Universidade de Aveiro
2018

Departamento de Educação e Psicologia

ANA RITA DOS SANTOS VALENTE

DESENVOLVIMENTO DE INSTRUMENTOS DE AVALIAÇÃO PARA ADULTOS COM PERTURBAÇÃO DE FLUÊNCIA

DEVELOPMENT OF ASSESSMENT TOOLS TO EVALUATE ADULTS WITH FLUENCY DISORDERS

Universidade de Aveiro
2018

Departamento de Educação e Psicologia

ANA RITA DOS SANTOS VALENTE

DESENVOLVIMENTO DE INSTRUMENTOS DE AVALIAÇÃO PARA ADULTOS COM PERTURBAÇÃO DE FLUÊNCIA

DEVELOPMENT OF ASSESSMENT TOOLS TO EVALUATE ADULTS WITH FLUENCY DISORDERS

Tese apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Doutor em Psicologia, realizada sob a orientação científica do Professor Doutor Luís Miguel Teixeira de Jesus, Professor Coordenador da Universidade de Aveiro e co-orientação científica da Professora Doutora Margaret M. Leahy, Associate Professor do Trinity College, Dublin, Irlanda.

Apoio financeiro da Fundação para a Ciência e Tecnologia do Ministério da Ciência, Tecnologia e Ensino Superior, e do Fundo Social Europeu, no âmbito do POPH-QREN – Formação Avançada – Tipologia 4.1. (Referência SFRH/BD/78311/2011)

Para o(s) meu(s) filho(s)
Para os meus pais.
Aos que me deram a vida.
Àqueles a quem dei a minha.

o júri

presidente

Doutor Fernando Manuel dos Santos Ramos
Professor Catedrático do Departamento de Comunicação e Arte
Universidade de Aveiro

vogais

Doutor Alexandre Lemos de Castro Caldas
Professor Catedrático do Instituto de Ciências da Saúde
Universidade Católica Portuguesa

Doutor Carlos Fernandes da Silva
Professor Catedrático do Departamento de Educação e Psicologia
Universidade de Aveiro

Doutor Luís Miguel Teixeira de Jesus
Professor Coordenador da Escola Superior de Saúde
Universidade de Aveiro (Orientador)

Doutora Ana Cardoso Allen Gomes
Professora Associada de Faculdade de Psicologia e Ciências da Educação
Universidade de Coimbra

Doutora Maria Selene Henriques da Graça Vicente
Professora Auxiliar da Faculdade de Psicologia e Ciências da Educação
Universidade do Porto

agradecimentos

Ao Professor Doutor Luís Jesus pelo acompanhamento e orientação científica ao longo destes quatro anos. Gostaria de agradecer e reconhecer a paciência e compreensão nas alturas duras e difíceis. Foi um gosto trabalhar assim!

I would like to thank my supervisor, Dr. Margaret Leahy for her invaluable support and supervision. Thanks for the availability to discuss my doubts and to give me your feedback. I learned a lot with you!

Ao IEETA, pelo espaço disponibilizado para o desenvolvimento deste trabalho. Às pessoas que fazem do IEETA uma casa da qual vou ter saudades.

À ESSUA, pela disponibilidade demonstrada aquando da necessidade de local para recolha de dados.

A todos os adultos que participaram e tornaram possível a validação dos instrumentos de avaliação desenvolvidos.

À Professora Doutora Andreia Hall pela disponibilidade demonstrada para orientação na análise estatística dos dados.

À Professora Doutora Maria Teresa Roberto pela colaboração no processo de tradução e adaptação cultural do *POSHA-S*.

À Professora Doutora Helena Alvelos pela ajuda na determinação da validade de construto do instrumento *ALUSCA*.

À Associação Portuguesa de Gagos pela promoção e divulgação do presente estudo e pela ajuda na recolha da amostra.

A todas as Freguesias que participaram ativamente no processo de recolha de dados do POSHA-S.

À professora Doutora Marisa Lousada pela sua participação no painel de peritos e contributo na validação dos instrumentos desenvolvidos. E pelas palavras amigas, o encorajamento e a confiança.

À Professora Doutora Joana Rocha e à Terapeuta Maria Manuel pela participação no painel de peritos e contributo para a validação dos instrumentos desenvolvidos.

Ao Centro de Tratamento de Gaguez pela disponibilidade demonstrada no processo de validação dos instrumentos desenvolvidos.

Ao Terapeuta Gonçalo Leal, companheiro de caminhada científica, pelas discussões e partilha de ideias sobre a gaguez, este tema que tanto nos fascina. Obrigada pela confiança!

Ao meu querido Tiago Ferraz, amigo de todas as horas, pela colaboração no processo de tradução e adaptação cultural do *POSHA-S*. E, acima de tudo, pela constante presença, preocupação, amizade e carinho.

À minha querida Ana Félix e à minha querida Elisabete Sousa pela ajuda no processo de recolha de dados do *POSHA-S*.

À minha amiga e colega de profissão Ercília Fernandes pela ajuda no processo de recolha de dados do *POSHA-S* e pelas palavras de incentivo ao longo deste processo.

I would like to thank to Ken St. Louis for his support and invaluable contribution in the translation /cross-cultural adaptation of POSHA-S and in the development of our POSHA-S paper.

Aos meus colegas Joaquim Santos e Joana Martinez, grandes companheiros durante metade deste processo. Obrigada Joana pela tua ajuda na recolha de dados para o ALUSCA e pelas tuas palavras de encorajamento. Obrigada Joaquim pelos cafés, pelas chaves que entregaste por mim (!!!) e, claro, pela disponibilidade que sempre demonstras quando precisei de ajuda.

Aos meus pais, a quem dedico esta tese. Esta é uma forma singela de reconhecer o tanto que fizeram/fazem por mim. Obrigada mãe pela tua exigência (que faz parte de mim), pelos valores fortes que me transmites e pelo apoio/amor incondicional que tiveste ao longo deste processo. Obrigada pelo suporte nas horas trágicas. Obrigada pai por te saber sempre presente, sempre ali, sempre pronto a ajudar.

Ao meu irmão Manel, por estar sempre presente. Gosto muito de ti, mano!

Ao David, pelo seu coração enorme. E pelo amor que me tem. Obrigada pela paciência. Obrigada pelo crescente apoio ao longo de todo este processo. Obrigada por Seres comigo.

Ao meu filho João. Porque veio trazer *aquele sentido* que faltava.

À Força que me suporta e ampara. Seja Ela o que/quem for.

Palavras-chave Gaguez, gravidade, pragmática, atitudes, adultos

resumo No presente estudo foram desenvolvidos dois instrumentos (*Severity Assessment Based on Events of Stuttering – SABES*; *Assessment of Language Use in Social Contexts for Adults – ALUSCA*), para avaliar adultos que gaguejam (AQG). Os instrumentos foram desenvolvidos com base na adaptação da Classificação Internacional de Funcionalidade, Incapacidade e Saúde (CIF) ao estudo da gaguez. Foi ainda realizada a tradução e adaptação cultural de um questionário (*Public Opinion Survey of Human Attributes-Stuttering – POSHA-S*) que tem como finalidade a determinação das atitudes da sociedade relativamente à gaguez e às pessoas que gaguejam.

O instrumento *SABES* avalia a gravidade da gaguez através da avaliação da frequência, duração, comportamentos associados, grau de tensão e naturalidade de cada momento de gaguez, em quatro amostras de fala, através do uso de um *software* de anotação. Foi determinada a validade de conteúdo, a consistência interna e a validade de construto do instrumento *SABES*. A validade de conteúdo foi analisada através de um processo de duas etapas. O estudo piloto foi conduzido com 5 AQG para analisar a praticabilidade dos procedimentos. A consistência interna foi analisada através do alfa de *Cronbach*. Os procedimentos do SABES foram aplicados a 92 amostras de fala para determinar a validade de critério e a validade de construto. Relativamente à validade de conteúdo, os instrumentos sistematicamente revistos para esta Tese avaliam entre 1 e 7 tipos de medidas comportamentais. As tabelas de conteúdo desenvolvidas revelaram que a maioria dos instrumentos de avaliação mede a frequência dos momentos de gaguez em percentagem de palavras gaguejadas, a duração em quantidade de tempo/unidades de repetição, os comportamentos associados utilizando descritores qualitativos/lista com diferentes tipos e os tipos de disfluência utilizando diferentes classificações; o grau de tensão e a naturalidade são avaliados apenas pelos dois mais recentes instrumentos desenvolvidos para avaliar a gravidade da gaguez. A análise quantitativa e qualitativa revelou desacordo entre os peritos consultados relativamente à clareza, simplicidade e precisão das instruções para recolha das amostras de fala. A consistência interna encontra-se garantida, uma vez que o resultado obtido para cada amostra de fala ultrapassa o *cut-off* de 0.7.

Foi obtida uma significante e larga correlação entre o resultado do *SABES* e um critério externo. O constructo subjacente à construção do *SABES* foi manifestado pela contribuição singular de cada medida comportamental, revelado através da existência de uma correlação entre 0.30 e 0.70. Através dos resultados obtidos é possível concluir que o *SABES* é um instrumento que apresenta evidências de fiabilidade e de validade de conteúdo, construto e critério.

O *ALUSCA* é um questionário que estima o efeito dos fatores ambientais em adultos, especificamente no que diz respeito à auto-perceção do nível de facilidade na utilização de competências pragmáticas de linguagem numa troca comunicativa difícil. Foram determinadas a validade de conteúdo, realizada análise de itens e obtidos coeficientes de fiabilidade e de validade de construto. O estudo piloto foi conduzido com 5 AQG e 5 controlos com vista à análise dos itens e ao cálculo de coeficientes de fiabilidade. Evidências de validade de construto foram obtidas através da aplicação do questionário *ALUSCA* a 28 AQG e a 28 controlos, utilizando análise fatorial e o método de relações hipotéticas. Relativamente à validade de conteúdo, os questionários revistos analisam um máximo de 12 competências pragmáticas de linguagem. A análise quantitativa e qualitativa revelou ambiguidades na construção de alguns itens. O estudo piloto permitiu concluir que o instrumento apresenta bons níveis de consistência interna e estabilidade temporal. As diferenças significativas entre os resultados do *ALUSCA* dos AQG e dos controlos, bem como os diferentes perfis de resposta revelaram o construto subjacente à construção do *ALUSCA*. Pode ser concluído que o ALUSCA é um questionário fiável e que apresenta evidências de validade de construto.

A tradução e adaptação cultural do *POSHA-S* contribuiu para a determinação das atitudes e conhecimento do público relativamente à gaguez, através de uma amostragem probabilística de um país (Portugal). O *POSHA-S* foi traduzido para Português-Europeu através de um processo de 5 etapas. A amostra (N=311) foi obtida através de uma amostragem probabilística por clusters (em três estádios). As atitudes da população portuguesa encontram-se na sua maioria entre os percentis 25-75. As variáveis demográficas que predisseram atitudes mais positivas foram a idade, a região, anos de escolaridade completados, situação profissional e número de línguas faladas. As variáveis demográficas que não predisseram atitudes mais positivas foram o género, o estado civil e a paternidade. Pode ser concluído que a maioria das atitudes da população portuguesa se encontram acima da média, quando comparada com a amostra total. O esquema de probabilidade utilizado permitiu a generalização dos achados.

Os instrumentos desenvolvidos serão parte de um processo de avaliação multidimensional de um AQG. Os procedimentos do *SABES* contribuirão para a determinação da gravidade dos comportamentos observáveis de gaguez de forma precisa. O questionário de auto-avaliação *ALUSCA* proporcionará a obtenção de informação precisa relativamente ao impacto das exigências pragmáticas em AQG. As atitudes da sociedade e o conhecimento acerca da gaguez serão essenciais para informar e melhorar o conhecimento da situação de um AQG numa perspetiva ampla, contribuindo para o processo de dessensibilização quanto às atitudes dos interlocutores.

keywords Stuttering, severity, pragmatic, attitudes, adults

abstract

In this study two assessment instruments (Severity Assessment Based on Events of Stuttering – SABES; Assessment of Language Use in Social Contexts for Adults – ALUSCA), were developed with the aim to assess adults who stutter (AWS), based on an adaptation of the International Classification of Functioning, Disability and Health (ICF) to the study of stuttering. A questionnaire used internationally to assess society attitudes toward stuttering (Public Opinion Survey of Human Attributes-Stuttering – POSHA-S) was also translated and cross-cultural adapted to Portuguese.

The SABES assess severity through the measurement of frequency, duration, associated behaviours, tension degree and naturalness of each stuttering moment in four speech samples based on an annotation software. Content validity, internal consistency and evidences of construct and criterion validity were determined. The content validity was analysed using a two stage process. A pilot study was conducted with five AWS to analyse the feasibility of SABES procedures. Internal consistency was analysed through Cronbach's alpha. The SABES procedures were applied to 92 speech samples to assess criterion and construct validity. Related to content validity, the instruments systematically reviewed for this Thesis assessed between one and seven types of speech behavioural measures. Tables of content revealed that the majority of the instruments measure frequency in terms of percentage of stuttered words, duration in amount of time/repetition units, associated behaviours with qualitative descriptors/list type and types of disfluencies using different classifications; tension degree and naturalness were assessed solely by the two most recent severity instruments. Qualitative and quantitative analysis revealed disagreements between experts concerning clarity, simplicity and accuracy of the speech sample collection instructions. Internal consistency was guaranteed, as the result for each speech sample was higher than the cut-off threshold of 0.7.

A significant and large correlation between SABES total score and an external criterion was found. The SABES' underlying construct was revealed by the singular contribution of each behavioural measure, as shown by a useful degree correlation (i.e., between 0.30 and 0.70). It could be concluded that SABES is a reliable and presented evidences of content, construct and criterion validity.

The ALUSCA estimates the effects of environmental factors on adults, specifically related to the self-perception of the level of ease in using pragmatic language competencies (PLC) on a difficult communicative exchange. Content validity analysis, item analysis, reliability coefficients and evidences of construct validity were analysed. The content validity was analysed using a two stage process. A pilot study was conducted with five AWS and five controls to analyse items and to calculate reliability coefficients. Construct validity evidences were obtained through ALUSCA application to 28 AWS and 28 controls, using the hypothesised relationships method and factor analysis. Concerning content validity, the questionnaires reviewed assessed up to twelve PLC. Qualitative and quantitative analysis revealed ambiguities in items construction. The pilot study showed that the instrument presented internal consistency and temporal stability. Significant differences between AWS and controls, and different response profiles revealed ALUSCA's underlying construct. It could be concluded that ALUSCA is a reliable and presented evidences of construct validity.

The translation and cross-cultural adaptation of the POSHA-S contributed to the determination of public attitudes and knowledge toward stuttering in a probability sampling of an entire country. The POSHA-S was translated to European Portuguese through a five-step process. A sample (N= 311) was collected through a three-stage cluster probability sampling, with a local administrative office-based. The attitudes of the Portuguese population were generally in the interquartile range. The demographic variables that predicted more positive stuttering attitudes were respondents' age, region of the country, years of school completed, working situation, and number of languages spoken. Non-predicting variables were respondents' sex, marital status, and parental status. It could be concluded that the majority of attitudes of the Portuguese population fell in the interquartile, meaning that POSHA-S scores were about average, compared with the total data sample. The probability sampling scheme used allows the generalization of the findings.

The instruments developed will be part of a broader and multidimensional assessment process of an AWS. The SABES procedures will contribute to the accuracy of the severity determination. Information collected through the self-assessment ALUSCA questionnaire will provide accurate information regarding the impact of pragmatic demands on AWS. The society attitudes and knowledge toward stuttering will be essential to inform and improve the understanding of an AWS's situation in a broader perspective, contributing to the desensitization process concerning other's communication attitudes.

Table of Contents

List of figures

List of tables

Abbreviations

AWS – Adults who stutter

ICF – International Classification of functioning, Disability, and Health

WHO – World Health Organization

PWS – Person/People who stutter

ICIDH – International Classification of Impairment, Disabilities, and Handicaps

%SS – Percentage of stuttered syllables

%SW – Percentage of stuttered words

SLD – Stuttering like-disfluencies

LBDL – Lidcombe Behavioural Data Language

SR – Syllable repetition

ISR – Incomplete syllable repetition

MUR – Multisyllable unit repetition

FPWithAA – Fixed postures with audible airflow

FPWithoutAA – Fixed postures without audible airflow

VEB – Verbal extraneous behaviours

NVEB – Nonverbal extraneous behaviours

SLT/SLTs – Speech and language therapist/therapists

TOCS – Test of Childhood Stuttering

SDA – Systematic disfluency analysis

FIIS – Fluency Interview Index of Severity

TOM – Therapy outcome measure

SSI – Stuttering Severity Instrument

SSI-3 – Stuttering Severity Instrument – 3rd edition

SSI-4 – Stuttering Severity Instrument – 4th edition

CSSS – Computerized scoring of the stuttering severity

ELAN – EUDICO Linguistic Annotator

VAS – Visual analogue scale

SPSS – Statistical package for the social sciences

PLC – Pragmatic language competencies

IPATHA – Project on attitudes toward human attributes

OSS – Overall stuttering score

EP – European Portuguese

NUTS – Nomenclature of Territorial units for statistics

MANOVA – Multivariate analysis of variance

Chapter 1 – Research Overview

1.1 Introduction

The research presented in this Thesis aims to contribute to a comprehensive assessment process for Portuguese Adults Who Stutter (AWS) based on the International Classification of Functioning, Disability and Health (ICF) (WHO, 2001). The ICF is a framework developed by the World Health Organization (WHO) to contextualise the experience of health (both positive and negative), emphasizing functioning and quality of life, rather than difficulties and disadvantages associated with a disability. Stuttering is a complex disorder, with individual differences in both observable/nonobservable behaviours, in the self-experience and on the overall impact on individual's life (Yaruss & Quesal, 2006). The ICF, as a framework designed to express the overall human health experience, can be used to capture the entire range of behaviours and experiences of stuttering (Yaruss, 2007). Yaruss and Quesal (2004b, 2006) proposed an ICF based framework to the assessment/intervention on stuttering; in this framework, the observable stuttering behaviours (impairment in body function) influence and are influenced by personal reactions and environmental factors (which, in turn, influence each other) and these mutual influences can have an impact on communication activities participation and, globally, in quality of life.

The assessment and intervention process of someone who stutters should comprise detailed evaluations of all factors that influence the experience of stuttering.

This work contributes to a comprehensive assessment process of AWS, through the creation and validation of two assessment instruments; a study of the knowledge and attitudes of the Portuguese population toward PWS, to contextualise the influence of society responses (environmental factors) on the experience of stuttering is also presented.

1.2 Overview

1.2.1 The multidimensionality of stuttering

Stuttering is a multidimensional and heterogeneous communication disorder, with an early onset during the development of speech and language (Bloodstein & Bernstein Ratner, 2008; Manning, 2010). Stuttering is characterised by involuntary disruptions to the fluency of speech, that takes the form of syllable, word and part-word repetitions, prolongations and

blocking of sounds, that are inconsistent and variable (Bloodstein & Bernstein Ratner, 2008; Guitar, 2014; Yaruss, 2007). This fluency disorder can negatively impact functioning in different areas of daily life of both children and adults, who report frequently a feeling of "loss of control" before, during and after the stuttering event (Blood & Blood, 2016)g. Negative reactions to speech difficulties (e.g., embarrassment, shame, anxiety and fear) (Blumgart, Tran, & Craig, 2010; Iverach et al., 2009; Iverach & Rapee, 2014; Tran, Blumgart, & Craig, 2011) can lead to the appearance of associated behaviours, such as tension/struggle or avoidance, produced to minimise the observable stuttering events (Bloodstein & Bernstein Ratner, 2008; Coleman & Yaruss, 2014; Guitar, 2014). Simultaneously, a Person Who Stutters (PWS) can develop negative behavioural, cognitive, and affective reactions derived from years of stuttering experiences. Feelings and attitudes can limit PWS in familiar, social and professional situations (Craig, Blumgart, & Tran, 2009; Yaruss & Quesal, 2004b). As a social communicative disorder, reactions and attitudes of the interlocutors impact PWS and, consequently, play a role in the progression and maintenance of stuttering (Blood & Blood, 2016). PWS are aware of the society stigmatised and stereotyped view of stuttering, which can lead to the internalisation of stigma (Boyle, 2013; Boyle & Blood, 2015). The anticipation of negative societal reactions due to stuttering is common (related to higher levels of social anxiety), leading to avoidance of speaking in certain situations, which can affect quality of life and difficulties in achieving life goals (Boyle, Dioguardi, & Pate, 2016; Butler, 2013; Plexico, Manning, & Levitt, 2009).

1.2.2 Stuttering and the International Classification of Impairments, disabilities, and handicaps (ICIDH)

The characteristics of stuttering mirror the broad-based and multifaceted nature of this disorder (Yaruss, 2007). To comprehensively define stuttering is critical to consider this fluency disorder within a broad-based framework, that could be, simultaneously, broader in the inclusion of a wide range of behaviours, experiences and consequences and specific enough to incorporate individual differences (Yaruss, 2007). As Yaruss (2007, p. 314) stated, the "(…) availability of a comprehensive method of describing stuttering would yield significant benefits, not only for conveying the nature of the disorder, but also for the evaluation of the treatment approaches that focus on different components of the disorder".

To capture the concept that "stuttering is more than just stuttering", Yaruss (1998a) applied two frameworks developed by the WHO to the study of stuttering (i.e., the original framework, ICIDH, and a revised framework, ICF).

The ICIDH (WHO, 1980) was developed by the WHO in 1980 and had the main objective of classifying the consequences of disease, injuries and other disorders and their implications for the lives of individuals This model described the consequences of diseases and disorders in terms of impairment, disability and handicap. Impairment was defined as "(…) any loss or abnormality of psychological, physiological, or anatomical structure or function" (WHO, 1980, p. 27); disability represented "(…) any restriction or lack (resulting from an impairment) of ability to perform an activity in the manner or within the range considered normal for a human being" (WHO, 1980, p. 28); handicap referred to "(…) a disadvantage for a given individual, resulting from an impairment or a disability that limits or prevents the fulfilment of a role that is normal (depending on age, sex, and social and cultural factors) for that individual" (WHO, 1980, p. 29).

The ICIDH intended to offer a common language for the discussion of multiple aspects of disability, within a conceptual framework for information, applicable to the long-term consequences of diseases or disorders, to personal health care and to the mitigation of environmental and societal barriers (WHO, 1980).

