At any level of equivalence analysis, if there are differences found between scales, the research team must decide if these differences are likely to be true population differences or differences in the instrument itself (Reichenheim & Moraes, 2007; Ware & Gandek, 1998; Ware et al., 1996). In the case of the latter, the adapted instrument may need to be revised (by returning items to the reproduction and pre-testing phases), depending on its intended purposes. As previously indicated, lack of measurement invariance is less important if diagnostic accuracy is maintained for an instrument designed solely for clinical use. As in each of the other steps, transparent documentation of psychometric analyses should be sufficient to allow the reader to examine the reliability and validity of the new version (Acquadro et al., 1996).

Once acceptable psychometric properties have been achieved, the new instrument can be used in the target population. In an effort to maximize transparency, the entire translation and adaptation process should be described with as much detailed information as possible in peer-reviewed sources (Acquadro et al., 1996; Maneesriwongul & Dixon, 2004). However, due to the complexity and length of the process, such information may be published in multiple articles (Sousa & Rojjanasrirat, 2011).

4. A quality appraisal tool

Using these guidelines presented in the literature, we propose a quality appraisal tool (Table A1). This checklist has been designed to systematically analyze the instrument translation and cultural adaptation process of any one assessment. It could be used in a clinical setting by those who seek to compare the quality of the translation of a particular instrument or set of instruments. Empirically, it could be used by researchers to systematically analyze the translation methods of a set of instruments for a systematic evidence map. Clinicians and researchers alike who wish to translate and culturally adapt an instrument themselves could use the quality appraisal tool as a way to guide their own process. The appraisal tool does not have a quantitative scoring system, nor is there consensus on what level of quality would indicate acceptable level of rigor in methodology. However, it provides a set of minimum steps thought necessary to ensure appropriate translation of self-report developmental screening and assessment tools.

5. Final considerations

Parent-report developmental screening and assessment tools are widely used to identify children showing signs of ASD or other developmental disabilities. Because of their utility, these tools are being translated into multiple languages and used internationally. Reliable and valid measures are a critical need for improving our global capacity to serve individuals with ASD and other developmental disabilities. Yet traditional forward-back translation methodology is not sufficient to translate and culturally adapt these measures such that their reliability and validity is retained in the new target population. In this article, we have described the various dimensions of equivalence most relevant when translating and culturally adapting self-report developmental screening and assessment tools, and provided a description of the basic stages in the process, along with various strategies that can be used in each stage. Due to the scarcity of studies comparing translation methodologies, it is not possible to recommend any one specific set of strategies based on empirical evidence. Clearly, more research is needed in this area to determine which strategies are the most efficient and effective in translating and culturally adapting tools so that the new tool functions appropriately in the new target population. For now, we hope that shedding light on the translation and cultural adaptation process will encourage research teams within the fields of ASD and developmental disabilities to increase the rigor in their translation methods and describe their methodology in greater detail in research reports.

Conflict of interest

Both authors declare that they have no conflicts of interest. This paper was prepared as a portion of a doctoral dissertation, as provided by U.S. Department of Education Doctoral Leadership Grant, H325D130041.