Several authors (Curlee, 1993; McClean, 1990; Prins, 1991) applied the ICIDH framework to stuttering, considering impairment as the underlying aetiology of the observable events of stuttering (described as the disability) and the disadvantages that result from reactions to stuttering events as the handicap (Yaruss, 1998a). Yaruss (1998a) pointed out that the definitions used with consensus among some researchers were not consistent with those provided in the ICIDH, whereas the main objective of the framework was to define the consequences of disorders and not the aetiology. As the differences in terminology could increase confusion in the use of the framework, the terms impairment, disability and handicap for stuttering were redefined to match the ICIDH purposes: Impairment were the disruptions of speech-language production and the tension associated; disability were considered as the limitations in the ability to communicate or to engage in social/work-related activities; handicap comprised the limitation experienced by an individual who stutters, as a result of impairment or reactions by himself or others.

To unify the terminology related to the outcomes of stuttering treatment in the context of disorder consequences, Yaruss (1998a) proposed a flexible model representing the application of ICIDH to stuttering (see Figure 1-1). The model follows the general structure of ICIDH (i.e., impairment can lead to disability, that can lead to handicap), but also include a presumed aetiology, personal reactions and environmental influences that influences the development of handicaps. Affective, behavioural and cognitive reactions (ABCs reactions

of stuttering) can mediate the occurrence of disabilities in individuals who stutter, so “any unifying framework for discussing treatment outcomes in stuttering must be flexible enough to incorporate the effects of the ABC reactions of an individual’s ability to enter social situations or achieve life goals” (Yaruss, 1998a, p. 254).

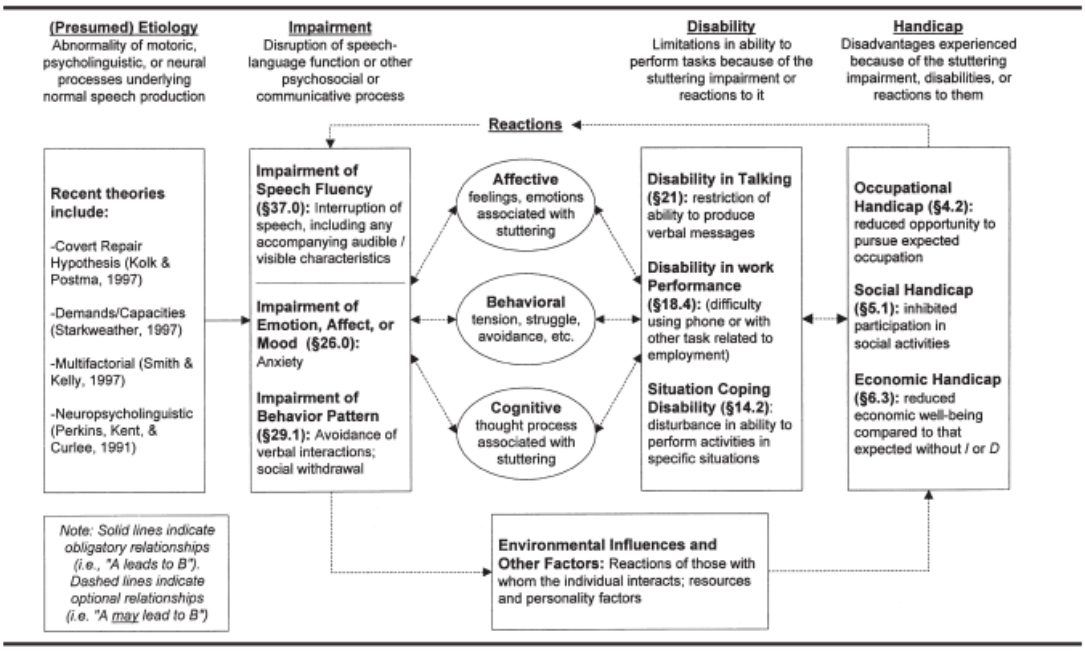


Figure 1-1 Model of the ICIDH applied to the study of stuttering (Yaruss, 1998a). The model presented specific ICIDH codes appropriate for stuttering. From Yaruss (1998a, p. 254).

1.2.3 Stuttering and the International Classification of Functioning, disability, and Health (ICF)

Several authors from different areas of health, disability and rehabilitation identified aspects that were not adequately represented and described within the ICIDH framework (Yaruss & Quesal, 2006). The complexity of the three-tiered model, difficulties in the definition of terms and the failure to account for individual differences between subjects that could improve or worsen the experience of disability or handicap are examples of ICIDH weaknesses (Brandsma, Lakerveld-Heyl, Van Ravensberg, & Heerkens, 1995; Yaruss & Quesal, 2006).

In 2001 the WHO presented the ICF, a revision and update of the ICIDH (WHO, 2001). The ICF simplified the ICIDH three-part classification in two primary levels, and added contextual factors (omitted on the ICIDH framework). The two primary levels that described health-related experiences comprise the structure and body functions, activities and participation

and a set of contextual factors. Difficulties with structure and body functions are denominated impairments (similar to ICIDH), and difficulties in the activity and participation are called activity limitations and participation restrictions. The framework had codes to indicate the specific nature of the disorder and subcodes to specify the severity (Yaruss, 2007). As an expansion of the ICIDH, the ICF was not limited to the description of the consequences of disorders, but incorporated both positive and negative experiences, affected by obstructing and facilitating forces. This revised framework incorporates personal and environmental factors (contextual factors), to capture individual differences.

Concerning the structure, the ICF described the components of health in two parts: 1) functioning and disability and 2) contextual factors. Part 1 (Functioning and Disability) provides a summary of both positive and negative aspects of the experiences that individuals might have regarding their health at different levels and in a variety of domains. The first part was separated into two components: Body functions and structures and activities and participation (WHO, 2001; Yaruss, 2007; Yaruss & Quesal, 2004b, 2006). Body functions and structures replaces the list of impairments proposed on the ICIDH and covers physiological and psychological functions as well as the anatomical structures that assist those functions. Sets of codes described body functions ("b" codes) or structures ("s" codes) and impairments are defined using additional qualifying codes. Qualifiers (i.e., 0 - no problem or within normal limits; 1- mild; 2- moderate; 3- severe; 4- complete or profound) represent the severity degree of the impairment (WHO, 2001; Yaruss & Quesal, 2004b). Activities and participation substitutes the list of disabilities and handicap of ICIDH and covers a comprehensive variety of life areas. Different items within the nine basic aspects of life covered by activities and participation can be qualified in terms of individual's performance and capacity. Performance is defined as what the person does in the current environment and capacity as what the person can do in an optimal environment. Difficulties in activities are defined as limitations and problems in the engagement in life situations as restrictions in participation (WHO, 2001; Yaruss & Quesal, 2004b).

The ICF's part 2 is divided in two components (environmental factors and personal factors) that comprises a variety of factors that may influence functioning and disability. The inclusion of contextual factors captures individual differences concerning the experience of health and/or disorder (Yaruss, 2007). Environmental factors encompass individual (i.e., the immediate environment) and societal external influences (social structures or attitudes) that affect in a positive or negative way the individual health experience (WHO, 2001; Yaruss & Quesal, 2004b). The environmental factors are divided into 5 chapters. Positive influence is indicated by a plus sign and negative influence by a period, which follows a qualifier (0-

4). The ICF framework also considers the positive or negative influence of personal factors (e.g., past experiences, coping style, personality and character characteristics) on functioning and disability and, generally, on health-related experience. Based on the huge number of personal factors, with social and cultural differences that needs to be considered, no specific list is proposed on the ICF (WHO, 2001; Yaruss & Quesal, 2004b, 2006).

Due to the specific characteristics of stuttering as a fluency disorder that exhibits observable speech disruptions in addition to broader consequences in different aspects of life (e.g., social communication or job-related tasks), the WHO ICF framework "provides an ideal framework for considering the overall experience of the stuttering disorder" (Yaruss, 2007, p. 312), with sufficient specificity for describing the variety of experiences from the perspective of individuals who stutter.

Each component of the ICF framework can be related to stuttering (Yaruss, 2007; Yaruss & Quesal, 2004b). Concerning body functions, there are several sections that address communication. Voice and speech functions (chapter 3) is the most relevant chapter for stuttering, which incorporate aspects related to fluency and rhythm of speech functions (Section b330). Impairments in body function in PWS can be related to fluency of speech (b3300), rhythm of speech (b3301), speed of speech (b3302) and melody of speech (b3303) (Yaruss, 2007; Yaruss & Quesal, 2004b). A qualifier can be associated to the codes, to specify the severity. Some individuals who stutter may also presented affective and cognitive reactions that can affect mental functions, which can include impairment in global psychosocial functions (b122), temperament and personality functions (b126) or emotional function (b152). Concerning body structures, as research findings indicated a possible structural difference in the nervous system of children who stutter and AWS (Chang, Erickson, Ambrose, Hasegawa-Johnson, & Ludlow, 2008; Foundas, Bollich, Corey, Hurley, & Heilman, 2001; Sommer, Koch, Paulus, Weiller, & Büchel, 2002) a code related to structures of the brain (section s110) may be applied for stuttering. Sections within the list of structures involved in voice and speech (chapter 3) would not be used for classifying stuttering, as there is no clearly identified structural deficit in these structures on developmental stuttering (Yaruss, 2007; Yaruss & Quesal, 2004b). The classification for activities and participation includes several components that might be affected by stuttering. Chapter 3 (Communication) is the most relevant and directly related to stuttering, as the person may experience difficulties in, e.g., starting a conversation (d3500), ending a conversation (d3502) or participate in a discussion with one person (d3550). To contemplate the broader effect of stuttering in activity and participation, other limitations throughout chapter 6 (domestic life), chapter 7 (interpersonal interaction and relationship),

chapter 8 (education and employment) and chapter 9 (community, social, and civil life) could be classified (Yaruss, 2007; Yaruss & Quesal, 2004b). The two qualifiers of the Activities and Participation component comprises the performance and the capacity. Related to stuttering, it may be impractical to identify an environment that is standard and similar for all persons (WHO, 2001; Yaruss, 2007; Yaruss & Quesal, 2004b), because an optimal fluency situation differs for different individuals. Consequently, it is more appropriated to assess the person's performance (not capacity) in daily activities and participation across common environments and social contexts (Yaruss, 2007; Yaruss & Quesal, 2004b).

Through the Contextual factors components (Environmental Factors and Personal factors), individual differences in the experience of stuttering can be described (Yaruss, 2007). Concerning environmental factors, there are several factors that can influence positively or negatively the experience of PWS: Products and Technology (Chapter 1) describe devices that can influence the fluency of PWS; Support and Relationships (Chapter 3) presented a list of individuals that can facilitate or prejudice an individual who stutter; Attitudes (Chapter 4) describes the attitudes of particular person on the individual's environment as well as the influence of societal attitudes; Services, systems and policies (Chapter 5) describes communication services, education and training services and association and organization services that can affect the experience of stuttering. The addition of personal reactions on the ICF improves the framework in the ability of accurately describe the nature of stuttering and "facilitates the description and ultimately assessment of treatment procedures and outcomes" (Yaruss, 2007, p. 317). Due to differences in personal experience across individuals and cultures, there isn't a specific list of personal contextual factors (WHO, 2001; Yaruss & Quesal, 2004b). Personal factors were defined by the WHO as "gender, race, age, other health conditions, fitness, lifestyle, habits, upbringing, coping style, social background, education, profession, past and current experience (past life events and concurrent events), overall behaviour pattern and character style, individual psychological assets and other characteristics, all or any of which may play a role in disability at any level" (WHO, 2001, p. 23). Researchers on fluency disorders address personal factors in terms of affective, behavioural and cognitive (ABC) reactions to stuttering These reactions can affect not only the fluency in a specific situation, but the overall experience of the stuttering disorder (Manning, 2010; Yaruss & Quesal, 2004b, 2006). Examples of negative feelings are embarrassment, fear, anxiety, shame and examples of positive feelings are hope, acceptance and optimism; behavioural reactions can include physical tension, struggle and avoidance; cognitive reactions can incorporate negative self-evaluation and reduced self-esteem or confidence in speaking ability (Yaruss & Quesal, 2004b). Yaruss (2007, p. 317)

referred that the ABC reactions “are particularly important in stuttering and other communication disorders because they determine, at least in part, the extent to which a speaker will experience limitations in daily activities or participation restrictions (…)”.

Yaruss and Quesal (2004b, 2006) adapted the ICF frameworks to the study of stuttering, developing a model, which was a meaningful way to interpret the experience of PWS (see Figure 1-2).

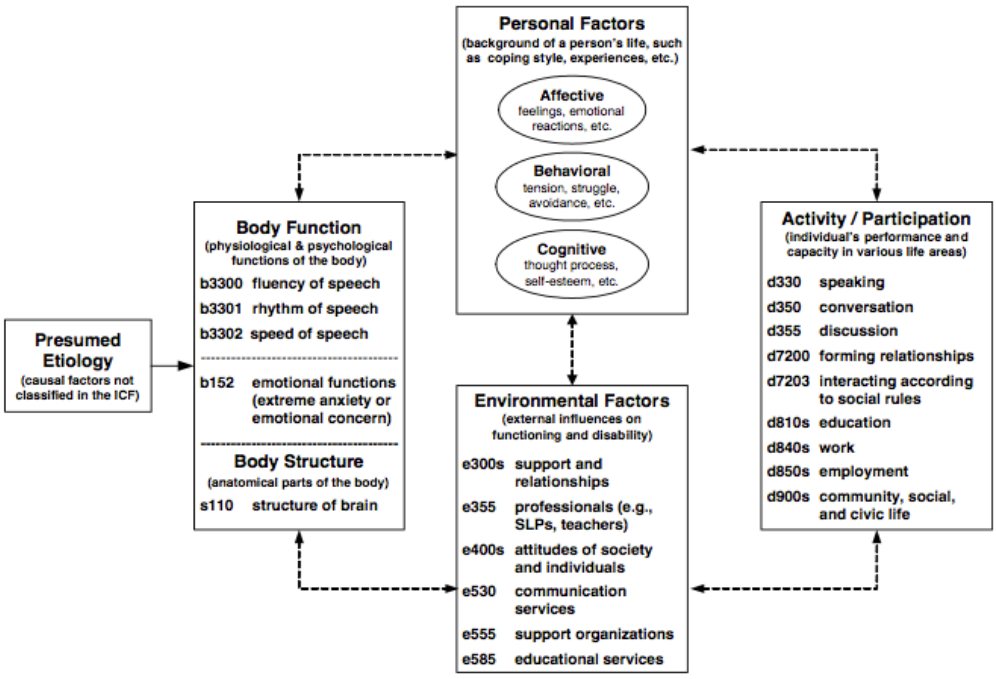


Figure 1-2 Graphical representation of the application of the World Health Organization’s International Classification of Functioning, Disability, and Health (ICF) to stuttering. The model presented specific ICF codes appropriate for stuttering. From Yaruss and Quesal (2004, p. 48).

This model can be considered in terms of several interacting components: The presumed aetiology or underlying cause(s) of the disorder (causal factors not classified in the ICF); the impairment in body function, indicated by the observable characteristics of stuttering; the speaker’s affective, behavioural, and cognitive (ABC) reactions to stuttering; the effects of the environment on stuttering, indicated by the difficulty in different speaking situations and the reactions of others; the overall impact of stuttering on the speaker’s life, indicated by limitations in communication activities and restrictions in participation in daily life (Yaruss & Quesal, 2004b, 2006). The model demonstrates that PWS can experience ABC reactions (personal factors) that can limit the participation in communication related activities. The

ability to participate can also be positively or negatively affected by attitudes and reactions provided by the environment. These limitations can also affect both the speaker's reactions to stuttering and the reactions of interlocutors in the speaker environment. The model also highlight that reactions received from other people can influence the individual ABC reactions and how the speaker's reaction to their stuttering experience can influence their communicative environment. Those reactions (both individual ABC reactions and environmental attitudes) can also affect the observable stuttering behaviours (Yaruss & Quesal, 2004b, 2006). It is also important to highlight that, isolated, the impairment severity (i.e., severity of the observable disfluencies of stuttering) does not regulate the degree of limitations or restrictions in activity/participation. Impairment and personal/contextual differences interact to determine the magnitude in which a PWS experience activity limitations or participation restrictions (Yaruss, 2007).

1.2.4 Assessment of stuttering in the context of the ICF framework

Through the ICF and the ICF adaptation to the study of stuttering developed by Yaruss and Quesal (2004b, 2006), it is clear that stuttering is more than the observable behaviours, and that should be perceived in all of its dimensions, concerning the observable stuttering behaviours (impairment in body function), the experience of stuttering from the perspective of the speaker (personal reactions and the overall impact of stuttering in activities and participation) and the effect of environment (indicated by difficulty in communicating in different situations and the reactions of other speakers).

Understanding the multiplicity of stuttering disorder involves the integration of all components mentioned above in a comprehensive assessment and intervention process (Coleman & Yaruss, 2014; Yaruss, 2007; Yaruss & Quesal, 2004b, 2006). The impairment in body function (i.e., the observable disfluencies) can be assessed through the measuring of speech disfluencies and the determination of severity; speech efficiency, spontaneity in communication, frequency and severity of disfluencies, physical concomitants and tension degree are examples of assessment areas related to impairment (ASHA, 2014). The assessment of the experience of stuttering from the perspective of the speaker implies the assessment of less observable characteristics, including emotional responses to stuttering and communication, the assessment of self-confidence and attitudes/reactions about stuttering (ASHA, 2014; Yaruss & Quesal, 2006). Several instruments (e.g., Andrews & Cutler, 1974; Brutten & Shoemaker, 1974; Ornstein & Manning, 1985; Riley, Riley, & Maguire, 2004; Wright & Ayre, 2000; Yaruss & Quesal, 2006) contain items to assess

affective, behavioural and cognitive reactions, as well as the impact on the speaker's quality of life. Determining the impact of environmental factors implies daily communication analysis, assessment of reactions, attitudes and opinions of the interlocutors or the assessment of society knowledge about stuttering (ASHA, 2014; Louis, Reichel, Yaruss, & Lukber, 2009; St. Louis, Lukber, Yaruss, Adkins, & Pill, 2008; Yaruss & Quesal, 2004b, 2006). Activities and participation assessment involves the determination of the individual ability to achieve education and personal objectives or the impact of the fluency disorders on quality of life (ASHA, 2014; Yaruss, 2007).

Through a comprehensive assessment process, it is possible to obtain baseline data to assess the efficacy of the intervention and to determine strengths and weaknesses, in order to developed an adequate treatment plan to the person that is being assessed (Yaruss, 2007).

1.3 Aims of the present study

Internationally, there are several instruments that can be used in an assessment process of AWS. The characterisation of the observable stuttering behaviours (to assess impairment in body functions) can be made using simple severity scales or validated instruments based on the analysis of a restricted number of speech samples. The influence of environmental factors, specifically the difficulties felt on entering speaking situations, are assessed using different questionnaires that include some items related to the use of language (i.e., pragmatic competencies needed to enter a communicative situation). However, the assessment procedures currently available are limited in different aspects: Concerning the assessment of stuttering events, a restricted number of speech samples are usually collected and assessed with standardised procedures that are substantially based on perception; related to the impact of stuttering on the use of language, the available questionnaires do not cover all aspects of pragmatic competencies, which may lead to an incomplete assessment of stuttering influence on pragmatics.

The aims of the present study are therefore:

1. To improve the assessment of the impairment on body functions by developing an assessment instrument design to measure frequency, duration, associated behaviours, tension degree and naturalness of each stuttering event, based on a large number of speech samples (four audio and video recorded samples) and to clearly define the procedures (with a free annotation software);

2. to estimate the effects of environmental factors, particularly associated with the competencies to participate in communicative situations, by the creation and validation of a self-report to assess the level of ease in pragmatics on a communicative exchange
3. to contribute to the determination of public attitudes toward stuttering in Portugal with a questionnaire completed by a representative sample, as there are no published research studies that explore attitudes in an entire country, anywhere in the world.

1.4 Thesis overview

This Thesis describes the development and validation of two assessment instruments, and the translation/cross-cultural adaptation and data collection through a probability sampling of an international published questionnaire. The assessment instrument design and the validation process of the *SABES*, an instrument created to assess impairment in body functions, is described in Chapter 2. Chapter 3 presents the description of the validity and reliability analysis of the *ALUSCA*, an assessment instrument developed to assess language use (pragmatics) on a communicative situation. The translation and adaptation process of the *POSHA-S* for European-Portuguese, the country-wide probability sample data collection and the conclusions concerning the attitudes, knowledge and predictive demographic variables of an entire country (Portugal) are presented in Chapter 4. Conclusions and future work are presented in Chapter 5.

Chapter 2 – Severity assessment Based on Events of Stuttering (*SABES*)

2.1 Introduction

The Severity Assessment Based on Events of Stuttering (*SABES)* is an assessment tool designed to assess the impairment on body functions, in terms of the severity of the observable stuttering behaviours. The *SABES* has the objective to improve the procedures used to determined stuttering severity on AWS. In this chapter, the validity and reliability analysis of the *SABES* is presented.

2.2 Background

2.2.1 Surface and intrinsic features of stuttering

The American Psychiatric Association defined stuttering as a disturbance in the normal fluency and time patterning characterised by frequent repetitions, prolongations of sounds and syllables, monosyllabic whole-word repetitions, broken words, audible or silent blocks, circumlocutions and excess of physical tension, that interfere with academic/occupational achievement and with social communication (American Psychiatric Association, 2013). According to Yaruss and Quesal's (2004b, 2006) framework adaptation of the World Health Organization's ICF, this multi-dimensional disorder is characterised by observable behaviours, reactions of the PWS (affective, cognitive and behavioural) and environmental reactions. Those three factors influenced each other and are influenced by the overall impact of stuttering on speaker's life, indicated by limitations in communicating in activities and restrictions in daily life participation (Yaruss & Quesal, 2004a, 2004b, 2006).

The multidimensionality of stuttering is characterised by intrinsic (or nonbehavioural) and surface (or behavioural) features (Manning, 2010). The intrinsic features include feelings, attitudes and coping responses of PWS. The observable disruptions to speech fluency are the surface features of stuttering. The features that the listeners can see and hear (in audio and video recordings) are characterised by frequency, duration, tension associated with the stuttering moments and associated features used by the speaker to escape or avoid stuttering (Guitar, 2014; Manning, 2010). Clinically, it is appropriate to consider surface

features as momentary speech disruptions (i.e., stuttering moments) bounded by normal sounding speech (Onslow, 2016).

Severity, in a global sense, characterises/typifies the seriousness, magnitude or the significance of a certain disorder and is labelled as mild, moderate or severe (Flipsen Jr, Hammer, & Yost, 2005; Paul & Norbury, 2012). Measuring severity is important to establish intervention priorities, to establish a baseline, to determined prognosis, to have a benchmark to assess the effectiveness of an intervention and, globally, to predict accurately the level of functioning, the response to treatment and the impact of a disorder on the quality of life of an individual (Gordis, 2014; Moran & Crawford, 2013; Paul & Norbury, 2012). Concerning stuttering, severity assessment entails documenting and, especially to quantifying the impact of both intrinsic and surface characteristics of stuttering on PWS (Onslow, 2016). Both surface and intrinsic features are important to severity determination; the level of the problem of an individuals who stutter, or their response to the problem, may have different impacts in surface and intrinsic features (Manning, 2010). A person who stutters may presented frequent moments of stuttering with tension and struggle (considered as severe) and a low impact on reaction to communication and life choices; the other extreme is also possible, when PWS present a low frequency of moments of stuttering with a huge impact on quality of life; the majority of the speakers falling in the extremes described above (Manning, 2010). Due to the importance of both aspects of stuttering, "the primary task for the clinician is to help the client map both the surface behaviours of stuttering as well as the intrinsic features of the problem" (Manning, 2010, p. 162).

Specifically concerning surface features of stuttering (the focus of the present chapter), clinical measurement of stuttering is important and, consequently, necessary to determine the severity of moments of stuttering, to state and document treatment goals, to assess progress during the intervention process and to monitor the post-treatment progress (Guitar, 2014; Onslow, 2017). However, and despite the easiness of surface feature determination (because the behaviours are observable), the variable nature of stuttering causes difficulties in the assessment process. Stuttering severity varies in different speaking situations, with different interlocutors, audience sizes or conversation topics (Bothe, 2004; Guitar, 2014; Manning, 2010; Onslow, 2017). Combined with the variability of stuttering, the ability of some AWS to avoid the stuttering moments during the assessment can lead to unrepresentativeness, misdiagnosis or an erroneous perception of severity (Manning, 2010). This could lead to several difficulties to complete the assessment process. To overcome those difficulties, the assessment process should involve a variety of speaking situations, and "the more these situations simulate the speaker's daily communication

situations the more apt we are to obtain a true indication of the problem" (Manning, 2010, p. 150). To obtain a more representative perception of the impact of stuttering (both concerning surface or intrinsic features), self-assessment by those who stutter is also a very useful approach for the clinician to get knowledge and keep track of stuttering variability (Onslow, 2017). Self-assessment reveals the way a person understands his own stuttering and considers his/her severity level (Manning, 2010).

2.2.2 Assessment of surface features characteristics

The assessment process begins with the identification of both surface and intrinsic features that are characteristic of PWS. According to Manning (2010), there are three basic characteristics of the surface features: Frequency, duration and tension.

The frequency of the fluency breaks is one of the most noticeable features of the stuttering problem, that can, to some extent, impact the perception of stuttering severity, especially as perceived by the listener (Manning, 2010). Frequency can be calculated as percentage of stuttered syllables (%SS) or percentage of stuttered words (%SW) (Guitar, 2014; Manning, 2010). The examiner could obtain %SS through the calculation of total number of syllables produced by the speaker and the indication of syllables in which stuttering occurs.

Percentage of stuttered syllables is agreed as the preferable outcome measure, as the timing of the speech movements is related to syllable size and not to word size (Allen, 1975; Starkweather, 1987; Stetson, 1951). Guitar (2014) considers the %SS preferable to %SW because a multisyllabic word may encompass more than one syllable stuttered. Additionally, Guitar (2014) also referred that the number of syllables can be calculated more easily than words through the count of syllables beats while the person is talking.

Due to reliability difficulties related to frequency counts (Coyle & Mallard, 1979; Curlee, 1981; Emerick, 1960; Ham, 1989; Ingham & Cordes, 1992; Kully & Boberg, 1988; MacDonald & Haroldson, 1973; Martin & Haroldson, 1981; Martin, Haroldson, & Woessner, 1988; Young, 1975) that could lead to different diagnosis or severity descriptions, an alternative measurement system, based on behaviours occurring within a defined time-interval was proposed (Cordes, Ingham, Frank, & Ingham, 1992; Schloss, Freeman, & Smith, 1987). Time-interval analysis does not focus on individual events but on the presence of a stuttering event within a short interval of time, in which the judges made a binary judgment (i.e., each interval contains at least one moment of stuttering or no stuttering). A systematic review (Valente, Jesus, Hall, & Leahy, 2014) of the reproducibility

(i.e., absolute and relative reliability parameters) of event-based and time-interval measurement concluded that it was unviable to quantify the agreement between inter-judge reliability, intra-judge reliability and accuracy values determined using the two methods in order to quantify which methodology represents the more reproducible and more accurate method to assess frequency because the studies did not present the measures in the same metric scale. The review also revealed that the use of trained/experienced judges and the small number of samples used in the determination of inter and intra-judge reliability could contribute to agreement values beyond the references for good reproducibility values in both methodologies (Baer, Wolf, & RIsley, 1987; Hinkle, Wiersma, & Jurs, 2003; Landis & Koch, 1977; McHugh, 2012; Nunnally & Bernestein, 1994).

Stuttering moments can also be classified in terms of types of disfluencies. Johnson et al. (1959), Johnson (1961) and Johnson, Darley, & Spriestersbach (1963) described disfluencies in terms of interjections, part-word repetitions, word repetitions, phrase repetitions, revisions, incomplete phrases, broken words and prolonged sounds. Based on the classification described above and on research findings that PWS are more likely to produce certain types of disfluencies, several classification schemes recognise disfluencies that are most likely to be moments of stuttering (Yaruss, 1997). Conture (1990a, 1990b) defines “within-word disfluencies” (i.e., monosyllabic whole-word repetition; sound/syllable repetition; audible prolongation and inaudible prolongation) as those typically produced by PWS and “between-word repetition” (i.e., phrase repetition; polysyllabic whole-word repetition; interjection; revision) as the disfluencies that individuals who do not stutter are more likely to produce. As “not all within-word disfluencies are stuttered, and not all between-word disfluencies are nonstuttered” (Yaruss, 1997, p. 36), other labels were used to describe disfluencies that are characteristic of PWS. Meyers (1986) used the labels *stutter-type disfluencies* (i.e., part-word repetition; prolongation; broken word; tense pause) and *normal-type disfluencies* (i.e., whole-word repetition; phrase repetition; revision; incomplete phrase; interjection). Gregory (1986; 1993) and Campbell and Hill (1987) used the terms *less-typical disfluencies* to identify disfluencies typically produced by PWS (i.e., three or more units on monosyllabic word repetition; three or more units on part-word syllable repetition; sound repetition; prolongation; block) and *more-typical disfluencies* to characterise disfluencies typically produced by people who do not stutter (i.e., hesitation; interjection; revision; phrase repetition; two or fewer no tense monosyllabic word repetition; two or fewer no tense part-word syllable repetitions). The labels *less-typical disfluencies* and *more-typical disfluencies* are used on the Systematic Disfluency Analysis (Campbell & Hill, 1987). Stuttering-like disfluencies (SLD) are one of the most employed disfluency-type

classifications. They were developed by Yairi and Ambrose (1992), as an enhancement of the eight categories classification of Johnson (1961; 1959): Word repetition, sound/syllable repetition, phrase repetition, incomplete phrases, interjection, revision, broken word and prolongation. Stuttering-like disfluencies include part-word and monosyllabic word repetitions, dysrhythmic phonation (including sound prolongation, silent blocks, broken words and other within word interruptions that interrupt the continuity of a word) and tense pause (Yairi & Ambrose, 1992; Yaruss, 1997). Due to difficulties in a reliable measure, tense pauses were omitted from the SLD classification (Ambrose & Yairi, 1999). Einarsdóttir and Ingham (2005), in a review related to the contribution of disfluency-type measures to the understanding and treatment of developmental stuttering, concluded that, due to poor reliability of disfluency type classifications, they should be regarded as "imprecise descriptors of observable stuttering and not a fundamental measure of stuttering" (p. 260). Additionally, Einarsdóttir and Ingham (2005) concluded that differences between children who stutter and children who do not stutter are mostly related to the amount of observable stuttering and only partially with differences in types of disfluencies. As an attempt to improve the description accuracy of perceptually identified stuttering events, the Lidcombe Behavioral Data Language (LBDL) was developed (Bryant & Packman, 1999; Packman & Onslow, 1998; Packman, Onslow, & Bryant, 2000); this taxonomy presented an acceptable higher interjudge agreement in 10 seconds samples (Teesson, Packman, & Onslow, 2003). The LBDL was developed to be a valid and reliable descriptor of stuttering behaviours that can be used in all ages. The LBDL consisted of a taxonomy of stuttering that should be used with moments previously identified as stuttering, based on the idea that a stuttering moment is everything that is labelled stuttering and has a relatively good agreement between two judges (Bloodstein & Bernstein Ratner, 2008). This taxonomy joins the various disfluency descriptors into three categories and seven descriptors within the three categories. The three prime categories are *repeated movements*, *fixed postures* and *superfluous behaviours* and there are variations of terms, such as repetitions, prolongations and accessory features (Onslow, 2017; Teesson et al., 2003). The first prime category (*repeated movements*) describes recurring movements and includes syllable repetition (SR), i.e., repeated movement of an entire syllable, incomplete syllable repetition (ISR), i.e., repetition of parts of syllables and multisyllable unit repetition (MUR), i.e., repetitions that involve more than one syllable. The second prime category (*fixed postures*) describes an "abnormality of no movement" (Onslow, 2017; Teesson et al., 2003) as the movement of the elements that contribute to speech remains fixed. *Fixed postures With Audible airflow* (FPWithAA) is the first sub-category of fixed postures and refers to prolongation of sounds.

Fixed postures Without audible airflow (FPWithoutAA) is the second sub-category of fixed postures, also referred to by clinicians as “blocks”, because the impression is that something blocks the speech. The third prime category includes verbal and nonverbal extraneous behaviours. *Verbal extraneous behaviours* (VEB) are referred in the taxonomy of Wendell Johnson to interjections. *Nonverbal extraneous behaviours* (NVEB) are the most idiosyncratic features of stuttering and include “compressed lips, open mouth, breath holding, blinking, nostril dilating, eyebrow raising, grimacing, facial, head, and torso movements, inspiratory airflow, grunts and other inappropriate noises, and aberrant fluctuations in pitch and loudness” (Onslow, 2017, p. 10). Stuttering behaviours can be described using the three categories or the seven descriptors, according to the level of detail required. Onslow (2017) referred that the seven stuttering behaviours could occur within one stuttering moment.

The duration of the stuttering moments can also impact the severity of stuttering. Manning (2010, p. 166) concluded that “even when the speaker exhibits a relatively low frequency of stuttering, if any one of the events lasts for several seconds and is associated with considerable muscular tension, the ability of the person to communicate is severely compromised”. This behavioural measure refers to the duration of individual events of sound/syllable repetition and sound prolongation (Zebrowski, 1991, 1994). Duration combined with frequency, number and rate of repetitions per sound/syllable repetition indicate severity (Conture, 1990b; Cooper & Cooper, 1985; Starkweather, Gottwald, & Halfond, 1990). Zebrowski (1991) analysed acoustically the average duration of stuttering moments from children who stutter and compared these values with controls; it was concluded that duration of stuttering moments does not distinguish significantly children who stutter from children who do not stutter. Repetitions were measured from the onset of the acoustic energy associated with the disfluent initial sound to the end of acoustic energy for the final repetition of the repeated sound or syllable. Audible prolongations were measured from the onset of acoustic energy associated with the initial sound to the end of acoustic energy of the audible prolonged sound. Inaudible prolongations produced at the beginning of a word were measured from the offset of acoustic energy linked with the final sound in the previous word to the onset of the first fluent sound in the subsequent word (i.e., where the disfluency occurs). The use of acoustic measurement of duration increases accuracy and precision, as human reaction time does not affect the measurements (Kelly & Conture, 1988; Zebrowski & Conture, 1989).

Riley (2009) proposed the measurement of the duration of the three longest stuttering events using a stopwatch, and the use of their mean duration. Gillam, Logan and Pearson

(2009) recommended the use of software to assess duration of moments of stuttering to yield a more accurate measurement. For repetitions, they proposed the following procedure: “the onset is the point at which the repeated speech begins, and the offset is the point at which fluent speech resumes” (Gillam et al., 2009, p. 98). For stuttering moments with prolongations or blocks, “the onset is the point at which the disfluent speech sound begins, and the offset occurs at the beginning of the following sound” (Gillam et al., 2009, p. 98).

Duration of repetitions can also be measured as the number of extra productions of a segment. In a research study of preschool children, Boey et al. (2007) grouped the duration of repetitions into four categories, i.e., none, one or two repetition units, three or four repetition units and five or more repetition units. In the same study, for prolongations and blocks, the duration divided into four categories, i.e., none, less than 1 second, 1-2 seconds and 3 or more seconds.

Stuttering moments could be accompanied by associated behaviours (with different severity levels of effort and tension), that reveal reactions and coping strategies for fluency breaks (Gregory, Campbell, & Hill, 2003; Manning, 2010; Van Riper, 1982). Guitar (2014) denominated the associate behaviours as secondary behaviours. According to Guitar (2014), secondary behaviours are learned reactions that PWS use to end or avoid stuttering moments and are divided into *escape behaviours* and *avoidance behaviours*. *Escape secondary behaviours* occur during the stuttering moment, when a person tries to end stuttering and finish the word. Interjections, extra sounds, eye blink or movements of extremities are examples of escape behaviours. An *avoidance secondary behaviour* occurs before the stuttering moment, when PWS try to avoid stuttering and the associated feelings, using behaviours that were previously used to escape from the stuttering moment.

Manning (2010) referred to the avoidance behaviours as subtle surface features, since they are observable but difficult to identify. Subtle surface features are avoidance (i.e., use of avoidance behaviours to hide the fluency problem), substitution (i.e., substitution of a fear word for a non-feared word, with a minor change in the meaning of the sentence) and postponement (i.e., use of strategies, such as sounds or words, to postpone or support the initiation of a feared word).

Other authors developed typologies for secondary behaviours. Riley (2009) used a scale of tension from 0 (none) to 5 (severe and painful looking) to classify four types of secondary behaviours (the term physical concomitants was used to describe secondary behaviours): Distracting sounds, facial grimaces, head movements and movements of the extremities. Cooper and Cooper (1985) in the *Cooper Personalized Fluency Control Therapy* divided

the concomitant behaviours into the following categories: Posturing behaviours, respiratory behaviours, facial behaviours, syntactic and semantic behaviours and vocal behaviours. Shipley and MacAfee (2016) developed a list of several eyes, nose, forehead, head, lips, tongue, teeth, jaw, neck, fingers, hands, arms, legs and breathing behaviours. Boey et al. (2007) used a scale from 0 to 3 to score physical movements/tension and/or audible signs: 0 – no tension was observed; 1 – mild physical tension and represent a physical effort short in duration and subtle to observe; 2 – moderate physical tension, meaning that the tension was immediately distracting; 3 – severe physical tension, indicating that the tension observed was very distracting and produced with considerable effort.

The Test of Childhood Stuttering (TOCS) presented a few more common behaviours associated with stuttering moments in the *Observational Rating Scale*, one of the scales in the instrument (Gillam et al., 2009). Items such as "seem to become tense when called on to speak", "seem to run out of breath while completing a sentence" and "moving his or her body inappropriately when speaking (…)" report secondary behaviours. In more detail, the assessment tool presented an *Associated Behaviours Worksheet*, in which the associated behaviours are divided into 7 categories: Excessive physical tension, extraneous movements, atypical phonation, atypical respiration, atypical prosody, atypical rate of articulation and avoidance of words or participation.

Naturalness refers to the degree to which the speaker sounds similar to normal speakers of the same gender, age and dialect (Martin, Haroldson, & Triden, 1984; Riley, 2009). Martin, Haroldson, and Triden (1984) developed a 9-point scale (1 – highly natural sounding speech; 9 – highly unnatural sounding speech) with which unsophisticated listeners rated speech of nonstutterers significantly more natural than stutter free speech of PWS speaking under delayed auditory feedback conditions. The 9-point scale was used to rate the naturalness of speech, based on the judges' sense of what is natural in speech. Additional studies (Armson & Kiefte, 2008; Hargrave, Kalinowski, Stuart, Armson, & Jones, 1994; Ingham, Gow, & Costello, 1985; Onslow, Adams, & Ingham, 1992; Stuart & Kalinowski, 2004; Van Borsel & Eeckhout, 2008) have shown that judges can achieve acceptable levels of interrater reliability with the naturalness 9-point scale.

2.2.3 Speech sample collection

"Speech samples are the means by which fluency, disfluency, and associated conditions are observed and quantified for the sake of diagnosis and treatment decisions" (Yairi & Seery, 2014, p. 208). Given the variability of stuttering, ideally, three separate speech

samples are recommended (Gregory, 2003; Yairi & Seery, 2014). In the assessment of adolescents and adults, Yairi & Seery (2014) recommended spontaneous speech samples (monologue and/or dialogue), reading sample and other relevant context (e.g., phone call, group discussion or classroom). Differences can be found in monologue and dialogue speech samples; the characteristics of a dialogue (with the occurrence of turn-taking during conversation, shifting topics and interruptions) are more similar and representative sample of a daily speech, where more communicative pressure can occur. Monologue is not the most representative form of daily speaking context; however, the collection of a larger sample of speech can be achieved quickly with this sample (Yairi & Seery, 2014).

Telephone speech samples are also used frequently because they are considered by many researchers as one of the most difficult and challenging situations for PWS (e.g., Bloodstein & Bernstein Ratner, 2008; O'Brian, Onslow, Cream, & Packman, 2003). Telephone calls "represent real life speech performance, and "(…) are relatively free of clinic cues" (Langevin et al., 2006, p. 236). Sampling speech with the client using the telephone avoids the possibility of inflated improvements that may occur when the clinician is present or when the client self-records (Langevin, Kully, Teshima, Hagler, & Narasimha, 2010).

To elicit a speech sample (monologue or dialogue), the clinician prompts with a request such as "Tell me about your hobbies/vacations/interests" or through the choice of a topic from a set of topic cards or general topics (e.g., job or vocation) (e.g., Antipova, Purdy, Blakeley, & Williams, 2008; Armson & Stuart, 1998; Langevin et al., 2006). To collect a telephone speech sample, the clinician could make a surprise telephone call (e.g., Langevin et al., 2006) or ask the client to make a telephone call to an unfamiliar person (Huinck & Rietveld, 2007).

Several authors considered different motivations to collect several speech samples. Costello and Ingham (1984), Guitar (2014) and Manning (2010) referred that through the collection of different speech samples contexts it is possible to understand the variability of stuttering, in an attempt to manage it. Lincoln, Packman, Onslow, and Jones (2010) noted that different speech samples allow the assessment of the impact on stuttering severity of different motoric, linguistic and cognitive demands that arise from different speech tasks. Ward (2008) pointed that a variety of stuttering samples constitutes an advantage for the client who stutters to demonstrate his/her variability of stuttering, to increase the appropriate subjective identification of overt stuttering behaviours made by the clinician. Packman et al. (2004) mentioned that a variety of contexts increase the size of the speech samples and the face validity of measurement.

Collecting a variety of speech samples beyond clinical samples is also important, as they provide greater representation of the PWS's stuttering than within-clinic samples (Packman et al., 2004; Riley, 2009). The variety of speaking situations in which speech samples could be collected are characterised by different modalities of speech (e.g., monologue, dialogue or phone call), locations (e.g., home or clinical setting), conversation partners (e.g., family member or stranger) or communicative intents (e.g., request or chatting) (Packman et al., 2004). It is therefore more appropriate to speak in communicative context instead of speaking situations.

It is recommended to record the speech samples in video, due to inaudible stuttering behaviours and superfluous nonverbal stuttering behaviours that are not detectable on audio recordings if they do not occur with audible stuttering behaviours (Packman et al., 2004; Yairi & Seery, 2014). It is suggested to obtain an image that includes the upper body (waist to the head), at a sufficient distance to capture oral postures and facial expressions (Yairi & Seery, 2014). Rousseau, Onslow, Packman, & Jones (2008) used speech samples from preschool children who stutter and concluded that measures of %SS made with audio-only recording samples may underestimate the frequency of stuttering.

Related to speech sample size, it has been shown that larger samples in different contexts improve reliability and it is recommended that a sample should contain a sufficient number of the target behaviours (Lund & Duchan, 1993). For stuttering, the recommendation is even more important due to the fluctuation of frequency and severity within and between contexts (Sawyer & Yairi, 2006): "(…) few systematic investigations have examined the role of sample size on the disfluency data obtained, and no scientific basis for using any speech sample size has been offered" (p. 37). In research studies with children who stutter, a great variability of sample size has been used. Onlow, Costa, & Rue (1990) used speech samples with 60 seconds; Sawyer & Yairi (2006) used 1200 syllables in their research with children who stutter. Related to the sample size on AWS, Logan & Haj Tas (2007) and Roberts, Meltzer, & Wilding (2009) reported nonsignificant differences for the number of stuttering disfluencies in different sample sizes, ranging from 300 to 1800 syllables. Similarly to the procedures used with children who stutter, research studies with AWS also used different sample sizes with duration from 2-5 minutes to 10 minutes, mainly for dialogue and speech samples collected on telephone (Antipova et al., 2008; Armson & Kiefte, 2008; Armson, Kiefte, Mason, & De Croos, 2006; Huinck & Rietveld, 2007; Langevin et al., 2006, 2010; Langevin & Boberg, 1993; Menzies et al., 2008; O'Brian et al., 2003; O'Brian, Packman, & Onslow, 2008; O'Donnell, Armson, & Kiefte, 2008; Packman, Onslow, & van Doom, 1994; Roberts et al., 2009; Unger, Glückb, & Cholewaa, 2012). Other authors, such as Guitar

(2014) recommended at least 300 syllables in spontaneous speech samples and 200 syllables for reading; Shapiro (1999) suggested between 300-400 words and Riley (2009) used 200 syllables in the Stuttering Severity Instrument – Fourth Edition. Yairi & Seery (2014) referred that the speech sample size is an important question for the purpose of counting frequency of specific disfluency types, but this could be a minor question when the clinician is only interest in assigning an overall rating of stuttering severity.