Appendix A

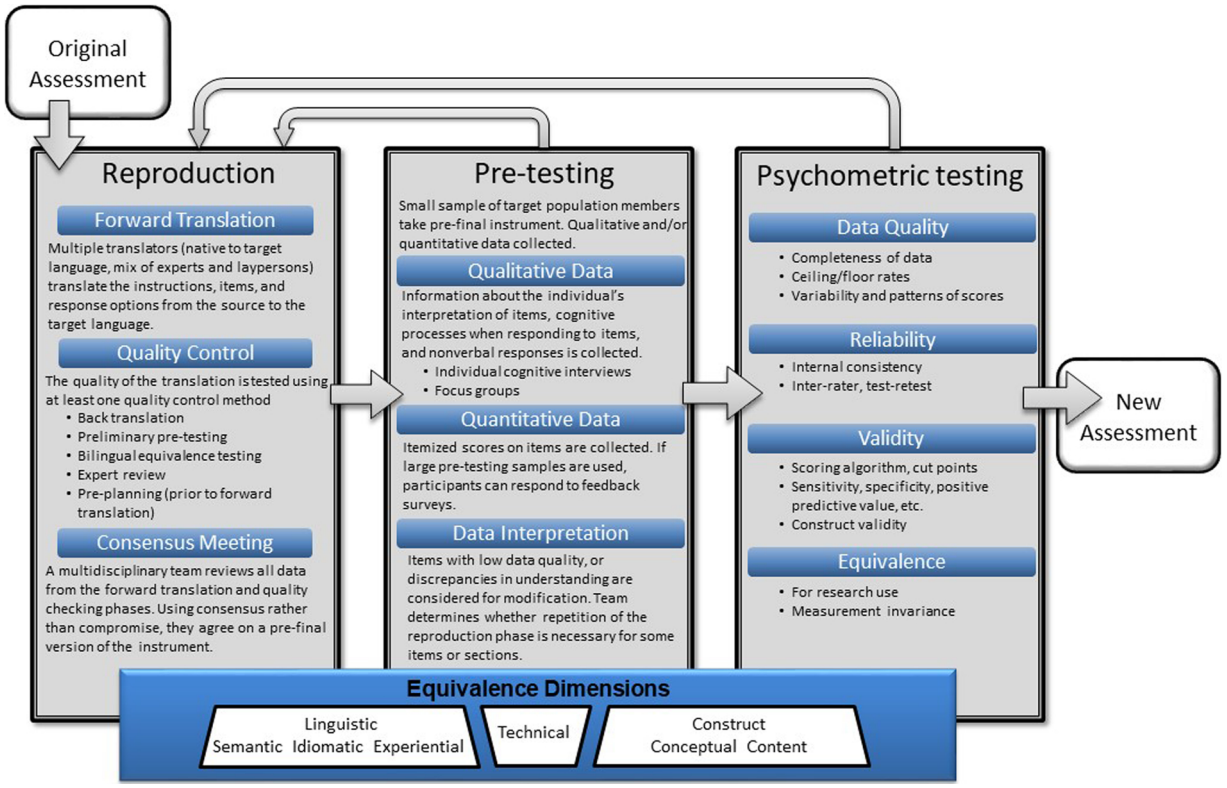

Fig. A1. Representation of the translation and cultural adaptation process.

Table A1
Quality Appraisal Tool for Translated and Adapted Caregiver-Report Developmental Instruments.

1. Was more than one translator used? Yes / No
- Did translators have varying backgrounds? Yes / No / NA
Description:

2. Was at least one sound quality assurance method used in the reproduction phase? (i.e. pre- planning, cognitive testing, equivalence testing, expert panel, back-translation with consensus committee review, dual panel) Yes / No
Description:

3. Was pretesting completed with an appropriate sample of members of the target population? Yes / No
Description:

4. Were appropriate revisions made using data from quality assurance and pre-testing steps?Yes / No
Description:

5. Was psychometric testing completed on the new version? Yes / No
- Was the sample sufficiently large and diverse to perform statistical analyses and adequately represent the target population? Yes / No / NA
Description:

- Was appropriate blinding and time delay used when applicable? Yes / No / NA
Description:

- Was the gold standard reference test culturally and linguistically appropriate (validated on the target population)? Yes / No / NA
Description:

- Were scoring procedures and cut-off scores reexamined and found to be appropriate? Yes / No / NA
Description:

- Were diagnostic accuracy data reported and acceptable (true/false negatives, true/false positives)? Yes / No / NA
Description:

- Was data quality acceptable? (e.g. ranges of scores, ceiling/floor rates, etc.) Yes / No / NA
Description:

- Were appropriate reliability and validity tests completed with results reported? Yes / No / NA
Description:

- Were appropriate item and scale equivalence tests completed with results reported? Yes / No / NA
Description:

References

Acquadro, C., Conway, K., Hareendran, A., & Aaronson, N. (2008). Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. *Value in Health : the Journal of the International Society for Pharmacoeconomics and Outcomes Research, 11*, 509–521.

Acquadro, C., Jambon, B., Ellis, D., & Marquis, P. (1996). Language and translation issues. In B. Spilker (Ed.). *Quality of life and pharmacoeconomics in clinical trials* (pp. 575–585). (2nd ed.). Phila: Lippincott-Raven Publishers.