2.2.4 Assessment of severity in PWS

Throughout the years, several assessment tools were developed to evaluate stuttering severity in AWS. The Sherman-Lewis Scale (Lewis & Sherman, 1951) included nine audiotape samples ranked from mild to severe. A given stuttered sample can be matched with the nine rating samples and rated accordingly. According to Riley (1994, 2009) this procedure was complex and ignored the influence of the visible manifestations related to stuttering.

The Scale for Rating Severity of Stuttering is a subjective measure, known as the Iowa Scale (Johnson et al., 1963). This scale is based on the listener's impression of a clients speech (rated from 0 to 7): No stuttering, very mild, mild, mild-moderate, moderate, moderate-severe, severe, and very severe. Each value presented a description based on frequency of stuttered words, tension, duration and associated movements. It has been shown that the Iowa Scale is reliable when a group of raters use the scale, but the reliability for use with single judges is questionable (Guitar, 2014). Van Riper (1982) revised the scale and organised the parameters in terms of frequency, tension, duration, and postponement-avoidance. Kully and Boberg (1988) used the Iowa scale limited to two or four words as a descriptor of each stuttering severity level and identified inter-clinic discrepancies. According to Yairi & Seery (2014), this scale cannot adjust the description to the client's stuttering. Additionally, the Iowa Scale considered that frequency and duration increase in the same direction, which might not be the case of a client's stuttering characteristics.

Wingate (1976) recommended, for diagnostic purposes, a 5-point scale of stuttering severity, that contemplated stuttering frequency (in percentage of stuttered words), the effort and the secondary behaviours. To each overall rating (i.e., very mild, mild, moderate, severe and very severe) corresponds a stuttering frequency (from 3% to 25%), a severity level related to effort (from no visible tension to very severe visible tension) and a level of secondary behaviours (from no noticeable movements to painfully agitated movements).

The Lidcombe Program's Severity Rating Scale was developed as a part of an intervention program based on the operant conditioning for preschool children (Onslow et al., 1990). The parents use this scale daily to rate the level of stuttering severity of the child, in a 1-10 scale (1 – no stuttering; 2 – extremely mild stuttering; 10 – extremely severe stuttering). The Lidcombe Program's Severity Rating Scale is a valid and reliable scale to obtain information concerning stuttering severity beyond the clinical environment (Eve, Onslow, Andrews, & Adams, 1995; Onslow et al., 1990; Onslow, Harrison, Jones, & Packman, 2002). After 2015, the Lidcombe Program's Severity Rating Scale used a scale from 0 (no stuttering) to 9 (extremely severe stuttering). A 9-point scale was also used by clinicians and adult clients within and beyond the clinic to measure stuttering severity in the Camperdown Program for Stuttering (O'Brian, Carey, Onslow, Packman, & Cream, 2010; O'Brian et al., 2003). The clients were trained by the clinicians to use the scale (1 –no stuttering; 2 – extremely mild stuttering; 9 – extremely severe stuttering) from the first clinic visit. The 9 point scale was shown to be a reliable and valid clinical tool for use with adults in measuring stuttering severity (O'Brian, Packman, Onslow, & O'Brian, 2004).

The Sherman-Lewis Scale (Lewis & Sherman, 1951), the Scale for Rating Severity of Stuttering (Johnson et al., 1963) and also the Lidcombe Program's Severity Rating Scale (Onslow et al., 1990) are severity rating scales, with equal interval ordinal scales (i.e., a sequence of numbers in which each scale division represents the same severity increment), where the listeners uses the scale to present an overall judgement of severity for a speech sample (Onslow, 2017). As the severity rating scales presented fewer potential scores, users can potentially achieve adequate levels of reliability with this measurement system, when compared with the percentage of stuttered syllables (Karimi, O'Brian, Onslow, & Jonesc, 2014; Onslow, 2017). Research studies with severity scales with different numbers of points (e.g., 5, 7, 9 or 15) did not reveal significant differences in mean scale values of reliability (Cullinan, Prather, & Williams, 1963; Curran & Hood, 1977). Additionally, rating scales are easy to use, do not depend of expensive equipment, require minimal training and reaching interjudge and intrajudge agreement is easier for the severity rating than for the %SS (O'Brian, Packman, Onslow, et al., 2004). O'Brian, Packman, & Onslow (2004) also reported that there are little differences in the severity rating when the points of the scale are defined, when the participants had repeated exposure to the task or feedback and when live samples were used (rather than recorded samples).

The Systematic Disfluency Analysis (SDA) is a technique developed to prepare speech samples transcriptions from audio or video (Campbell & Hill, 1987). Disfluencies are identified based on the typology *more typical* (i.e., interjections, revisions or phrase/word

repetitions) and *less typical* (i.e., sound/syllable/whole word repetitions, blocks or prolongations). Duration (number of repetition units or prolongation duration) as well as associated behaviours characteristics (e.g., increase in pitch or loudness or physical tension) are also registered. The disfluencies are scored on a weighted point system and a severity index is obtained (Campbell, Hill, & Driscoll, 1991).

The Real-Time Analysis of Speech Fluency (Yaruss, 1998b) measures the frequency of different types of disfluencies in real time, in an easy and quick way, suitable to use in assessments during intervention sessions. This procedure is not based on a transcription and, due to this fact, it cannot provide any information related to language output or speech fluency. The Real-Time Analysis of Speech Fluency is a flexible technique, allowing the clinician to choose a range of factors (e.g., unit of measurement or behaviours to be measured) and complement these with additional measures, such as duration. It uses a disfluency count sheet to register disfluencies by hand. The clinician should watch the client several minutes to become familiar with the communication style; then, the clinician codes the occurrence of each word with a “dot” or “dash” and each disfluent or stuttered word with an “x” or with the abbreviation of each type of disfluency.

The Fluency Interview Index of Severity (FIIS) is a severity rating proposed by Ryan (2001). It was based on ten speech samples (e.g., automatic, echoic, reading or conversation), from which the clinician calculated the stuttering rate (in stuttered words/minute) and the word rate (in syllables per minute). Each number of the severity scale (from 0 – no stuttering to 7 – severe stuttering) presented an equivalent stuttering rate, word rate and topography of stuttering (types of stuttering moments observed). As PWS could present different severity ratings in different measures, a number of the severity scale is attributed to each measure (i.e., stuttering rate, word rate and topography) that are added and divided to calculate the severity rating mean.

The Therapy Outcome Measure (TOM) (Enderly & John, 2015) was created to allow different professionals working in health, social care and education in the description of abilities and difficulties of clients related to impairment activity, participation and well-being/distress and in monitoring changes over time. The TOM is a reliable and valid tool. Related to the domain impairment of disfluency, the clinician needs to identify the descriptor that “best fits” the severity of stuttering observed in the client. The descriptors were presented in the following scale: 0 (severe stammer), 1 (severe/moderate stammer), 2 (moderate stammering), 3 (moderate/slight stammer), 4 (slight stammer) and 5 (no

stammer). The description of each severity rating was based on types of disfluencies, tension, duration, avoidance, associated behaviours and evidence of stuttering.

The Stuttering Severity Instrument (SSI) was developed in 1972 by Riley with the objective of determining stuttering severity for children and adults. The original SSI (Riley, 1972) was created as an instrument to monitor intervention effects in clinical practice and as a research tool for studies about stuttering. It was revised and modified in 1980, 1994 and 2009. In all different revisions of the SSI, the behavioural measures considered were frequency, duration, and physical concomitants. These three parameters were chosen based on the criteria of observability, measurability and ease of administration (Riley, 1972). The total overall score was obtained by adding the three subcomponents measured. In the first version of this instrument, frequency was measured in %WS and expressed on a task score from 2 to 18. Duration was also expressed in a task score from 1 to7 after the estimation of the length of the three longest blocks. Physical concomitants (distracting sounds, facial grimaces, head movements and extremities movement) were rated in a scale from 0 to 5 (0- none to 5 – severe and painful looking). The SSI was revised in 1980 (SSI-R) (Riley, 1980) without changes in the three parameters assessed. Al-Khaledi, Lincoln, McCabe, Packman, and Alshatti (2009) estimated interjudge and intrajudge reliability for the SSI with ten graduate students as judges. They had relatively high intrajudge agreement and poor interjudge reliability. Despite the results regarding interjudge reliability, the study partially supports the statement that the SSI differentiates among stuttering severities. Healey (1991) reviewed the SSI-R and concluded that the content validity, the low interjudge reliability and the limited number of parameters used for the assessment of stuttering severity were the weaknesses of the instrument.

Mowrer (1991) reviewed the SSI based on a list of eight characteristics of a “perfect” measuring instrument, concluding that the definition of stuttering used in SSI, which was based only on observable behaviours, lead to an inadequate representation of attitudes and feelings that may influence the communication process. The observable behaviours assessed was silent or audible prolongations and repetitions of a sound or syllable; the other types of disfluencies (e.g., word and phrase repetitions, interjections, broken words, revisions, and other verbal devices) were overlooked. Demographic information about the sample was absent and present a limited sampling of stuttering behaviour (i.e., only one or two speaking situations were represented). Mowrer (1991) also concluded that the instrument does not represent absolute scores, and that the index of reliability, intrajudge reliability or the standard error was not reported, as well as reliability and validity using adult population. The sample used in the reliability determination was also criticised, as only

students were used in the study (instead of practicing clinicians). Related to validity, the opinion of Mowrer (1991) is that the choice of the Iowa Scale of Severity of Stuttering to assess criterion validity was inappropriate because it was a poor test.

Based on the recommendations and comments from these reviews, the SSI was revised once more. The frequency count was changed in the Stuttering Severity Instrument– Third Edition (SSI-3) (Riley, 1994) to %SS, because according to Riley (1994) the use of stuttered syllables does not lead to the error that can result from two stuttering events on a single word and seems to be more objective due to the absence of linguistic interpretation (Riley, 1994). Frequency is converted to a scale score of 2-18. Duration is expressed as the average length of the three longest stuttered events and converted to a scale score of 4-18. Physical concomitants (distracting sounds, facial grimaces, head movements and movements of the extremities) are ranked on a scale from 0 to 5. The SSI-3 was revised by Lewis (1995) with twenty graduate students as judges. Lewis (1995) concluded that the SSI-3 (Riley, 1994) score is well correlated with judgments of overall severity.

The last version of the Stuttering Severity Instrument (SSI-4) retained the three behavioural measures and the scale scores included in SSI-3, as well as the normative data (Riley, 2009). Several measures and types of judgments were added: Recommendation of the collection of beyond clinical speaking samples and telephone samples, the integration of the Computerized Scoring of the Stuttering Severity Version 2 (CSSS-2.0), the addition of a speech naturalness rating scale to the Examiner Record Form and the inclusion of a scale to self- report of feelings related to stuttering: The Clinical Use of Self-Reports. The SSI-3 (Riley, 1994) and SSI-4 (Riley, 2009) were normed on a sample of 72 preschool-aged children, 139 school-aged children, and 60 adults (Riley, 2009).

All the versions of the SSI can be administered to nonreaders and readers. Frequency can be obtained through video- or audio-taped speech samples. For nonreaders, data from a speaking task (based on pictures designed to provide verbal stimulation) was collect. It was necessary to collect two clinic speaking samples. For readers, data was first obtained from a reading task of about 150-300 syllables (with reading material appropriate to a specific education level) and a speaking task (picture plates can be used for younger individuals or a conversation about an age-appropriate topic). Beyond clinic speaking samples can be obtained as a transfer measure, but the score is not related to the normative data. To determine the percentage of stuttered syllables, the number of stuttering behaviours was divided by the number of syllables. For reader, the percentages of stuttered syllables are separately computed for the reading task and the speaking task. Once the frequency score

is computed, it can be converted to a task score. Repetitions and prolongations of sounds or syllables (considering also silent prolongations) were considered as stuttering behaviours ("types of disfluencies that are abnormal in children and AWS but are found frequently among people who stutter" (Riley, 2009, p. 4). Rephrasing, repetition of whole words or sentences, and pausing without tension were not considered stuttering behaviours. Repetition of monosyllabic words are considered as a stuttering behaviour if the words sounds shortened, prolonged, in staccato, with tension, i.e., "abnormal" (Riley, 1994). To compute the duration score the three longest events are added and divided by three to obtain the mean duration. Once the mean duration was calculated, it can be converted to a scale score. The assessment of the physical concomitants was based on observations of the totality of the speaking samples. The auditory and visual concomitants associated with stuttering events (i.e., distracting sounds, facial grimaces, head movements and movements of the extremities) should be scored on a scale from 0 to 5. The total score is computed by adding the scores for frequency, duration, and physical concomitants (scores can be between 0 and 56). The value of the total score can be converted to a percentile rank or a severity equivalent.

The SSI-4 is an instrument easy to use and only measures of stuttering severity with standardised procedures for gathering and scoring speech samples in three dimensions (i.e., frequency, duration and physical concomitants) and for three different samples (i.e., preschool-age children, school-age children and adults). The numerical data obtained with the SSI-4 were considered "only a partial description of severity and that any diagnostic conclusions are the product of careful consideration of all available information" (Riley, 1994, p. 1). Other important diagnostic inputs included interviews to PWS, relatives, teachers, co-workers, assessment of speech and language, self-assessment of affective, behavioural, and cognitive reactions to stuttering, audiology assessments and medical or psychological assessments (Riley, 2009; Riley et al., 2004; Woolf, 1967; Wright & Ayre, 2000; Yaruss & Quesal, 2006).

Self-measurements of stuttering severity to quantify stuttering and concomitant speech behaviours are referred as important to reporting speech performance and to achieve and maintain treatment benefits (O'Brian, Packman, & Onslow, 2004). Results obtained by Naylor (1953) and Aron (1967) with severity self-rating scales from 1 to 9 suggest, in agreement with the literature related to severity rating scales use, that these are useful to "supplement traditional stutter-count measures, particularly for reporting on stuttering severity outside the clinic in the person's everyday speaking environment "(…) providing valuable information about the generalization of treatment effects in many different

situations" (O'Brian, Packman, & Onslow, 2004, p. 220). A study by O'Brian et al. (2004) with ten stuttering adults using a 9-point severity self-rating scale concluded that there were acceptable levels of agreement in the measurement made by the participants and the speech and language therapists (SLTs) and also between the initial measure and the severity assessment made six months later. O'Brian et al. (2004) also concluded that the use of this scale as a clinical measurement procedure is supported by the findings.

The TOCS (Gillam et al., 2009) is an assessment tool to be used in children between 4 and 12 years of age. TOCS encompasses three components: The standardised Speech Fluency Measure with four speech tasks (i.e., rapid picture naming, modelled sentences, structured conversation and narration), Observational Rating Scales (i.e., speech fluency rating scale and disfluency-related consequences rating scale) and Supplemental Clinical Assessment Activities (i.e., clinical interviews, comprehensive analysis of disfluency, speech rate analysis, disfluency duration analysis, repetition unit analysis, associated behaviours analysis, stuttering frequency analysis and speech naturalness analysis). The Speech Fluency Measures and Observational Rating Scales are norm-referenced and present index scores, percentiles and descriptive terms to characterise speech performance, diagnose stuttering and estimate severity. The Supplemental Clinical Assessment Activities can be useful if the clinical suspected of a false negative/false positive, or if he/she consider that the norm-referenced part of the TOCS underestimate or overestimate the child's severity of stuttering. The TOCS has content-description validity, criterion-related validity and construct validity. Related to reliability, the TOCS has internal consistency, test-retest and interscorer reliability guaranteed.

The instruments presented above allow the assessment of the severity of observable behaviours. Simple severity scales (Lidcombe Rating Scale and the Sherman-Lewis Scale) can be utilized as a global assessment, that are easy to implement and can be used by both clinicians and clients on a daily-basis evaluation. However, those scales do not capture the specific behavioural characteristics of the stuttering moments. The Real-time Analysis of Speech Fluency (Yaruss, 1998b) and the Fluency Interview Index of Severity (Ryan, 2001) are severity assessment tools designed to assess only a specific behavioural measure, i.e., frequency and rate of speech, respectively. The Iowa Scale (Johnson et al., 1963), the Severity Rating Scale (Wingate, 1976) and the SDA (Campbell & Hill, 1987) are similar scales, that assessed frequency in %SW and duration/secondary behaviours with qualitative descriptors; %SW, despite of being an easy-to-obtain behavioural measure, could not capture instances when a person stutters more than one time in a multisyllabic word. Additionally, duration is a quantitative measure that should be assessed in a

continuous metric scale. Concerning the disfluency component of the TOM scale, the limitation is similar: Behavioural measures are only described qualitatively. Despite the usability of the scales described (i.e., The Iowa Scale, the Severity Rating Scale, the SDA and TOM), the qualitative descriptors bring subjectivity to the process of severity determination.

The SSI-4 (Riley, 2009) is the only assessment tool that includes specific procedures to calculate frequency (in %SS), duration and physical concomitants. Despite the criticism regarding the reliability and validity process, SSI-4 is the most used instrument to determine the severity of stuttering, because it has standardised procedures and validation data for children and adults; additionally, SSI-4 also has flexibility in the assessment procedures to allow the use of the instrument in a variety of environments (from clinic to a research laboratory) (Howell, Soukup-Ascencao, Davis, & Rusbridge, 2011). However, several weaknesses can be identified in this instrument: Absence of demographic data descriptors; normative data calculated based on a restricted number of speech samples; imprecise procedure to measure duration (i.e., only based on perceptual assessment); absence of procedures to qualitatively classify stuttering events; different severity assessment procedures (i.e., digital versus live assessment) allowed in the SSI-4(Howell et al., 2011). The CSSS-2.0 (added to the SSI-4) contributes to the automatic calculation of frequency and duration, but retains the use of only perceptual information to the analysis (instead of the use of acoustic information to improve measurement accuracy and precision). Other weakness could also be pointed to CSSS-2.0: Absence of reliability, validity and accuracy determination (Howell et al., 2011).

The present research intends to create a severity assessment instrument that overcomes the weaknesses identified in different instruments. This new instrument for AWS intends to be carefully validated with different speech samples, to include accurate procedures to analyse all behavioural measures with precision (e.g., use of acoustic information) and to use an automatic method to calculate severity of stuttering events based on a previous analysis made on an annotation software.

2.2.5 Aims of the current study

This study describes the development of an assessment instrument to calculate severity of observable stuttering behaviours based on four video recorded speech samples (reading, monologue, dialogue and telephone speech) using the EUDICO Linguistic Annotator (ELAN). The development and the study of the content validity of the Severity Assessment

Based on Events of Stuttering (*SABES*), as well as the assessment of criterion-related validity and construct validity are reported and discussed in this chapter.

2.3. Methodology

2.3.1 Content validity analysis

Content validity represents the content relevance of the items of an instrument. It was determined by the application of a two-stage process (Lynn, 1986): 1) developmental; 2) judgment-quantification. Lynn (1986) claims that the two-stage process is essential for the validation process of an instrument. The developmental stage (initial stage) was comprised of three steps: Identification of full content domains, sampling and item generation and assimilation of items into a useable scale. The initial step in the developmental stage was the domain identification and consisted of the categorisation and description of the dimensions to be assessed and on a literature review of the measure (Lynn, 1986).

After an extensive literature review related to severity assessment of stuttering behaviours on ERIC, MEDLINE, PubMed and B-on and also on specific literature (Bloodstein & Bernstein Ratner, 2008; Guitar, 2014; Manning, 2010; Shapiro, 2011) the procedures to assess observable behaviours (i.e., frequency, duration, associated behaviours, effort/tension, naturalness, rate of speech and types of stuttering) were recognised and identify as the content domain. During the initial step, nine instruments (Campbell & Hill, 1987; Campbell et al., 1991; Enderly & John, 2015; Gillam et al., 2009; Johnson et al., 1963; D. Lewis & Sherman, 1951; Riley, 2009; Ryan, 2001; Wingate, 1976; Yaruss, 1998b) that assessed at least one speech behaviour were identified. Tables of contents were produced to assess the instruments, and the scales related to the behavioural measures assessed and the procedures used were reviewed.

The second step in the first stage involved the generation of procedures for the domains used in initial step to assess the assessment instruments. On the second step, a first version of the *SABES* was produced based on the instruments and research procedures reviewed. The third step of the developmental stage consisted in refining the procedures generated on second step into a useable form. On this step, a manual of procedures using ELAN to analyse stuttering moments and the specific procedures to collect speech samples were developed.

A second stage of the process was implemented to determine and ensure content validity using judgment-quantification. This stage comprised two steps: The first consisted of

judging the *SABES* procedures and the second judging the entire instrument, both by an expert panel. Lynn (1986) proposes three as the minimum number of experts that should incorporate the panel. The experts selected should have expertise in the domain studied; it should be given to the judges information about the full dimensions of the variables studied; a specific instrument should be used to assess relevant content of each item (Grant & Davis, 1997).

The *SABES* procedures, as well as the procedures to assess moments of stuttering in ELAN, were presented and submitted to judgment by a panel of experts (second stage – judgment-quantification stage). The author of this Thesis presented the content domain, the general goals, the conceptual basis, the procedures to assess stuttering moments from different speech samples (in terms of frequency, duration, associated behaviours, tension degree and naturalness) and the use of ELAN to the expert panel. The objectives of the expert meeting were: To analyse the *SABES*; to rate the *SABES* items using a visual analogue scale (VAS) (Crichton, 2001; Hasson & Arnetz, 2005); to add or remove words or sentences from the *SABES* manual, in order to achieve the relevant content (Perry et al., 2004). The three experts involved (with experience in stuttering and instrument development) assessed the items in terms of relevance, clarity, simplicity and accuracy, using an instrument review questionnaire with 28 questions. Using the thinking aloud method (Goldman, 1971), the opinions of the expert panel related to the *SABES* relevance, clarity, simplicity and accuracy were recorded while they filled in a questionnaire. The Bland and Altman (1986) modified approach (Jesus, Valente, & Hall, 2015, pp. 4–5) was used to analyse the agreement/disagreement between the three experts on each question of the instrument review questionnaire. The pre-final version of the *SABES* was developed after this phase.

The pre-final version of the *SABES* included an introduction, administration procedures (with print screens to explain the use of ELAN in the assessment of stuttering moments), speech sample collection instructions specific to reading, monologue, dialogue and telephone samples and two appendices: The materials used to collect speech samples (i.e., the “North Wind and the Sun” fable and the spontaneous speech samples cards) and the characterisation of the stuttering moments.

2.3.2 Pilot study

A pilot study was conducted using the pre-final version, to examine the feasibility of *SABES* to be used with a larger group (Leon, Davis, & Kraemer, 2011). All ethical procedures were

ensured (appendix 1) and informed consent (appendix 2) was collected from all participants prior to any data collection. Six AWS (mean age 26.8, range from 21 to 35 years old) were recruited through local SLTs with experience in stuttering intervention based on the following inclusion criteria: Self-reports of onset, development and chronicity of stuttering; classification as a PWS by two SLPs that agreed on the diagnosis; to be able to read. Four speech samples were collected from each participant. Frequency (in %SS and in %SW), mean duration of stuttering moments, classification of stuttering moments using the LBDL taxonomy, the number of associated behaviours and the mean effort/tension degree were obtained for each speech sample. Descriptive statistics were used to analyse the data. Scoring spreadsheets were developed in *Excel 2013*, based on pilot study measurements.

After the completion of the pilot study, the final version of *SABES* (manual and procedures) was developed.

2.3.3 Assessment of reliability

To assess *SABES*'s reliability, internal consistency was calculated, to ensure that the items are homogenous and measure the same trait (Polit & Beck, 2012). The internal consistency of *SABES* was analysed for each speech sample collected through the use of the Cronbach alpha. Cronbach's alpha values higher than 0.70 indicate that the items reliably measure the same attribute (Field, 2013; Kline, 2004; Polit & Beck, 2012).

2.3.4 Criterion and construct validity analysis

A sample size of 23 AWS (AWS) was deemed necessary, based on power calculations performed with GPower v 3.1.9.2 for a correlation test (Bivariate normal model) with $\alpha=0.05$ and 80% power. The 23 AWS were recruited through Portuguese Stutterers Association (Associação Portuguesa de Gagos) and through local SLTs with experience in stuttering intervention. Ethical procedures were guaranteed and informed consent was collected from all AWS prior to data collection. The AWS were recruited based on the following inclusion criteria: Self-reports of onset, development and chronicity of stuttering; classification as a PWS by two SLPs that agreed on the diagnosis; to be able to read; consented audio and video recordings. Four speech samples were collected from each adult who stutters. Frequency (in %SS), mean duration of stuttering moments, classification of stuttering moments using the LBDL, number of associated behaviours and the mean tension degree were assessed for each speech sample. Specific procedures will be referred to in the results

section, as an outcome of the content validity process. Behavioural measures were calculated using purpose built *Excel 2013* spreadsheets, developed as a result of the pilot study (referred in the results section).

Criterion-related validity refers to the "extent to which one measure estimates or predicts the values of another measure" (Salkind, 2007, p. 200). Criterion validity analysis implies that a raw score of a test/measurement should be compared with another measure related to what the test is designed to predict. In this study, the *SABES* total scores were correlated with the percentage of stuttered syllables (frequency) in the speech sample analysis. The correlation should be large to guarantee criterion validity (Gillam et al., 2009; Riley, 2009).

Construct validity concerns the extent to which the test scores reflect the theory behind the construct being measured (Polit & Beck, 2012). Based on the procedures implemented by Gillam et al. (2009) on the validity process of Test of Childhood Stuttering, a three-step procedure was used to analyse construct validity: Identification of a construct that contributes to the test performance; generation of hypothesis (*hypothesised relationships method*) based on the constructs developed; verification of the hypothesis based on empirical or statistical methods. The underlying construct is: "all the parameters of the *SABES* should be related to each other because they are parts of the overall construct of severity".

The task score (i.e., the conversion scores of frequency, duration, number of associated behaviours and mean tension degree of associated behaviours into a similar scale) was developed based on the results of the validity process.

2.3.5 Statistical procedures

The IBM Statistical Package for the Social Sciences (SPSS) version 23 (IBM, Armonk, USA) was used to calculate descriptive statistics (on data from the pilot study and the validity study), internal consistency and the Spearman's rank correlation coefficient were used to assess the correlation between behavioural measures results of the *SABES*.

2.4. Results

2.4.1. Content validity process

2.4.1.1 Stage I

During the first stage of the content validity analysis process, scales and instruments used to assess stuttering severity were identified and reviewed. Through the construction of two table of contents, domains and procedures to assess stuttering severity were identified (see Table 2-1 and Table 2-2). Table 2-1 presents the type of observable speech behaviours assessed by scales and instruments. Table 2-2 presents the specific procedures used by the instruments and scales to assess severity.

Table 2-1. Observable speech behaviours assessed by nine instruments. “x” indicates that the instrument assessed the behavioural measure and “-----” indicates that the behavioural measures are not assessed.

Instrument	Frequency	Duration	Associate Behaviours	Effort/tension	Naturalness	Rate of speech	Types of stuttering
Sherman-Lewis Scale (Lewis & Sherman, 1951)	x	-----	-----	-----	-----	-----	-----
Scale for Rating Severity of Stuttering (Johnson et al., 1963)	x	x	x	x	-----	-----	-----
Severity Rating Scale (Wingate, 1976)	x	-----	x	x	-----	-----	-----
Systematic Disfluency Analysis (SDA) (Campbell & Hill, 1987; Campbell, Hill & Driscoll, 1991)	x	-----	-----	x	-----	-----	-----
The Real-Time Analysis of Speech Fluency (Yaruss, 1998b)	x	-----	-----	-----	-----	-----	x
Fluency Interview Index of Severity (Ryan, 2001)	-----	-----	-----	-----	-----	x	x
Therapy Outcome Measures (Enderly & John, 2015)	-----	x	x	x	-----	-----	x
Stuttering Severity Instrument-4 (Riley, 2009)	x	x	x	x	x	-----	-----
Test of Childhood Stuttering (Gillam et al., 2009)	x	x	x	x	x	x	x

Table 2-2. Procedures used by nine instruments to assess observable behaviours. "----" indicates that the behavioural measures are not assessed by the severity instrument.

Instrument	Frequency	Duration	Associate Behaviours	Tension	Naturalness	Rate	Types of stuttering
Sherman-Lewis Scale (Lewis & Sherman, 1951)	Matching with 9 audio speech samples	-----	-----	-----	-----	-----	-----
Scale for Rating Severity of Stuttering (Johnson et al., 1963)	Percentage of stuttered words	Time (in seconds)	Qualitative descriptions associated with a specific level (e.g., a few distracting sounds and facial grimaces for level 5)	Qualitative descriptions associated with a specific level (e.g., very little relevant tension for level 1)	-----	-----	-----
Severity Rating Scale (Wingate, 1976)	Percentage of stuttered words	-----	Qualitative descriptions associated with a specific level (e.g., noticeable movements for moderate level)	Qualitative descriptions associated with a specific level (e.g., moderate visible tension for moderate level)	-----	-----	-----
Systematic Disfluency Analysis (Campbell & Hill, 1987; Campbell et al., 1991)	Percentage of less typical type of disfluencies and more typical type of disfluencies	-----	-----	Qualitative descriptions associated with a specific level (e.g., signs of visible and audible tension for mild level)	-----	-----	-----

Table 2-2 (cont.). Procedures used by nine instruments to assess observable behaviours. "---" indicates that the behavioural measures is not assessed by the severity instrument.

Instrument	Frequency	Duration	Associate Behaviours	Tension	Naturalness	Rate	Types of stuttering
The Real-Time Analysis of Speech Fluency (Yaruss, 1998b)	Using the Vanderbilt Disfluency count sheet (Conture, 1990b), percentage of stuttered words are used. Percentage of within-word disfluencies and between-word disfluencies.	-----	-----	-----	-----	-----	In the Vanderbilt Disfluency count sheet (Conture, 1990b), within-word disfluencies and between-word disfluencies are assessed (Conture, 1990b; Yaruss, 1997) In the Northwestern Form "more typical" and "less typical (Gregory, 1986; Yaruss, 1997) disfluencies are assessed.
Fluency Interview Index of Severity (Ryan, 2001)	-----	-----	-----	-----	-----	Stuttered words/minute Word rate (in syllables per minute	Different type of stuttering associated with each severity level (e.g., single part-word repetitions for level 2)

Table 2-2 (cont.). Procedures used by nine instruments to assess observable behaviours. "----" indicates that the behavioural measures are not assessed by the severity instrument.

Instrument	Frequency	Duration	Associate Behaviours	Tension	Naturalness	Rate	Types of stuttering
Therapy Outcome Measures(Enderly & John, 2015)	-----	Qualitative descriptions associated with a specific level (e.g., long prolongations for moderate level)	Qualitative descriptions associated with a specific level (e.g., occasional associated gestures and behaviours for moderate/severe level)	Qualitative descriptions associated with a specific level (e.g., slight tension for moderate/slight level)	-----	-----	Qualitative descriptions associated with a specific level (e.g., many repetitions, long prolongations, marked and repeated blocks for severe level)
Stuttering Severity Instrument-4 (Riley, 2009)	Percentage of stuttered syllables	Mean of the three longest stuttered moments (in seconds)	5 types: distracting sounds, facial grimaces, head movements and movements of the extremities	Scale: 0 (none), 1 (not noticeable unless looking for it), 2 (barely noticeable to casual observer), 3 (distracting), 4 (very distracting) and 5 (severe and painful looking)	Scale from 1 (highly natural sounding speech) to 9 (highly unnatural sounding speech)	-----	-----

Table 2-2 (cont.). Procedures used by nine instruments to assess observable behaviours. "----" indicates that the behavioural measures is not assessed by the severity instrument.

Instrument	Frequency	Duration	Associate Behaviours	Tension	Naturalness	Rate	Types of stuttering
Test of Childhood Stuttering (Gillam et al., 2009)	Percentage of syllable stuttered Number of disfluent segments with repetitions, prolongations and blocks per 100 words Number of disfluent segments with intervisions and revisions per 100 words Number of disfluent segments per 100 words Percentage of disfluent segments with repetitions, prolongations and blocks Percentage of disfluent segments with intervisions and revisions	Average duration in seconds using a stopwatch or speech analysis software (in seconds) Average number of units per repetition (duration is assessed on three different speech tasks – structured conversation, narration and other)	5 categories: excessive physical tension, extraneous movements, atypical phonation, atypical respiration, atypical prosody and atypical rate of articulation in a scale from 0 (never), 1 (rarely), 2 (sometimes) and 3 (often) (associated behaviours are assessed on three different speech tasks – structured conversation, narration and other) Questions on the disfluency-related consequences rating scale on a scale from 0 (never), 1 (rarely), 2 (sometimes) and 3 (often)	Questions on the disfluency-related consequences rating scale on a scale from 0 (never), 1 (rarely), 2 (sometimes) and 3 (often)	Scale from 1-9 (naturalness is assessed on three different speech tasks – structured conversation, narration and other)	Words per minute Words per second Syllables per minute Syllables per second (rate of speech is assessed on three different speech tasks – structured conversation, narration and other)	Presence (1) or absence (0) of repetitions and prolongations/blocks on four speech tasks (rapid picture naming, modelled sentences, structured conversation and narration)

Results shown in Table 2-1 revealed that the instruments assessed between one and seven types of speech behaviours. Data presented in Table 2-2 shows that, in terms of frequency, the majority of the instruments assessed the behavioural measure in %SW and that the most recent instruments (i.e., SSI-4 and TOCS) assessed this in terms of %SS. Related to duration, the instruments used the amount of time or number of repetition units. The associate behaviours were analysed in three instruments (Campbell & Hill, 1987; Campbell et al., 1991; Johnson et al., 1963; Wingate, 1976) using qualitative descriptors, and three instruments (Cooper & Cooper, 1985; Gillam et al., 2009; Riley, 2009) used a list type divided in five categories. The TOCS (Gillam et al., 2009) and SSI-4 (Riley, 2009) were the

only two instruments that used a scale to characterise effort and tension; the instruments that assessed tension used specific qualitative descriptors for each severity level. The TOCS and SSI-4 were also the only instruments that assessed naturalness, using Martin, Haroldson and Triden's (1984) 1-9 scale. Concerning types of disfluencies, the instruments used very different classifications: Within words/between words, more typical/less typical, specific qualitative descriptors associated with a severity level or presence/absence of repetitions and prolongations/blocks.

Concerning the corpus used to assess severity of stuttering, the instruments used different audio speech samples. Three instruments (Cooper & Cooper, 1985; Gillam et al., 2009; Riley, 2009) proposed specific speech samples for the assessment of stuttering moments. Cooper and Cooper (1985) used six different speech samples: Reciting a poem, responding to questions, repeating sentences, reading, naming pictures and spontaneous speech sample elicited through pictures; the SSI-4 (Riley, 2009) proposed the use of a spontaneous speech sample (collected through picture description or conversation) and reading (with an appropriate educational level text); the TOCS analyses data of four speech production tasks (rapid picture naming, modelled sentences, structured conversation and narration).

The first version of the *SABES* was based on the results shown in Table 2.1 and Table 2.2. Methods to assess observable speech behaviours were developed for the *SABES*: Stuttering moments are assessed in terms of frequency of words and stuttered syllables, duration in mean time duration, associated behaviours and types of stuttering moments using a categorisation list (the LBDL taxonomy), effort/tension in a severity scale (from 0-3, similar to the scale used by Boey et al, 2007) and naturalness using the Martin's et al. (1984) scale from 1 to 9. As important information to the reliable assessment of stuttering moments could be obtained through video recordings (e.g., type of associated behaviours or severity of tension/effort), the ELAN, a tool that allows the creation, edition, search and visualisation of annotations of video and audio, was used to analyse the recordings. The ELAN is a free software, developed by the Max Planck Institute for Psycholinguistics (Nijmegen, The Netherlands). According to the user guide of ELAN, the tool is "specifically designed for the analysis of language, sign language, and gesture, but it can be used by everybody who works with media corpora, i.e., with video and/or audio data, for purposes of annotation, analysis and documentation" (Tacchetti, 2013, p. iv). Specific procedures were established to analyse four speech samples (reading, monologue, dialogue and telephone speech sample) using ELAN.

A reading speech sample was collected using the text “The North Wind and the Sun”, a fable from Esopo. “The North Wind and the Sun” that has been shown to be phonetically balanced (Jesus et al., 2015).

To elicit spontaneous speech samples (i.e., monologue, dialogue and speech on the telephone), a set of speech cards were developed. A total of twelve cards with different themes (e.g., vacations, leisure time or favourite books) and questions related, are given to the person to allow the collection of a speech sample of 2-4 minutes (between 300 and 400 syllables, as recommended by Guitar, 2014). The person should seat on a chair and a camcorder should be placed on a tripod located 85 cm from the ground and 1m in front of the chair.

Before recording of spontaneous speech samples, the person chooses two cards, according to his/her preferences. One card is used to obtain a monologue and the other to elicit dialogue with the Speech and Language Therapist (SLT). Before recording the monologue, the person can organise the ideas about the theme chosen during 1 minute. To collect a speech sample on the telephone, a dialogue with a minimum duration of 2 minutes should be recorded using the second card chosen. The AWS stays in the room where the assessment is taking place, in front of the camcorder and the SLT establishes telephone contact, from another room, with the AWS.

The speech samples were analysed in terms of six observable behaviours: Frequency, duration, types of stuttering moments, associated behaviours, effort/tension and naturalness. Frequency was analysed in terms of the percentage of syllables stuttered; the total number of syllables produced by the PWS was measured with FreP version 3.2.2.6 (Martins, Vigário, & Frota, 2009), using an orthographic transcription of the total speech sample produced (for monologue, dialogue and spontaneous speech sample on the telephone). Mean duration was obtained using the acoustic waveform (available on the ELAN) and the procedures described for the TOCS instrument (Gillam et al., 2009); associated behaviours and types of stuttering moments were classified using the LBDL descriptors. Effort/tension was classified in a scale from 0 to 3, previously used by Boey et al. (2007). Each stuttering event identified was select by dragging the mouse (using information contained on the waveform associated with the video, following the recommendations by Gillam et al. 2009); duration was automatically calculated and the stuttering event was classified in terms of type of stuttering moment (using the LBDL descriptors), type of associated behaviour (using the LBDL descriptors) and effort/tension degree. A step-by-step manual (with a theoretical background, procedures, specific

instructions and print screens from ELAN (see Figure 2-1) and the material to collect speech samples was developed to guide the clinician in the analyses of behavioural measures using ELAN.

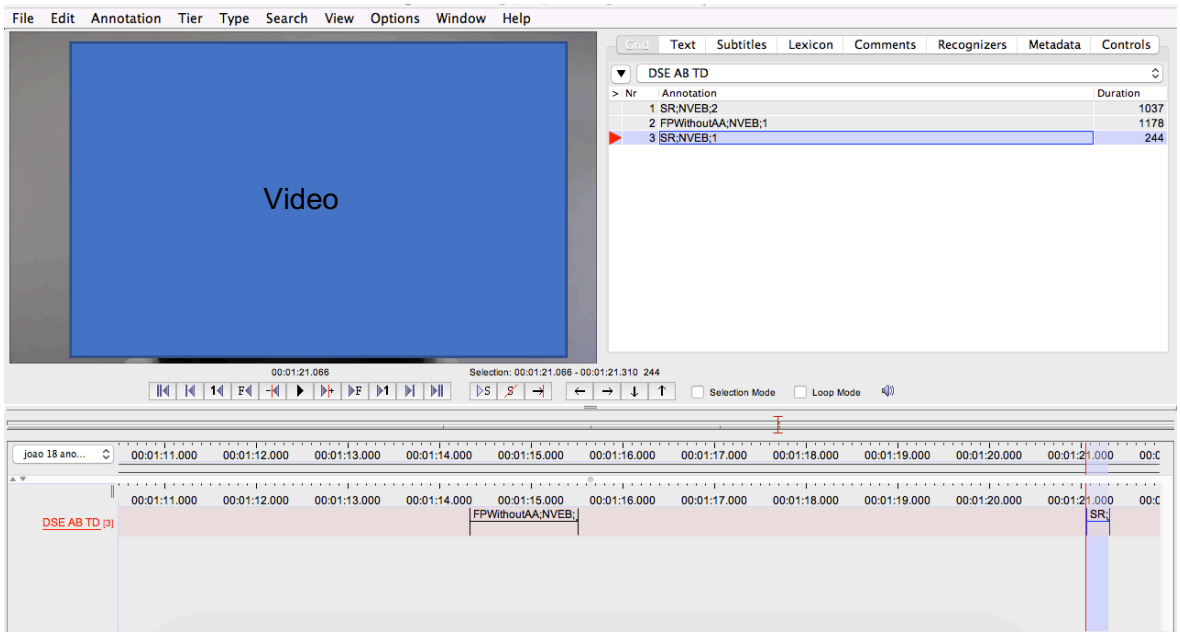


Figure 2-1. Example of a stuttering event analysis with ELAN (DSE – descriptor stuttering event; AB – associated behaviours; TD – tension degree).

2.4.1.2 Stage II

On the second stage of the content validity analysis, the manual was presented to the three experts and the results of the thinking aloud methodology implemented revealed that procedures for the analysis of speech samples with the ELAN program presented lack of clarity and simplicity. Additionally, the experts proposed the creation of a summary table with all the acronyms used on the ELAN program, to facilitate the application of the *SABES*. The modified Bland and Altman approach (Jesus et al., 2015, pp. 4–5), results shown in Figure 2.2, revealed that the experts disagreed in three questions regarding clarity, simplicity and accuracy of the speech sample collection instructions. The instructions were then modified and procedures clarified based on these results.

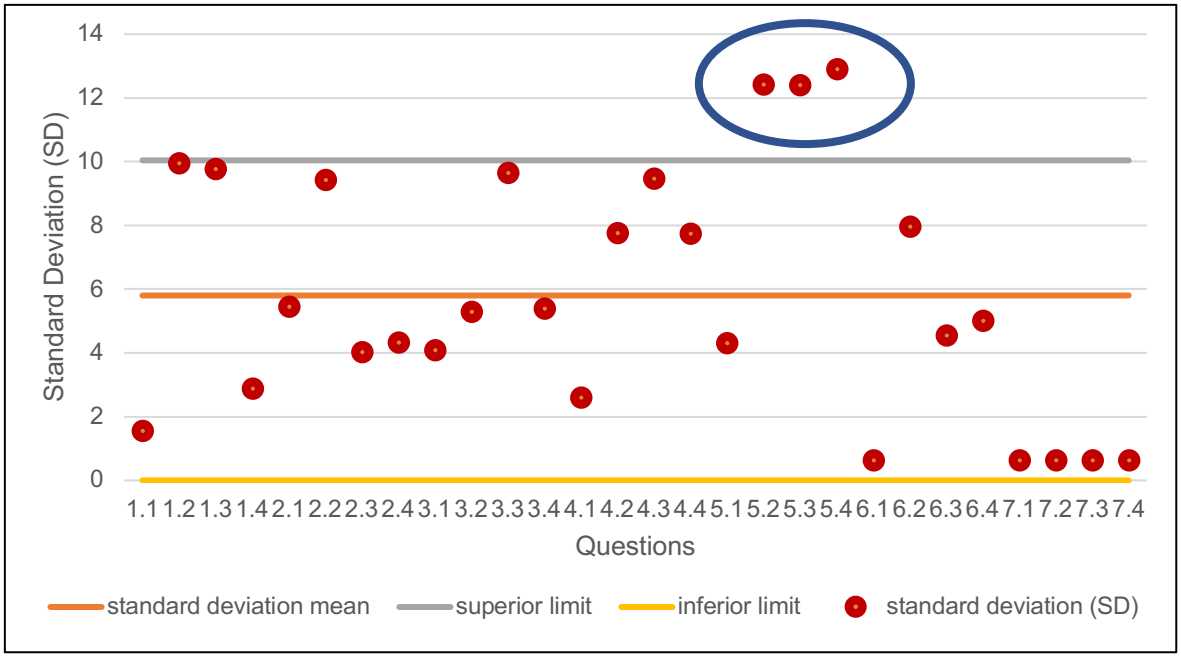


Figure 2-2 The Modified Bland and Altman analysis (Jesus et al., 2015, pp. 4–5) results. Disagreements between experts are signalled with an ellipsis.