Acquadro, C., Patrick, D., Eremenco, S., Martin, M., Kuliś, D., Correia, H., ... International Society for Quality of Life Research (ISOQOL) Translation and Cultural Adaptation Special Interest Group (TCA-SIG) (2018). Emerging good practices for Translatability Assessment (TA) of Patient-Reported Outcome (PRO) measures. *Journal of Patient-Reported Outcomes, 2*, 8.

Al Maskari, T., Melville, C., & Willis, D. (2018). Systematic review: Cultural adaptation and feasibility of screening for autism in non-English speaking countries. *International Journal of Mental Health Systems, 12*, 1–19.

Anderson, R., Mcfarlane, M., Naughton, M., & Shumaker, S. (1996). Conceptual issues and considerations in cross-cultural validation of generic health-related quality of life instruments. In B. Spiker (Ed.). *Quality of life and pharmacoeconomics in clinical trials* (pp. 605–612). (2nd ed.). Philadelphia: Lippincott-Raven Publishers.

Baio, J., Wiggins, L., Christensen, D., et al. (2018). Prevalence of autism Spectrum disorder among children aged 8 years — Autism and developmental disabilities monitoring network, 11 sites, United States, 2014. *Morbidity and Mortality Weekly Report Surveillance Summaries (Washington, DC : 2002), 67*(SS-6), 1–23.

Beaton, D., Bombardier, C., Guillemin, F., & Ferraz, M. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine, 25*, 3186–3191.

Behr, D. (2017). Assessing the use of back translation: The shortcomings of back translation as a quality testing method. *International Journal of Social Research Methodology, 20*, 573–584.

Behr, D. (2018). Translating questionnaires for cross-national surveys: A description of a genre and its particularities based on the ISO 17100 categorization of translator competences. *Translation and Interpreting, 10*, 5–20.

Behr, D., & Scholz, E. (2011). Questionnaire translation in cross-national survey research: On the types and value of annotations. *Methoden, Daten, Analysen, 5*.

Berkanovic, E. (1980). The effect of inadequate language translation on Hispanics' responses to health surveys. *American Journal of Public Health, 70*, 1273–1276.

Bracken, B., & Barona, A. (1991). State of the art procedures for translating, validating and using psychoeducational tests in cross-cultural assessment. *School Psychology International, 12*, 119–132.

Brislin, R. (1970). Back-translation for cross-cultural research. *Journal of Cross-cultural Psychology, 1*(3), 185–216.

Campbell, J., Klugar, M., Ding, S., Carmody, D., Hakonsen, S., Jadotte, Y., ... Munn, Z. (2015). Diagnostic test accuracy: Methods for systematic review and meta-analysis. *International Journal of Evidence-Based Healthcare, 13*(3), 154–162.

Carter, J., Lees, J., Murira, G., Gona, J., Neville, B., & Newton, C. (2005). Issues in the development of cross-cultural assessments of speech and language for children. *International Journal of Language & Communication Disorders, 40*, 385–401.

Castro, F., Barrera, M., & Holleran Steiker, L. (2010). Issues and challenges in the design of culturally adapted evidence-based interventions. *Annual Review of Clinical Psychology, 6*, 213–239.

Chapman, D., & Carter, J. (1979). Translation procedures for the cross cultural use of measurement instruments. *Educational Evaluation and Policy Analysis, 1*(3), 71–76.

Colina, S., Marrone, N., Ingram, M., & Sanchez, D. (2017). Translation quality assessment in health research: A functionalist alternative to back-translation. *Evaluation & the Health Professions, 40*, 267–293.

da Mota Falcao, D., Ciconelli, R., & Ferraz, M. (2003). Translation and cultural adaptation of quality of life questionnaires: An evaluation of methodology. *The Journal of Rheumatology, 30*(2), 379–385.

El-Behadli, A., Neger, E., Perrin, E., & Sheldrick, R. (2015). Translations of developmental screening instruments: An evidence map of available research. *Journal of Developmental and Behavioral Pediatrics : JDBP, 36*, 471–483.

Flaherty, J., Gaviria, M., Pathak, D., Mitchell, T., Wintrob, R., Richman, J., ... Birz, S. (1988). Developing instruments for cross-cultural psychiatric research. *The Journal of Nervous and Mental Disease, 176*(5), 257–263.