2.4.2. Pilot study

Results from the pilot study were used to reformulate the *SABES*, to create the final version of the manual and to develop an *Excel 2013* score sheet. Tables 2-3, 2-4, 2-5 and 2-6 show the quantitative analysis results (i.e., %SW, %SS, mean duration of stuttering moments, number of associated behaviours and mean of tension degrees) obtained from 24 speech samples collected during the pilot study.

Table 2-3. Pilot study reading task results.

AWS	%SW (%)	%SS (%)	Duration mean (s)	Mean number of associated behaviours	Mean tension degree (0-3)	Naturalness (1-9)
AC	1.0	0.5	0.5	0	0.0	8
GM	2.0	1.0	4.0	0	0.0	9
MV	0.0	0.0	0.0	0	0.0	9
RF	11.2	5.6	0.3	3	0.5	7
RN	0.0	0.0	0.0	0	0.0	9
FP	6.1	4.1	1.4	3	0.7	6
Mean	3.4	1.9	1.0	1	0.2	8

Table 2-4. Pilot study monologue task results.

AWS	%SW (%)	%SS (%)	Duration mean (s)	Mean number of associated behaviours	Mean tension degree (0-3)	Naturalness (1-9)
AC	25.1	16.2	0.6	60	1.4	4
GM	12.3	7.5	0.7	5	0.0	6
MV	0.6	0.6	1.0	0	0.0	9
RF	20.1	14.4	0.7	29	0.5	4
RN	2.9	1.9	0.8	2	0.3	7
FP	49.0	46.9	1.1	53	0.9	4
Mean	18.3	14.6	0.8	24.8	0.5	6

Table 2-5. Pilot study dialogue task results.

AWS	%SW (%)	%SS (%)	Duration mean (s)	Mean number of associated behaviours	Mean tension degree (0-3)	Naturalness (1-9)
AC	22.0	9.0	0.5	47	1.3	4
GM	9.1	5.6	0.7	5	0.3	6
MV	0.0	0.0	0.0	0	0.0	9
RF	16.3	11.2	0.6	22	1.4	5
RN	0.2	0.2	0.5	0	0.0	9
FP	56.1	41.1	1.4	50	0.8	4
Mean	17.3	11.2	0.6	20.7	0.6	6

Table 2-6. Pilot study telephone sample task results.

AWS	%SW (%)	%SS (%)	Duration mean (s)	Mean number of associated behaviours	Mean tension degree (0-3)	Naturalness (1-9)
AC	22.0	15.2	0.7	58	1.5	4
GM	9.1	5.6	0.9	5	0.2	7
MV	0.0	0.0	0.0	0	0.0	9
RF[1]	-	-	-	-	-	-
RN	1.7	1.8	0.8	9	1.2	8
FP	56.1	51.2	1.	15	0.3	4
Mean	17.8	14.8	0.8	17.4	0.6	6

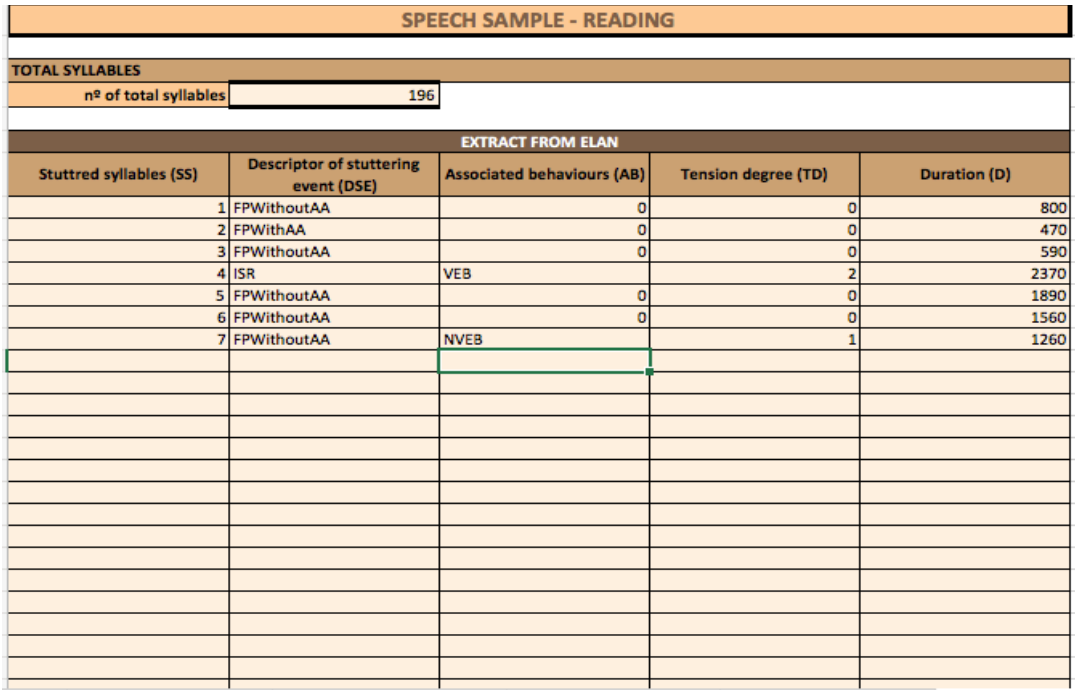
[1] The participant refuses to be recorded.

Results obtained revealed that, in terms of frequency, the telephone speech task presented the highest score, followed by the monologue task, the dialogue task and, lastly, the reading task. Duration presented higher values on the monologue and telephone tasks. Concerning the number of associated behaviours, the monologue task presented higher mean values. For the mean tension degree, the dialogue and the telephone task presented higher values. Concerning naturalness, the reading sample presented the more natural sounding speech.

The description of stuttering moments using the LBDL taxonomy revealed that the reading sample presented fixed postures with audible airflow with the highest mean percentage (24.0%); the monologue speech sample presented multisyllable unit repetition as the stuttering type with the highest mean percentage (29.0%). For dialogue and speech sample on the telephone the highest mean percentage of stuttering types were incomplete syllable repetition (25.5% and 21.2%, respectively). The classification with the taxonomy for associated behaviours revealed that nonverbal extraneous behaviours presented the mean highest percentage (33.3%, 69.9%, 61.6% and 67.6%) for reading, monologue, dialogue and speech sample on the telephone, respectively).

The purpose built *Excel 2013* file used to analyse the results contained 11 score sheets, in which is possible to calculate the stuttering severity index. Spreadsheets 1-4 allow the determination of %SW in the speech samples. Spreadsheets 5-8 allow the determination of %SS, percentage of types of stuttering events, mean duration, percentage of each type of associated behaviours and naturalness. The %SS, percentage of types of stuttering events, mean duration, percentage of each type of associated behaviours are exported automatically from the analysis made on the ELAN. Spreadsheet 9 consists of a global

profile and sheet 10 and 11 present normative data. Figure 2-3 presents an example of an Excel spreadsheet built to analyse the details of a stuttering event extracted from the ELAN.



SPEECH SAMPLE - READING

TOTAL SYLLABLES	
nº of total syllables	196

EXTRACT FROM ELAN				
Stuttred syllables (SS)	Descriptor of stuttering event (DSE)	Associated behaviours (AB)	Tension degree (TD)	Duration (D)
1	FPWithoutAA	0	0	800
2	FPWithAA	0	0	470
3	FPWithoutAA	0	0	590
4	ISR	VEB	2	2370
5	FPWithoutAA	0	0	1890
6	FPWithoutAA	0	0	1560
7	FPWithoutAA	NVEB	1	1260

Figure 2-3. Part of spreadsheets 5-8, with the detailed analysis of a stuttering event extracted from the ELAN.

2.4.3. Reliability analysis

Data obtained for the validity process were used to calculate the internal consistency. Cronbach's alpha was 0.93 for reading sample, 0.75 for monologue, 0.87 for dialogue and 0.88 for telephone sample.

2.4.4. Validation process

During the validation process, a total of 92 speech samples were collected from 23 AWS. Mean age was 29 (range from 18 to 46 years old) and sex ratio of 5 women to 18 men. Thirty five percent of the sample were from Lisbon district, 39% completed high school, 48% were employed and 74% were single. Tables 2-7, 2-8, 2-9 and 2-10 present values of the behavioural measures (i.e., %SW, %SS, mean duration of stuttering moments, number of associated behaviours and mean tension degree) for each speech sample collected. The percentage of stuttered words (%SW were not used to calculate *SABES*'s total score;

however, the behavioural measure were maintained on *SABES* due to the ease of counting of words.

Table 2-7. Validation study results of the reading task.

AWS	%SW (%)	%SS (%)	Duration mean (s)	Mean number of associated behaviours	Mean tension degree (0-3)	Naturalness (1-9)
TF	0.0	0	0	0	0	9
CS	13.3	7.7	0.9	10	0.7	5
JP	0.0	0	0	0	0	9
JF	2.0	1.0	0.4	2	1.0	9
JG	1.0	0.5	0.3	1	1.0	9
NA	1.0	0.5	0.9	0	0	9
MB	6.1	3.1	0.4	3	0.5	8
JP	32.7	16.3	1.1	30	1.7	3
DV	0.0	0	0	0	0	9
HL	42.9	25.0	0.4	45	1.5	3
RM	0.0	0	0	0	0	9
LM	3.1	2.0	0.7	1	0.3	8
JS	13.3	6.6	0.2	4	0.4	6
TB	44.9	27.6	0.6	47	1.5	3
JN	1.0	0.5	0.2	1	1.0	9
PM	3.1	2.0	0.3	2	0.8	8
FC	6.1	7.7	1.21	7	0.7	8
LA	0.0	0	0	0	0	9
AS	15.3	9.2	0.5	7	0.4	6
NM	3.1	1.5	1.4	0	0	8
JD	8.2	4.1	0.9	22	1.3	4
DC	0.0	0	0	0	0	9
AC	6.1	3.6	1.3	2	0.4	8
Mean	8.8	5.2	0.5	8.0	0.6	7

Table 2-8. Validation study results of the monologue task.

AWS	%SW (%)	%SS (%)	Duration mean (s)	Mean number of associated behaviours	Mean tension degree (0-3)	Naturalness (1-9)
TF	14.7	10.1	0.7	3	0.2	5
CS	26.7	18.9	1.4	36	1.2	3
JP	5.3	2.8	0.5	3	0.4	8
JF	10.3	6.6	0.6	9	0.5	7
JG	2.4	1.0	0.8	3	1.3	9
NA	11.0	6.2	0.2	7	0.5	7
MB	16.1	8.9	0.6	9	0.5	5
JP	18.0	10.5	0.5	35	1.3	3
DV	7.5	5.2	0.3	4	0.3	8
HL	13.0	8.1	0.2	24	1.2	4
RM	0.6	0.3	0.3	1	2.0	9
LM	2.6	1.6	0.5	2	0.6	9
JS	10.3	5.3	0.2	8	0.5	7
TB	48.4	32.7	0.5	49	1.5	2
JN	1.6	0.9	0.3	0	0	9
PM	1.8	1.0	0.2	3	1.0	9
FC	14.0	13.3	0.3	15	1.4	5
LA	28.6	9.9	1.2	38	0.7	3
AS	8.2	5.3	0.4	4	0.3	7
NM	13.4	7.5	0.6	3	0.8	6
JD	11.4	5.0	0.7	49	1.2	2
DC	8.5	5.5	0.4	12	0.6	5
AC	8.2	4.6	0.6	35	0.5	2
Mean	12.1	7.4	0.5	15.2	0.8	6

Table 2-9. Validation study results of the dialogue task.

AWS	%SW (%)	%SS (%)	Duration mean (s)	Mean number of associated behaviours	Mean tension degree (0-3)	Naturalness (1-9)
TF	9.9	5.5	0.7	6	0.3	5
CS	4.5	3.1	1.1	10	1.3	4
JP	4.9	3.4	0.3	4	0.5	6
JF	5.9	4.8	0.6	3	0.2	6
JG	2.5	1.8	0.3	1	0.2	7
NA	5.5	3.2	0.2	2	0.3	7
MB	15.3	9.6	1.2	18	1.2	4
JP	21.8	13.3	0.3	24	0.8	2
DV	1.1	0.6	0.3	0	0	9
HL	8.7	5.1	0.4	16	1.8	4
RM	1.3	0.6	0.2	0	0	9
LM	3.1	1.8	0.6	2	0.3	9
JS	10.2	5.7	0.3	12	0.9	4
TB	46.2	31.5	0.6	71	1.6	2
JN	1.1	1.2	0.2	0	0	9
PM	3.3	1.4	0.3	1	0.2	9
FC	13.0	5.5	1.2	24	1.5	3
LA	5.9	3.5	0.3	5	0.5	7
AS	5.6	3.5	0.4	4	0.4	7
NM	10.4	6.0	0.4	16	1.0	4
JD	5.8	2.7	1.2	32	1.3	3
DC	7.1	4.2	0.5	10	0.9	5
AC	7.5	3.8	0.5	30	1.0	3
Mean	8.7	5.3	0.5	12.7	0.7	6

Table 2-10. Validation study results of the telephone task.

AWS	%SW (%)	%SS (%)	Duration mean (s)	Mean number of associated behaviours	Mean tension degree (0-3)	Naturalness (1-9)
TF	11.0	6.1	1.1	18	1.3	4
CS	10.3	7.7	1.1	10	1.0	4
JP	5.3	3.3	0.5	2	0.3	7
JF	12.4	5.9	0.5	8	0.6	5
JG	2.7	1.4	0.9	3	1.5	8
NA	8.9	4.8	0.4	3	0.2	6
MB	12.5	6.7	0.4	4	0.5	5
JP	35.7	15.6	0.7	31	1.0	2
DV	6.0	3.1	0.2	0	0	7
HL	9.2	9.2	0.4	15	1.3	4
RM	2.1	1.1	0.3	0	0	9
LM	1.7	0.9	0.2	2	0.7	9
JS	8.9	5.2	0.5	7	0.7	5
TB	46.2	37.1	0.8	48	1.5	2
JN	0	0	0	0	0	9
PM	6.8	3.3	0.4	7	1.5	6
FC	19.6	7.8	1.6	29	1.5	2
LA	9.2	6.4	0.4	15	0.7	4
AS	10.6	7.6	0.4	9	0.4	5
NM	16.0	9.5	0.7	25	0.9	3
JD	8.9	4.5	1.0	54	1.2	2
DC	8.6	4.2	0.4	11	0.8	5
AC	13.2	6.4	0.6	38	0.7	2
Mean	11.6	6.9	0.6	14.7	0.8	5

Results of the validation study indicated that the monologue task presented the highest values of frequency (in %SS), followed by the telephone task, the dialogue task and the reading task. Concerning duration, the telephone task presented, on average, longer stuttering moments, followed by the other three speech tasks with similar duration values. The number of associated behaviours are higher in the monologue task, followed by the telephone, dialogue ad reading task. The tension degree mean is similarly higher in the monologue and telephone speech sample, followed by the dialogue sample and the reading sample. In general, the monologue turned out to be the speech task with more frequent and longer stuttering moments, with more tension associated behaviours. The naturalness mean level decrease from the reading speech sample (mean 7) to the telephone speech sample (mean 5).

Concerning the descriptor of the stuttering event in terms of type of stuttering (using the LBDL taxonomy), all speech samples presented fixed postures without audible airflow with the highest mean percentage (36.3%, 31.7%, 25.9% and 27.5%), for reading, monologue, dialogue and speech sample on the telephone, respectively. Regarding the associated behaviours, the classification with the LBDL taxonomy revealed that nonverbal extraneous behaviours presented the mean highest percentage (49.6%, 72.6%, 69.2% and 66.7% for reading, monologue, dialogue and speech sample on the telephone, respectively).

As the behavioural measures determined through *SABES* procedures for each speech sample presented different metrics, a task score was derived to allow the calculation of a final raw score and, therefore, a severity equivalence.

The percentage of syllables stuttered is not expected to be normally distributed (see Figure 2-4), because there are more mild than severe cases, which makes the measure positively skewed (Jones, Onlow, Harrison, & Packman, 2000; Jones, Onslow, Packman, & Gebski, 2006; O'Brian, Packman, Onslow, et al., 2004; Tumanova, Conture, Lambert, & Walden, 2014). Jones et al. (2006) has previously shown that the %SS is not normally distributed and that it can be adequately modelled with a gamma distribution based on the following arguments: The shape of the gamma distribution closely corresponds to the shape of the distribution of %SS; goodness-of-fit tests; Q-Q plot suggest that the gamma distribution describes the distribution of %SS; one of the gamma distribution parameters (i.e., the location parameter) can be zero, simulating that all scores are greater than zero as it is the case of %SS. Figure 2-5 illustrates the gamma distribution, in which the yy axis represents density and the xx axis the variable values.

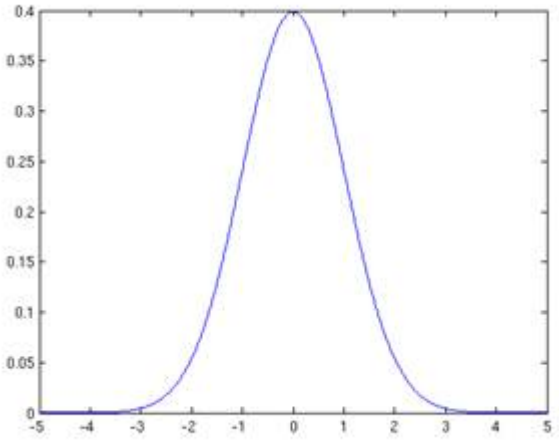


Figure 2-4 Graphical representation of a normal distribution. From Asher (2010).

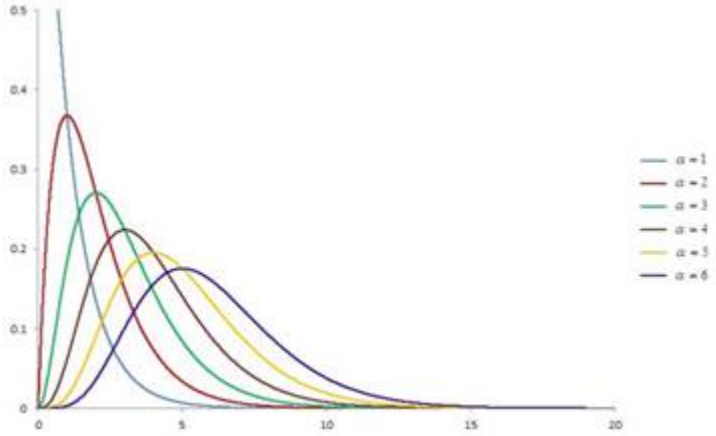


Figure 2-5 Graphical representation of different gamma distributions. From Ma (2015).

The *SABES* scores from the behavioural measures were studied in terms of their distribution for a population of 23 AWS (used during the validity analysis). Considering a gamma distribution fit to the behavioural measures distribution (similar to what happen in frequency in %SS), the parameters of gamma distribution were calculated using Excel 2013 (i.e., shape parameter and scale parameter).

To calculate the task scores, all the percentiles of the gamma distribution were obtained for each behavioural measure within each speech sample. Stanine scales were used to convert scores from different behavioural measures and represent “an examinee’s position relative to other test takers’ positions” (Salkind, 2007, p. 951). According to Salkind (2007) the stanine value is: 1 for the first 4% of the individuals in the reference group; 2 for the next

7%; 3 for the next 12%; 4 for the next 17%; 5 for the next 20%; 6 for the next 17%; 7 for the next 12%; 8 for the next 7%; 9 for the last 4% of the reference group percentage.

Tables 2-11 to 2-14 present the stanine for each possible value in all different behavioural measures for each speech sample assessed by *SABES*. SABES's total score was obtained by adding the stanines from the behavioural measures of all speech sample (%SS, mean duration, associated behaviours and tension degree mean).

Table 2-11. Stanine values for each behavioural measure on the reading speech task.

%SW (%)		**%SS (%)**		**Duration mean (s)**		**Mean number of associated behaviours**		**Mean tension degree (0-3)**	
%	Stanine	%	Stanine	Mean	Stanine	Sum	Stanine	Mean	Stanine
1.0-2.0	4	0.5-1.2	4	0.1	3	1	4	0.1	3
2.1-5.8	5	1.3-3.4	5	0.2	4	2-4	5	0.2-0.3	4
5.9-12.5	6	3.5-7.3	6	0.3-0.4	5	5-10	6	0.4-0.5	5
12.6-22.8	7	7.4-13.3	7	0.5-0.7	6	11-21	7	0.6-0.8	6
22.9-38.5	8	13.4-22.4	8	0.8-1	7	22-38	8	0.9-1.1	7
>38.6	9	>22.5	9	1.1-1.4	8	>39	9	1.2-1.7	8
				>1.5	9			>1.8	9

Table 2-12. Stanine values for each behavioural measure on the monologue speech task.

%SW (%)		**%SS (%)**		**Duration mean (s)**		**Mean number of associated behaviours**		**Mean tension degree (0-3)**	
%	Stanine	%	Stanine	Mean	Stanine	Sum	Stanine	Mean	Stanine
0.4-1.0	1	0.1-0.3	1	0.1	1	1	2	0.1	1
1.1-2.2	2	0.4-1.0	2	0.2	3	2-3	3	0.2	2
2.3-4.2	3	1.1-2.2	3	0.3-0.4	4	4-7	4	0.3-0.4	3
4.3-7.3	4	2.3-4.0	4	0.5	5	8-14	5	0.5	4
7.4-11.8	5	4.1-6.9	5	0.6-0.7	6	15-23	6	0.6-0.8	5
11.9-17.6	6	7.0-10.8	6	0.8	7	24-35	7	0.9-1.0	6
17.7-25.1	7	10.9-15.8	7	0.9-1.0	8	36-52	8	1.1-1.4	7
25.2-35.0	8	15.9-22.5	8	>1.1	9	>53	9	1.5-1.7	8
>35.1	9	>22.6	9					>1.8	9

Table 2-13. Stanine values for each behavioural measure on the dialogue speech task.

%SW (%)		%SS (%)		Duration mean (s)		Mean number of associated behaviours		Mean tension degree (0-3)	
%	Stanine	%	Stanine	Mean	Stanine	Sum	Stanine	Mean	Stanine
0.1	1	0.1-0.2	2	0.1	1	1	2	0.1	2
0.2-0.7	2	0.3-0.8	3	0.2	3	2-3	3	0.2	3
0.8-1.8	3	0.9-2.0	4	0.3-0.4	4	4-6	4	0.3-0.4	4
1.9-4.0	4	2.1-4.3	5	0.5	5	7-11	5	0.5-0.6	5
4.1-7.6	5	4.4-7.7	6	0.6-0.7	6	12-18	6	0.7-0.9	6
7.7-12.7	6	7.8-12.5	7	0.8-0.9	7	19-27	7	1.0-1.3	7
12.8-19.7	7	12.6-19.3	8	1.0-1.1	8	28-40	8	1.4-1.8	8
19.8-29.4	8	>19.4	9	>1.2	9	>41	9	>1.9	9
>29.5	9								

Table 2-14. Stanine values for each behavioural measure on the telephone speech task.

%SW (%)		%SS (%)		Duration mean (s)		Mean number of associated behaviours		Mean tension degree (0-3)	
%	Stanine	%	Stanine	Mean	Stanine	Sum	Stanine	Mean	Stanine
0.3-0.8	1	0.1	1	0.1	2	1	2	0.1	1
0.9-1.9	2	0.2-0.6	2	0.2	3	2-3	3	0.2	2
2.0-3.7	3	0.7-1.5	3	0.3-0.4	4	4-7	4	0.3	3
3.8-6.6	4	1.6-3.2	4	0.5-0.6	5	8-13	5	0.4-0.5	4
6.7-11.0	5	3.3-6.0	5	0.7-0.8	6	14-21	6	0.6-0.8	5
11.1-16.7	6	6.1-10.0	6	0.9-1.0	7	22-33	7	0.9-1.0	6
16.8-24.0	7	10.1-15.4	7	1.1-1.3	8	34-48	8	1.1-1.4	7
24.1-33.8	8	15.5-22.9	8	>1.4	9	>49	91	1.5-1.7	8
>33.9	9	>23.0	9					>1.8	9

Verbal descriptors can also be attached to individual stanine scores or to groups of adjacent stanine scores. The stanine scale can be divided in three parts (above average, average and below average), five parts (outstanding, above average, average, below average and poor) or one could label each stanine score (high, well above average, above average, somewhat above average, about average, somewhat below average, below average, well below average and low) (Salkind, 2007). The SABES total scores were first converted into percentile scores and then into 1-9 stanine scores. Groups of adjacent stanine scores were

labelled. The *SABES* total scores were divided into five parts, in which the label very mild was added to stanine 1, mild to stanine 2 and 3, moderate to stanines 4, 5 and 6, severe to stanine 7 and 8 and very severe to stanine 9 (Table 2-15). The labels adopted were similar to those used in the SSI-4 (Riley, 2009).

Table 2-15. Severity levels for SABES total score.

SABES total score	**Label**
27-35	Very mild
36-54	Mild
55-93	Moderate
94-129	Severe
>130	Very severe

Related to criterion-related validity, the *SABES* total score was correlated with percentage of syllable stuttered (%SS) obtained on each speech task. Spearman correlation values were 0.85, 0.60, 0.61 and 0.76, for the reading, monologue, dialogue and speech telephone tasks, respectively. Correlation values were significant at the p> 0.01 level.

Concerning construct validity, a matrix of correlations was made to verify the following underlying construct hypothesis: All the parameters of the *SABES* are related to each other on a useful degree (i.e., between 0.30 and 0.70), because they are parts of the overall construct of severity). Table 2-16 presents results of the Spearman correlation between the behavioural measures of the total sample collected, which were all statically significant beyond the 0.01 level. The coefficients ranged from 0.472 to 0.831 with a median of 0.55.

Table 2-16. Spearman correlation between the behavioural measures for the total speech sample.

	Frequency	Duration	Associated behaviours	Tension degree
Frequency	1.000	0.496	0.831	0.548
Duration	0.496	1.000	0.549	0.472
Associated behaviours	0.831	0.549	1.000	0.761
Tension degree	0.548	0.472	0.761	1.000

2.5. Discussion

2.5.1. Content validity analysis

The representativeness of the items of the *SABES* instrument to the content that it is intended to be measure was analysed. Conceptualisation of the content domain was obtained through different sources, such as literature review, expert consultation or a qualitative inquiry (Polit & Beck, 2012). *SABES*'s two-stage process of the content validity process (developmental stage and judgment stage) results showed that the procedures are representative and appropriate for the content that is intended to measure (Lynn, 1986). The exhaustive collection and exploration of instruments that assessed severity of stuttered events in addition to bibliography analysis captured the full content domain and include the majority of the behavioural measures used in the more recent instruments and the procedures with better established reliability (Polit & Beck, 2012).

Results obtained from the modified Bland and Altman (1986) analysis (Jesus et al., 2015, pp. 4–5) guided the modifications in the pre-final version of the *SABES* to achieve clarity, simplicity and accuracy of the non-agreed items. Using this method, we ensured that items were modified in agreement with to expert's opinion, to develop a final version of the instrument that was valid.

The use of a heterogeneous expert panel to assess content validity allowed the development of an instrument that is valid and acceptable to the professionals that could use the instrument (Schilling et al., 2007).

2.5.2. Pilot study

Procedures developed during the first stage of the content validity analysis and modified in terms of clarity and simplicity after the expert panel meeting proved to be feasible and easy to use during the pilot study. The absence of difficulties in using the procedures lead us to retain all the instructions in the final version and conclude that *SABES* was feasible to be used in a larger population (Leon et al., 2011).

2.5.3. Reliability analysis

Cronbach's alpha results indicated that the *SABES* items/behavioural measures presented good internal consistency, as results were higher than the cut-off threshold of 0.7 (Bland &

Altman, 1997; Field, 2013; Kline, 2004). Values higher than 0.7 indicated that the items that constitute the *SABES* are homogenous and measured the same underlying construct (i.e., severity of stuttering events). As the behavioural measures are related to each other, the amount of error in the instrument is likely to be residual (Gillam et al., 2009).

2.5.4. Criterion and construct validity analysis

Criterion-related validity was assessed through the correlation of *SABES* total score with an external criterion, i.e., with frequency of stuttering moments. According to Hopkins (2002) the correlation values obtained presented a magnitude between large (0.5 to 0.7) to very large (0.7 to 0.9). The magnitude of the correlation obtained indicated that the *SABES* total score is associated with a common and most used behavioural measure to assess stuttering moments, which stablished criterion validity (Gillam et al., 2009; Guitar, 2014; Jones et al., 2006; Riley, 2009).

Evidences of construct validity were obtained through the correlation between the values of *SABES* behavioural measures obtained for the total speech sample collected. The median values obtained were between 0.30 and 0.70, meaning that they correlate in a useful degree, but they do not measure the same behaviour, meaning that each behavioural measure presented a singular contribution to the total score (Gillam et al., 2009; Riley, 2009). The results indicate that the behavioural measures assess aspects of fluency that are different but related to each other, allowing "(…) diagnostic decisions or investigating intraindividual speech fluency differences" (Gillam et al., 2009, p. 74).

Concerning the classification of events of stuttering, it was shown that, globally, the monologue speech sample presented more frequent, longer and with more tension events of stuttering and the reading sample presented the less frequent, long and with less tension moments. Traditionally, reading is perceived as an easier speech sample, when compared with monologue (Venkatagiri, 2005), which has been observed in the present validation study. Regarding the classification of the stuttering moments using the LBDL taxonomy (for types of stuttering and associated behaviours), it was revealed that all speech samples presented a higher percentage of fixed posture without audible airflow when compared with repeated movements. This finding is consistent with the description of the characteristics of stuttering in adults, that point to tense and long blocks (i.e., a term often used by clinicians to denominate fixed postures without audible airflow (Onslow, 2017)) as the most frequent behaviour (Guitar, 2014).

2.6. Conclusions

The *SABES* is an assessment tool designed to assess the characteristics of the observable features of stuttering, in order to determine the severity level of the observable stuttering moments. Despite the existence of several instruments to assess severity of stuttering, none use several speech samples (that are used to capture and to deal, at least in part, with the variability of stuttering severity) assessed with specific procedures based in annotated speech samples. Through the assessment based on a software, the examiner can overcome the inherent difficulties of a perceptual evaluation.

The *SABES* went through a two-stage process to guarantee that procedures adequately measure the domains identified. The *SABES* content validity has been ensured through the analysis of instruments that assess moments of stuttering in terms of type of behavioural measures and the procedures used, the development and quantitative and qualitative judgement of the *SABES* procedures by an expert panel.

Pilot study results allowed the refinement of a standardised methodology to assess stuttering moments, based on video and audio recordings using the ELAN. The appropriateness of the use of the *SABES* on larger scale research was ensured.

Evidences of criterion validity were revealed by the large significant statistical correlation between *SABES*'s total score and an external criterion: Frequency (in %SS). In terms of construct validity evidence, the correlation results indicated that all of the *SABES*'s behavioural measures are related and contribute differently to the severity level obtained, which probes the construct basis of the assessment tool. The validity evidence found allow us to propose that the interpretation of the scores obtained is grounded by the theory that underlies the *SABES* construction (AERA 2014). Study limitations include the small sample size of the pilot study. Future research will include additional reliability studies, specifically related to inter and intra-judge reliability and the usability of the *Excel* spreadsheets.

Chapter 3 - Assessment of Language Use in Social Contexts for Adults (*ALUSCA*)

3.1 Introduction

The Assessment of Language Use in Social Contexts for Adults (*ALUSCA*) is a specific, adult focused and self-administered questionnaire designed to assess pragmatic language competencies (PLC) in AWS and adults who do not stutter. In the present chapter, the content validity process, item analysis, reliability coefficients and evidences of construct validity will be present.

3.2 Background

3.2.1 Language and Pragmatics

Language can be defined as a complex and dynamic system of conventional symbols, shared socially (Owens, 2011). Language has three major components: Form, content and use. Form encompasses syntax, morphology and phonology; content encompasses semantics (or word meaning); and use encompasses pragmatics. The rule system of language functions with this set of components. When people use language, they code ideas (semantics) through a symbol (e.g., a sound or a word) using forms that include the use of sound units (phonology), word order (syntax) and word formation (morphology) to convey the meaning and to achieve different communication ends (pragmatics) (Owens, 2011). The components of language are intrinsically linked. According to the Functionalist Model of language, pragmatics is considered as the organising principle of language (Owens 2011), as context influences language and determines communication options in terms of content and form.

Pragmatics is an important aspect of language functioning, regulated by sociolinguistic rules related to language use, and manipulation of language competencies in different communicative contexts (Owens, 2011; Weiss & Zebrowski, 1994). Pragmatics is based on communication intentions, conversational principles or rules and on the use/construction of different types of narratives (Owens 2011). The appropriate and effective use of language in context depends on the speaker's pragmatic language competencies (PLC) (Russell &

Grizzle, 2008). The appropriateness of an utterance in a context is dependent, in part, of grammar, vocabulary and meaning; however the way that a child used language in different contexts may have a greater contribute to adjustment and social success than mastering the traditional language areas (Ninio & Snow, 1999; Russell & Grizzle, 2008). The idea is corroborated by Fey (1986), who referred that the ability to use language in context is not always reflected in the semantic or syntax output of the children.

Persistent difficulties with the social use of verbal/nonverbal communication are characteristics of "social (pragmatic) communication disorder", a new diagnostic category added to the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* (DSM-V) by the American Psychiatric Association (American Psychiatric Association, 2013). Deficits can be expressed in terms of difficulties in understanding and following social rules of communication, changing language according to the listener or communicative situation, obeying the rules related to conversation or storytelling, and making inferences or understanding nonliteral language Social (pragmatic) communication disorder is common in individuals with attention-deficit/hyperactivity disorder, autism spectrum disorder, language impairment or specific learning disorders (American Psychiatric Association, 2013).

3.2.2 Stuttering and pragmatics

Stuttering is a multi-faceted fluency disorder characterised by observable disruptions in speech, with different physiological, behavioural, emotional and cognitive reactions to communication (Bloodstein & Bernstein Ratner, 2008; Guitar, 2014; Manning, 2010; Shapiro, 2011). Stuttering can cause negative social and psychological consequences in everyday life (Bleek et al., 2012). Developmental stuttering usually begins between the ages of 2 and 7 when vocabulary is growing rapidly and many aspects of language mastery are occurring. Psycholinguistically-oriented models of stuttering suggest there is a direct relationship between stuttering and language (Bernstein, 1981; Bernstein Ratner, 1997; Bloodstein, 2006; Bloodstein & Bernstein Ratner, 2008; Howell, 2004; Howell & Au-Yeung, 1995; K. Logan & Conture, 1995; Ntourou, Conture, & Lipsey, 2011; Postma & Kolk, 1993; Wingate, 1988). A meta-analysis of studies investigating evidences of language abilities in children who stutter concluded that the scores on norm-referenced assessments of language, receptive and expressive vocabulary and mean length of utterance are significantly lower than controls (Ntourou et al., 2011). Regarding the relationship between syntax and developmental stuttering, Crystal (1987) suggested that syntactic demands

increase the vulnerability to maintain fluency; Weiss (2004) further suggested that "if excessive demands on a child's syntax repertoire can have a negative effect on fluency maintenance, then it follows that excessive demands on pragmatic skills could also negatively affect fluency maintenance" (p. 36).

Various studies of the relationship between pragmatic components and stuttering in children have been published (Nippold, Schwarz, & Jescheniak, 1991; Weiss, 2004; Weiss & Zebrowski, 1994). Nippold *et al*. (1991) and Weiss and Zebrowski (1994) studied the narrative productions of children who stutter. Nippold *et al*. (1991) concluded, with a small number of children (N=10 children who stutter and 10 children who do not stutter) that children who stutter were not more vulnerable to language disorders nor different in narrative tasks when compared with children who do not stutter. Weiss and Zebrowski (1994) studied the narrative production of 16 children, half of them with developmental stuttering, and observed nonsignificant differences between both groups; however, stories produced by children who stutter were shorter and less elaborate whether the listener was more familiar or less familiar with the story, but with the same critical information. Weiss and Zebrowski (1994) suggest that the production of utterances with a limited length by children who stutter could be due to an increased possibility that stuttering occurs in longer, more complex sentences. Additionally, the authors highlight that the non-recognition that an unfamiliar listener needs more information "(…) can lead to communication breakdowns and may indicate a deficit in the speaker's knowledge of the pragmatic aspects of language (…)" (Weiss & Zebrowski, 1994, p. 55).

Weiss (2004) showed that there is a propensity for delay or differences in certain areas of language in children who stutter in comparison to children who do not stutter, but also indicated that individual differences on different subjects should be considered to decide the importance of contemplating pragmatic competencies in the treatment plan.

The impact and the relationship of pragmatics and stuttering in adults were not well established. Two studies from Spencer, Packman, Onslow and Ferguson (2005, 2009) with AWS and adults who do not stutter used a sociolinguistic approach to determined differences in the use of language concerning quantitative discourse measures (e.g., number of words, utterances and clauses), complexity (i.e., quantification of information into an utterance, reporting by the average number of clauses in an utterance), salience (i.e., resources that the speaker can use to highlight information to the interlocutor) and modality (i.e., resources to engage the interlocutor through language use, such as convey of opinions, attitudes and politeness). Results of the comparison study between a group of 10

AWS matched with 10 adults who do not stutter (Spencer et al., 2009) revealed that the language used by the experimental group was significantly less complex and uses less resources for engagement than the control group. Spencer et al (2005) studied linguistic resources for engagement pre- and post an intervention period and found an increase in modality after intervention in conversation. The findings highlight that stuttering may adversely impact verbal expression of someone who stutter, affecting their communication effectiveness. The authors imply that linguistic differences can potentially impact interactions, social participation and communication opportunities and, in a broader view, the use of language in contexts, as the AWS "may be withholding themselves from engagement in interactions, perhaps to minimize the uncertainty inherent in normal conversation exchange (…)" due to stuttering itself and to the way that language was used (Spencer et al., 2009, p. 485). Findings from Spencer et al (2009) were observed on a monologue task, in which the range of opportunities to use linguistic engagement resources were less likely to occur when compared with conversation. As pragmatics consist in combining all language components by conquer the rules for use language in functional and socially appropriate way (ASHA, 1993), it seems rational to consider the impact of pragmatics in someone who stutter, given the additional demanding related to the use of language on a social context that can emphasize and intensify the restrictions on engagement raised on the Spencer et al (2009) study. Those contextual challenges, such as talking in different speaking situations, were dreaded by adolescents and AWS and inducing, in some cases, to the exacerbation of disfluencies and to the avoidance of situations (Brundage, Bothe, Lengeling, & Evans, 2006; Manning, 2010; Silverman & Ratner, 1997). Avoidance are referred by Cream et al (2003) as an adaptation or a way to control communication aspects (e.g., speech output or contribution in a speaking exchange) in order to protect from harm and lack of control in situations where stuttering is more likely to occur. This observation leads to the consideration, by some clinicians, that the demands to use language components in social contexts (pragmatic) presents an impact on stuttering (Cummings, 2014; Weiss, 2004). Based on the results of Silverman and Ratner (1997), Weiss (2004) suggested that school-age children and adolescents may be less influenced by semantic and syntactic linguistic demands than by the pragmatic demands (i.e., contextual challenges) on speaking situations. Tanner, Belliveau, and Seibert, (1999) referred, in their intervention approach, that the anxiety associated with specific communicative situations leads to difficulties in the use of language in context, which can lead to avoidance on AWS that, consequently affects language experiences and the development of pragmatic skills. Cummings (2014) stated that "as more becomes known

about the pragmatic problems of this clinical population, and the specific influence of communicative context on speech fluency, clinicians will increasingly have to consider the contribution of pragmatics to the adverse psychological outcomes of people who stutter" (pp. 150).

The influence of pragmatic competencies is expressed in several items of published instruments used to assess the person's speaking or stuttering in various contexts. Although they do not mention the term pragmatics, they are referring to language use in context in AWS. Those instruments include self-assessment tools of communication attitudes, perceptions and reactions of AWS to speaking situations (Andrews & Cutler, 1974; Brutten & Shoemaker, 1974; Johnson, Darley, & Spriestersbach, 1952; Ornstein & Manning, 1985; Riley et al., 2004; Tanner et al., 1999; Woolf, 1967; Wright & Ayre, 2000; Yaruss & Quesal, 2006).

A conclusion can be drawn that a specific investigation of pragmatic competencies in AWS may provide additional information for clinicians to develop a therapeutic intervention process based on the level of impact of stuttering on language use.

3.2.3 The pragmatic language competencies framework

Russel and Grizzle (2008) developed a coding system, to evaluate the core PLC domains commonly assessed using checklists, questionnaires and tests related to pragmatics. This was "(…) drawn from the theoretical and research literatures on pragmatics" (Russell & Grizzle, 2008, p. 61), identifying several pragmatic domains, and structuring these into a framework, composed by three main categories: 1) precursors/enablers; 2) basic exchanges; 3) extended literal and non-literal discourse. The first category includes basic competencies to participate in an interaction; the second category comprises PLC that facilitate discourse exchanges; the third category contains domains that assist the participation in extended discourses (Russell & Grizzle, 2008). Table 3-1, Table 3-2 and Table 3-3 presents Russel and Grizzle's (2008) definitions for seventeen PLC as well as the PLC included in each framework category.

Table 3-1. Russel and Grizzle's (2008) framework of pragmatic language competencies.

Category	PLC	Definition
Precursors/enablers	Nonverbal communication	Capacity to attend and understand nonverbal behaviours
	Discourse attentiveness and empathy	Capacity to focus and concentrate on the interlocutor and/or their communication; involve in empathic listening
	Speech characteristics and fluency	Capacity to make speech sounds, with correct articulation and/or prosody
	Rituals, greetings and goodbyes	Capacity to use greetings and farewells and politeness formulas
	Vocabulary	Capacity in using appropriate vocabulary and/or to understanding and using different classes of words
	Comprehensibility	Capacity to be engaged in an understandable communicative exchange

Table 3-2. Russel and Grizzle's (2008) framework of pragmatic language competencies (continued).

Category	**PLC**	**Definition**
Basic exchanges	Conversational turn taking	Capacity to take turns in a communication exchange and to adopt roles as a speaker or as a listener appropriately
	Topic control and maintenance	Capacity to initiate, maintain or change and to finish a conversation topic in a proper manner
	Requests	Capacity to formulate simples and complex requests, questions and answers
	Speech acts	Capacity to use different speech acts
	Syntax and grammar	Capacity to construct correct sentences (in terms of tenses, subject or verbs agreement)
	Interlocutor variety	Capacity to participate in a conversation with a variety of interlocutors and in different contexts

Table 3-3. Russel and Grizzle's (2008) framework of pragmatic language competencies (continued).

Category	PLC	Definition
Extended Literal and Nonliteral Discourse	Negotiations, directions and instructions	Capacity to negotiate in a communicative exchange, offer directions, instructions and recipes
	Theory of mind and emotion language	Capacity to use internal state cognitive or emotional language; ability to taking perspectives or attribute intentions
	Narrative	Capacity to use narratives and describe events
	Nonliteral language, use of indirections and presupposition	Capacity to use nonliteral sentences and indirect language; ability to use and understand presupposed knowledge in a conversation
	Gricean principles	Capacity to satisfy the Gricean maxims of quantity, quality, relation, manner and co-operativeness

Their framework presented a value of 0.84 of Cohen's Kappa, meaning that there is a substantial agreement between the two raters that used the framework to classify questionnaire/checklist and test item (Kundel & Polansky, 2003; Russell & Grizzle, 2008). Using the framework described, Russel and Grizzle (2008) identified the core PLC (i.e., the PLC that are commonly assessed by pragmatics instruments) based on two indices of salience. The first index was calculated by counting the number of pragmatics instruments that had at least one item for each of the 17 PLC domains identified. Core or central PLC "(…) would be assessed by the most checklists/questionnaires and tests, whereas peripheral PLC domains would be assessed by only a few" (Russell & Grizzle, 2008, p. 61). The second index was calculated by counting how many items assessed each PLC on the pragmatic instruments. According to Russel and Grizzle (2008, p. 62), a core domain "(…) would be probed by many items". The index of salience was calculated though the attribution to each PLC a ranking based on the number of questionnaires/checklists and tests that include probes to assess it and in terms of the total number of items that probe the PLC in the totality of the assessment instruments considered. The rankings were summed for each type of assessment instrument to conclude the relative salience (i.e., the

importance) of each PLC. Russell and Grizzle (2008) evaluated thirteen questionnaires that assess PLC in children and adults and identified the six core or most salient domains, namely: Requests; Speech Characteristics and Fluency; Nonverbal Communication; Topic Control and Maintenance; Communication Turn-taking; Negotiations, directions and instructions.