Fourie, J., & Feinauer, I. (2005). The quality of translated medical research questionnaires. *Southern African Lingiustics and Applied Language Studies, 23*, 349–367.

Gjersing, L., Caplehorn, J., & Clausen, T. (2010). Cross-cultural adaptation of research instruments: Language, setting, time and statistical considerations. *BMC Medical Research Methodology, 10*(13), 1–10.

Granas, A., Norgaard, L., & Sporrong, S. (2014). Lost in translation? Comparing three Scandinavian translations of the Beliefs about Medicines Questionnaire. *Patient Education and Counseling, 96*, 216–221.

Guillemin, F. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology, 46*, 1417–1432.

Habbema, J., Eijkemans, R., Krijnen, P., & Knottnerus, J. (2009). Analysis of data on the accuracy of diagnostic tests. In J. Knottnerus, & F. Buntinx (Eds.). *The evidence base of clinical diagnosis: Theory and methods of diagnostic research* (pp. 118–145). (2nd ed.). Wiley-Blackwell Publishing Ltd.

Hagell, P., Hedin, P., Meads, D., Nyberg, L., & McKenna, S. (2010). Effects of method of translation of patient-reported health outcome questionnaires: A randomized study of the translation of the rheumatoid arthritis quality of life (RAQoL) instrument for Sweden. *Value in Health : the Journal of the International Society for Pharmacoeconomics and Outcomes Research, 13*, 425–430.

Harrison, A., Long, K., Tommet, D., & Jones, R. (2017). Examining the role of race, ethnicity, and gender on social and behavioral ratings within the Autism Diagnostic Observation Schedule. *Journal of Autism and Developmental Disorders, 47*, 2770–2782.

Kim, K., Lee, J., Choi, S., Huh, J., & Park, S. (2015). Systematic review and meta-analysis of studies evaluating diagnostic test accuracy : A practical review for clinical researchers – Part I. General guidence and tips. *Korean Journal of Radiology, 16*(6), 1188–1196.

Kondolot, M., Ozmert, E., Oztop, D., Mazicioglu, M., Gumus, H., & Elmali, F. (2016). The modificed checklist for autism in Turkish toddlers: A different cultural adaptation sample. *Research in Autism Spectrum Disorders, 21*, 121–127.

Küçükdeveci, A., Sahin, H., Ataman, S., Griffiths, B., & Tennant, A. (2004). Issues in cross-cultural validity: Example from the adaptation, reliability, and validity testing of a Turkish version of the Stanford Health Assessment Questionnaire. *Arthritis Care & Research, 51*(1), 14–19.

Lenderking, W. (2005). Comments on the ISPOR Task Force report on translation and adaptation of outcomes measures: Guidelines and the need for more research. *Value in Health : the Journal of the International Society for Pharmacoeconomics and Outcomes Research, 8*, 92–93.

Maneesriwongul, W., & Dixon, J. (2004). Instrument translation process: A methods review. *Journal of Advanced Nursing, 48*(2), 175–186.

McKenna, S., & Doward, L. (2005). The translation and cultural adaptation of patient-reported outcome measures. *Value In Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Research, 8*(2), 89–91.

Committee on the Evaluation of the Supplemental Security Income (SSI) Disability Program for Children with Speech Disorders and Language Disorders; Board on the Health of Select Populations; Board on Children, Youth, and Families; Institute of Medicine; Division of Behavioral and Social Sciences and Education; National Academies of Sciences, Engineering, and Medicine; Rosenbaum S, Simon P, editors. Speech and Language Disorders in Children Implications for the Social Security Administration's Supplemental Security Income Program. Washington (DC): National Academies Press (US); 2016 Apr 6. 2, Childhood Speech and Language Disorders in the General U.S. Population. Available from: https://www.ncbi.nlm.nih.gov/books/NBK356270/.

Nelson McDermott, M., & Palchanes, K. (1994). A literature review of critical elements in translation theory. *Image the Journal of Nursing Scholarship, 26*, 113–117.

Perneger, T., Leplège, A., & Etter, J. (1999). Cross-cultural adaptation of a phychometric instrument: Two methods compared. *Journal of Clinical Epidemiology, 52*, 1037–1046.

Reichenheim, M., & Moraes, C. (2007). Operationalizing the cross-cultural adaptation of epidemological measurement instruments. *Revista de SaÃºde PÃºblica, 41*, 665–673.

Robins, D., Casagrande, K., Barton, M., Chen, C., Dumont-Mathieu, T., & Fein, D. (2014). Validation of the modified checklist for Autism inToddlers, revised with follow-up (M-CHAT-R/F). *Pediatrics, 133*(1), 37–45.

Rutter, M., Bailey, A., & Lord, C. (2003). *The social communication questionnaire.* Los Angeles: Western Psychological Services.

Sartorius, N., & Kuyken, W. (1994). Translation of health status instruments. In J. Orley, & W. Kuyken (Eds.). *Quality of life assessment: International perspectives* (pp. 3–

18). Berlin: Springer-Verlag.
Soto, S., Linas, K., Jacobstein, D., Biel, M., Migdal, T., & Anthony, B. J. (2015). A review of cultural adaptations of screening tools for autism spectrum disorders. *Autism : the International Journal of Research and Practice, 19*(6), 646–661.
Sousa, V., & Rojjanasrirat, W. (2011). Translation, adaptation and validation of instruments or scales for use in cross-cultural health care research: A clear and user-friendly guideline. *Journal of Evaluation in Clinical Practice, 17*(2), 268–274.
Squires, A., Aiken, L., van den Heede, K., Sermeus, W., Bruyneel, L., Lindqvist, R., ... Zikos, D. (2013). A systematic survey instrument translation process for multi-country, comparative health workfor studies. *International Journal of Nursing Studies, 50*(2), 264–273.
Stewart, A., & Napoles-Springer (2000). Health-related quality-of-life assessments in diverse population groups in the United States. *Medical Care, 38*(9), II102–II124.
Swaine-Verdier, A., Doward, L., Hagell, P., Thorsen, H., & McKenna, S. (2004). Adapting quality of life instruments. *Value in Health: The Journal Of The International Society For Pharmacoeconomics And Outcomes Researchealth, 7*(Suppl. 1), S27–30.
van Kleeck, A. (1994). Potential cultural bias in training parents as conversational partners with their children who have delays in language development. *American Journal of Speech-language Pathology, 3*(1), 67–78.
Wang, W., Lee, H., & Fetzer, S. (2006). Challenges and strategies of instrument translation. *Western Journal of Nursing Research, 28*(3), 310–321.
Ware, J., & Gandek, B. (1998). Methods for testing data quality scaling assumptions, and reliability: The IQOLA Project approach. *Journal of Clinical Epidemiology, 51*, 945–952.
Ware, J., Keller, S., Gandek, B., Brazier, J., Sillivan, M., & The IQOLA Project Group (1995). Evaluating translations of health status questionnaires: Methods from the IQOLA project. *International Journal of Technology Assessment in Health Care, 11*, 525–551.
Ware, J., Gandek, B., & Keller, S. (1996). Evaluating instruments used cross-nationally: Methods from the IQOLA proyect. In B. Spilker (Ed.). *Quality of life and pharmacoeconomics in clinical trials* (pp. 681–692). (2nd ed.). Philadelphia: Lippincott-Raven Publishers.
Wetherby, A., & Prizant, B. (2002). *The infant toddler checklist from the communication and symbolic behavior scales.* Baltimore: Brookes Publishing.
Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., ... Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health : the Journal of the International Society for Pharmacoeconomics and Outcomes Research, 8*(2), 94–104.
Windham, G., Smith, K., Rosen, N., Anderson, M., Grether, J., Coolman, R., ... Harris, S. (2014). Autism and developmental screening in a public, primary care setting primarily serving Hispanics: Challenges and results. *Journal of Autism and Developmental Disorders, 44*, 1621–1632.
World Health Organization (2014). *WHA67.8: Comprehensive and coordinated efforts for the management of autism spectrum disorders. Sixty-Seventh World Health Assembly, Geneva, Switzerland.* Retrieved fromhttp://www.who.int/mental_health/maternal-child/WHA67.8_resolution_autism.pdf.