2.4 Aims

To date, there is no specific instrument that assess the totality of PLC in adults, therefore we developed the *ALUSCA,* a specific and adult focused self-administered questionnaire created to assess a person's ease in using language in context. The procedure is described in outline, along with the development and the establishment of the content validity of *ALUSCA*. In addition, assessment of construct validity and reliability coefficients are reported and discussed.

3.3. Method

3.3.1 Content validity analysis

Content validity represents the content relevance of the items of an instrument. It was determined by the application of a two-stage process (Lynn, 1986): 1) developmental; 2) judgment-quantification. Lynn (1986) claims that the two-stage process is essential for the validation process of instrumentation. The developmental stage (initial stage) comprised three steps: Identification of full content domains, sampling and item generation and assimilation of items into a useable scale. The domain identification consisted of the categorisation and description of the dimensions to be assessed (Lynn 1986).

Following an extensive literature review related to pragmatic development and pragmatic language competencies on ERIC, MEDLINE, PubMed and B-on and on specific literature (e.g., Cummings, 2014; Horn & Ward, 2006) the framework of Russel and Grizzle (2008) was identified as a broad and systematic coding system, and it was used as the framework basis to questionnaires assessment and the development of the *ALUSCA* items. Eleven instruments were assessed initially (two related to pragmatics and nine related to communication attitudes and reactions to speaking situations in AWS to locate the items that probe each PLC, through the development of tables of contents.

The second step in the first stage involved item generation for the domains identified in step 1. This resulted in producing a first version of the ALUSCA items based on the items identified on the eleven instruments reviewed. The rating scale was a Likert-type scale, in which 1 indicates “never easy” and 5 indicates “always easy”; the ease level was chosen (instead of the difficult level) in keeping to the philosophy of ICF (WHO, 2001) that focuses on what clients can do rather than what they cannot do (Perry et al., 2004). The third step of the developmental stage consisted in refining the items generated on step two into a useable form.

A second stage of the process was implemented to determine and ensure content validity using judgment-quantification. This stage comprised two steps: The first consisted of judging items and the second, judging the entire instrument, both done by an expert panel. Lynn (1986) proposes three as the minimum number of experts that should integrate the panel. The experts selected should have expertise in the domain studied; they should be informed about the full dimensions of the variables studied, and a specific instrument should be used to assess relevant content on each item (Grant & Davis, 1997).

This initial version of the *ALUSCA* was submitted to a focus group, comprising three AWS and three controls. Using the thinking aloud method (Goldman, 1971) the opinions of the focus groups elements were recorded while they were expressed during the questionnaire completion. A second version of the *ALUSCA* was constructed and submitted for review to a panel of experts (stage 2 – judgment-quantification stage) comprised by three experts with experience in stuttering, instrument development and pragmatics. The contents of the questionnaire (i.e., the items developed), the general goals, the conceptual basis, and the measurement model of *ALUSCA* were presented to the expert panel. The objectives of the expert meeting were to scrutinise the *ALUSCA* item-by-item; to rate the *ALUSCA* items with a VAS (Crichton, 2001; Hasson & Arnetz, 2005); to add or remove words or sentences from the *ALUSCA* as necessary to achieve the relevant content (Perry et al., 2004). The experts involved assessed the items in terms of relevance, clarity, simplicity and accuracy, using an instrument review questionnaire with 104 questions. Again, the thinking aloud method was used to record the opinions related to each item’s relevance, clarity, simplicity and accuracy. The Bland and Altman (1986) modified approach (Jesus, Valente, and Hall 2015, pp. 4–5) was used to observe the agreement/disagreement between the three experts on each question of the instrument review questionnaire. The pre-final version of the *ALUSCA* was developed after the judgment phase by the focus groups and the expert panel.

3.3.2 Pilot study

A pilot study was conducted using the pre-final version, in order to examine the feasibility of the questionnaire to be used with a larger group (Leon et al., 2011). Five AWS and five controls matched by gender and age completed the paper and pencil questionnaire. AWS were recruited based on the following inclusion criteria: 1) self-reports of onset, development and chronicity of stuttering; 2) classification as a PWS by two SLPs that agreed on the diagnosis, 3) be able to read and 4) provide informed consent. In addition to age and gender matching, the inclusion criteria for controls were: 1) absence of stuttering; 2) be able to read and 3) provide informed consent.

The mean of each part 1 variable and part 2 categories were calculated as the mean of each set of items that composed the variables and the PLC. The pilot study intended to calculate item analysis through pairwise correlation, to compare ALUSCA scores between AWS and controls and to check any difficulties that respondents might face in completing the questionnaire, in order to develop the final version. Descriptive statistics and item analysis (pairwise correlation using Pearson product moment correlation) were performed.

3.3.3 Reliability analysis

To assess *ALUSCA*'s reliability, test-retest and internal consistency were calculated, to ensure that it was a consistent questionnaire (Polit & Beck, 2012).
To assess test-retest reliability, ten adults who do not stutter (five males and five females) completed the questionnaire twice with an interval of 4 weeks. Pearson product moment correlation was used to assess the correlation between the two scores.
Internal consistency was calculated for ALUSCA part 2 scores and final score using pilot study results and also scores from *ALUSCA*'s final version. Cronbach alpha were the coefficients chosen to establish internal consistency of *ALUSCA*.

3.3.4 Construct validity analysis

Construct validity evidence concerns the extent to which the test scores reflect the theory behind the construct being measured (Polit & Beck, 2012).

The sample size to be used on the construct validity study was calculated with the *GPower* (v 3.1.9.2) for α=0.05, 80% power and a medium effect size of 0.5, for a Wilcoxon signed-rank test (for matched pairs). The total sample size resulting from this calculation was 28

AWS and 28 controls. AWS were recruited based on the inclusion criteria: 1) self-reports of onset, development and chronicity of stuttering; 2) classification as a PWS by two SLPs that agreed on the diagnosis, 3) be able to read and 4) provide informed consent. The inclusion criteria for controls were: 1) absence of stuttering; 2) be able to read and 3) provide informed consent, in addition to age and gender matching.

Construct validity evidence were obtained through hypothesised relationships method and factor analysis. With the hypothesised relationship method hypotheses were tested, based on theory or prior research. This method does not constitute a proof of construct validity, but creates important evidence for it (Polit & Beck, 2012). Based on the theoretical approach described by several authors (Cummings, 2014; Tanner et al., 1999; Weiss, 2004) that AWS could present pragmatic difficulties, the hypotheses tested were:

- According to theory, PLC performance was different when compare AWS with controls;
- As *ALUSCA* measures PLC in adult, the questionnaire scores should reflect significant differences between AWS and controls.

We then can infer that the ALUSCA questionnaire presents evidence of construct validity for the assessment of PLC. As Shapiro-Wilk Test revealed that *ALUSCA*'s data is normally distributed (p=0.461, for the entire sample; p= 0.179, for AWS; p= 0.858 for adults who do not stutter), parametric statistics (independent sample t-test) were chosen to assess differences between AWS and controls in all categories and in *ALUSCA*'s final score.

Concerning factor analysis, two Principal Component Analysis (PCA) – one for AWS and another for the control group – were performed to identify the number of underlying dimensions using scores from part 1 variables and the three categories of part 2 (Polit & Beck, 2012). PCA allowed the definition of the underlying structure among different variables by grouping variables that are highly correlated and have a similar profile (Hair Jr, Black, Babin, & Anderson, 2009).

3.5 Statistical procedures

The IBM SPSS Statistics version 23 (IBM, Armonk, New York) was used to calculate Cronbach alpha (internal consistency), Pearson product moment correlation (item analysis and test-retest), PCA (construct validity) and independent sample t-test (to assess differences between AWS and controls.

3.4 Results

3.4.1 Content validity analysis

During the first stage, a systematic search on ERIC, MEDLINE, PubMed, B-on was performed to identified articles related to pragmatic assessment tools and questionnaires used for AWS to assess attitudes to speaking situations. Two questionnaires related to pragmatics and nine questionnaires related to communication attitudes and reactions to speaking situations in AWS were analysed. Concerning pragmatic questionnaires, it was assessed the *Pragmatic profile for adults* (Dewart & Summers, 2003) and the *Pragmatic Profile* (Wiig, Semel, & Secord, 2013). Regarding specific questionnaires developed for AWS, tables of contents were design to study the *Stutterer's Self-Rating of Reactions to Speech Situations* (Johnson et al., 1952), the *Perceptions of Stuttering Inventory* (Woolf, 1967), the *Scale of Communication Attitudes* (Andrews & Cutler, 1974), the *Speech Situation Checklist* (Brutten & Shoemaker, 1974), the *Self-efficacy scaling by adult stutterers* (Ornstein & Manning, 1985), the *Pragmatic Stuttering Intervention* (Tanner et al., 1999), the *Wrigth and Ayre Stuttering Self-rating profile* (Wright & Ayre, 2000), the *Subjective Screening of stuttering severity, locus of control and avoidance* (Riley et al., 2004) and the *Overall Assessment of the Speaker's Experience of Stuttering* (Yaruss & Quesal, 2006).

The two pragmatic questionnaires chosen were considered in the content validation process since they assessed the totality of the six core PLC identified in Russel and Grizzle's (2008) study. Related to the questionnaires chosen about attitudes and reactions in speaking situations for AWS, all published questionnaires available to assess attitudes and feelings about stuttering on adults were reviewed.

The definitions proposed by Russell and Grizzle (2008) were used to develop tables of contents to identify the specific items that probe each PLC domain. Table 3-4 and Table 3-5 were constructed for the eleven questionnaires chosen, which revealed examples of items that probe each PLC in a particular assessment tool.

Table 3-4. Items that probe each PLC in the Pragmatic Profile for adults and in the Pragmatic Profile (from CELF-5).

Instrument	NVC	DAE	SCF	RGG	V	C
Pragmatic Profile for adults	<u>Attention directing to self:</u> "If (name) is busy with something, and you want to get his/her attention, how do you usually do it?" <u>Attention directing to events, objects, other people:</u> "If you and (name) are out somewhere and you see something interesting, how does he/she point it out to him/her?"					

Table 3-4 (cont). Items that probe each PLC in the Pragmatic Profile for adults and in the Pragmatic Profile (from CELF-5).

Instrument	NVC	DAE	SCF	RGG	V	C
Pragmatic Profile (from the Clinical Evaluation on Language Fundamentals – CELF-5)	- Maintaining eye contact/gaze - Using strategies for getting attention - Knowing how someone is feeling based on nonverbal cues - Read and interpret the following nonverbal messages accurately: "facial cues/expressions" "making/responding to greetings to/from others) "making/responding to farewells to/from others" "beginning/ending conversations" "tone of voice" - Appropriate use of the following nonverbal support: "facial cues/expressions" "body language/gestures" "express messages by using gestures or facial expressions" "uses gestures and/or facial expressions according to the situation" "adjusts body distance /sits/stands) according to the situation" "presents matching gestures/facial expressions and verbal messages"		Appropriate use of the following nonverbal support: "voice intonation"	Making/responding to greetings to/from others Using strategies for responding to interruptions and interrupting others		

NVC- nonverbal communication; DAE- discourse attentiveness and empathy; SCF- speech characteristics and fluency; RGG- rituals, greetings, and goodbyes; V- vocabulary; C- comprehensibility.

Table 3-4 (cont). Items that probe each PLC in the Pragmatic Profile for adults and in the Pragmatic Profile (from CELF-5).

Instrument	CTT	TCM	R	SG	SA	IV
Pragmatic Profile for adults	Maintaining an Interaction or Conversation: "When you are chatting with someone, are there things you notice yourself doing which interfere with the flow of the conversation?"	Initiation: "When you want to start up a conversation with someone, what do you generally do?" Maintaining and Interaction or Conversation: "When (name) is chatting with you, how does the conversation flow?" Joining a conversation: "If you want to join in a conversation others are having, how do you go about it?"	Requests for assistance: "If you need help with something you are doing, how do you usually let (name) know?" Requests for information: "If you need to find out what is planned, for example for the day or for the weekend, how do you go about it?" Gaining attention: "when you haven't understood something someone has said to you, how do you let them know?" Conversational Repair: "When someone asks you to clarify something you have said that they haven't understood, what do you usually do?"		Rejecting "If you are offered something to eat or drink that you don't want, what do you usually do?" Giving information "If something happens that (name) isn't aware of and you tell him/her about it, how do you go about it? (for example, if someone visited or something got broken)" "if you are feeling unwell or uncomfortable, how do you let others know?"	Sociolinguistic Awareness: "Do you sometimes change your way of talking for particular people or situations? What changes do you make?" Person: "Are there people you like to be with or to talk to more than others?" Place: "where are most likely to be relaxed and communicative?" Topic: "What do you like to talk about most?" Situations Causing particular difficulty: "Are there situations involving speaking or communicating that cause you particular difficulty or anxiety? What are these?"

CTT- conversational turn-taking; TCM – topic control and maintenance; R – requests; SA – speech acts; SG – syntax/grammar; IV –interlocutor variability

Table 3-4 (cont). Items that probe each PLC in the Pragmatic Profile for adults and in the Pragmatic Profile (from CELF-5).

Instrument	CTT	TCM	R	SG	SA	IV
Pragmatic Profile (from the Clinical Evaluation on Language Fundamentals – CELF-5)	Observing turn-taking rules in the classroom or in social interactions	Beginning/ending conversations (face-to-face, phone, etc.) Introducing appropriate topics of conversation Maintaining topics using typical responses Making relevant contributions to a topic during cconversation Joining or leaving an ongoing communicative interaction	Asking for/responding to requests for clarification during conversation		Several specific speech acts, such as: - Giving (e.g., information) - Asking (e.g., for permission) - Offering (e.g., help) - Responding (e.g., to advice) - Agreeing - Disagreeing - Accepting - Rejecting - Apologize	Adjusting/modifying language based on the communication situation (communication partners, topic, place) Participating/interacting in structured group activities Participating/interacting in unstructured group activities

CTT- conversational turn-taking; TCM – topic control and maintenance; R – requests; SA – speech acts; SG – syntax/grammar; IV –interlocutor variability

Table 3-4 (cont). Items that probe each PLC in the Pragmatic Profile for adults and in the Pragmatic Profile (from CELF-5).

Instrument	NDI	TMEL	N	NLLIP	GP
Pragmatic Profile (from the Clinical Evaluation on Language Fundamentals – CELF-5)	Giving/asking for directions Starting/responding to verbal and nonverbal negotiations	Responding to teasing, anger, failure, disappointment Offering/responding to expressions of affection, appreciation	Telling/understanding jokes/stories that are related to the situation Showing sense of humour during communication situations		Avoiding use of repetitive/redundant information
Pragmatic Profile for adults (PP)	<u>Response to conflicting views</u> "If you feel strongly about something that should happen and others have a different point of view, how do you generally react?"	<u>Expression of emotion (pleasure):</u> "If you are really pleased about something how do you let people know?" <u>Expression of emotion (upset):</u> "If you are hurt or upset about something, how do you show it?" <u>Expression of emotion (self-assertion)</u> "If someone offers to help you to do something but you want to do it yourself, how do you let them know?"	<u>Narrative: Telling Stories and Jokes</u> "If you want to tell someone a story or a joke, what usually happens?"	<u>Response to Nonliteral Language:</u> "If someone says something in a roundabout way, such as using an expression like 'That's a bit of a dog's breakfast', do you ever have difficulties in understanding what is intended? How do you respond?" Responding to hints: "If (name) hints at something, rather than saying it directly, how do you generally respond?" <u>Presuppositions and Shared Knowledge:</u> "When you are telling (name) about something he/she doesn't know about, how clearly can you put him/her in the picture?"	<u>Presuppositions and Shared Knowledge:</u> "When you are telling (name) about something he/she doesn't know about, how clearly can you put him/her in the picture?"

NDI- negotiations, directions, and instructions; TMEL – theory of mind and emotion language; N – narrative; NLLIP – nonliteral language, use of indirection, and presupposition; GP – Gricean principles.

Table 3-5. Items that probe each PLC in the reactions/attitudes questionnaires for AWS.

Instrument	NVC	DAE	SCF	RGG	V	C
Stutterer's Self-rating of reactions to speech situations				"Saying hello to a friend passing by"		
Perceptions of Stuttering Inventory	"Using gestures as a substitute for speaking (e.g., nodding your head instead of saying "yes" or smiling to acknowledge a greeting)" "Acting in a manner intended to keep you out of a conversation or discussion (e.g., being a good listener, pretending not to hear what was said, acting bored, or pretending to be in deep thought")		"Making the pitch of your voice higher or lower when you expect to get "stuck" on words" "Making your voice louder or softer when stuttering is expected" "Speaking so that no word or sound stands out (e.g., speaking in a singsong voice or in a monotone")			
Modified Erickson Scale of Communication Attitudes	"I find it very easy to look at my audience while speaking to a group"		"I find it easy to keep control of my voice when I speak" "My speaking voice is rather pleasant and easy to listen to"			
Speech situation Checklist				"Saying hello"		
A self-efficacy scale for adults stutterers						

NVC- nonverbal communication; DAE- discourse attentiveness and empathy; SCF- speech characteristics and fluency; RGG- rituals, greetings, and goodbyes; V- vocabulary; C- comprehensibility.

Table 3-5 (cont). Items that probe each PLC in the reactions/attitudes questionnaires for AWS.

Instrument	NVC	DAE	SCF	RGG	V	C
Pragmatic Stuttering Intervention for Adolescents and Adults				"Greets others"		
The Wright & Ayre Stuttering Self-Rating Profile	"Loss of eye contact"					
Clinical Use of Self- reports						
The Overall Assessment of the Speaker's Experience of Stuttering	"How often do you break eye contact or avoid looking at your listener"					

NVC- nonverbal communication; DAE- discourse attentiveness and empathy; SCF- speech characteristics and fluency; RGG- rituals, greetings, and goodbyes; V- vocabulary; C- comprehensibility.

Table 3-5 (cont). Items that probe each PLC in the reactions/attitudes questionnaires for AWS.

Instrument	CTT	TCM	R	SG	SA	IV
Stutterer's Self-rating of reactions to speech situations		"Introducing myself (face to face)"			"Ordering in a restaurant" "Asking the instructor a question in class" "Asking girl for date" "Asking a secretary if I can see the employer" "Asking for a job" "Asking a desk for book or card to be filled out, etc." "Answering roll call in class"	Several items are related to communication with different interlocutors (e.g., mother, father, instructor, a good friend or a stranger)
Perceptions of Stuttering Inventory		"Avoiding introducing yourself, giving your name, or making introductions"			"Avoiding asking for information (e.g., asking for directions or inquiring about a train schedule)	"Avoiding talking to people in authority (e.g., a teacher, employer, or clergyman"
Modified Erickson Scale of Communication Attitudes					"I dislike introducing one person to another" "I often ask questions in group discussions"	"A person who is my teacher or my boss is hard to talk to" "I do not mind speaking in front of a group"

CTT- conversational turn-taking; TCM – topic control and maintenance; R – requests; SA – speech acts; SG – syntax/grammar; IV –interlocutor variability

Table 3-5 (cont). Items that probe each PLC in the reactions/attitudes questionnaires for AWS.

Instrument	CTT	TCM	R	SG	SA	IV
Speech situation Checklist		"Making introductions"			"Answering a specific question" "Asking for information" "Asking the teacher a question" "Answering questions about speech" "Asking if someone is at home" "Asking a gas station attendance for a specific amount of gas"	Several items are related to interlocutor variability (e.g., talking with a young child, an animal or a close friend) or situation (e.g., group)
A self-efficacy scale for adults stutterers		"Approach your boss and initiate a conversation at work" "Initiate a conversation with a stranger of the opposite sex at a party" "Introduce yourself to a stranger" "Introduce yourself to a group of strangers" "Initiate a conversation wih the person sitting next to you on an airplane"	"Requests help in an uncrowded department store" "Request help in a crowded department store when all the salespeople seem busy"		"Order food at McDonald's when there are no other customers" "Answer questions during a group discussion" "Ask questions during a group discussion" "Order food from your car through a speaker at McDonald's" "Order a drink from a bartender at a noisy, crowded bar" "Ask for directions over the phone" "Order food in a restaurant when the waitress is obviously in a hurry" "order a pizza over the phone" "Order exactly what you want in a restaurant even though you might stutter on the words"	Several items are related to interlocutor variability (e.g., family member, close friend or physician) or situation (e.g., telephone, noisy bar)

CTT- conversational turn-taking; TCM – topic control and maintenance; R – requests; SA – speech acts; SG – syntax/grammar; IV –interlocutor variability

Table 3-5 (cont). Items that probe each PLC in the reactions/attitudes questionnaires for AWS.

Instrument	CTT	TCM	R	SG	SA	IV
Pragmatic Stuttering Intervention for Adolescents and Adults	"Take turns appropriately during conversation"	"Initiates conversations" "Maintains a topic over a series of utterances"	"Requests assistance when help is needed" "Offers suggestions to others"		"Asks questions" "Gives compete answers when asked for information"	"Participates verbally during group activities"
The Wright & Ayre Stuttering Self-Rating Profile						
Clinical Use of Self-reports						All items are rated concerning different interlocutors (close friend, parent, stranger, authoritative figure) and a specific communicative situation (telephone)
The Overall Assessment of the Speaker's Experience of Stuttering		"How difficult is for you to communicate in the following general situations: Initiating conversations with other people (e.g., introducing yourself)" "How difficult is for you to communicate in the following general situations: Continuing to speak regardless of how your listener responds to you"			"How difficult is for you to communicate in the following social situations: Asking for information (e.g., asking for directions or other people's opinions" "How difficult is for you to communicate in the following social situations: Ordering food in a restaurant" "How difficult is for you to communicate in the following general situations: Ordering food at a drive-thru"	Several items are related to interlocutor variability (e.g., friends, strangers, large group, small group) or situation (e.g., telephone, social event or a restaurant)

CTT- conversational turn-taking; TCM – topic control and maintenance; R – requests; SA – speech acts; SG – syntax/grammar; IV –interlocutor variability

Table 3-5 (cont). Items that probe each PLC in the reactions/attitudes questionnaires for AWS.

Instrument	NDI	TMEL	N	NLLIP	GP
Stutterer's Self-rating of reactions to speech situations	"Giving directions to a stranger"		"Telling a funny story with one stranger in a crowd"		
Perceptions of Stuttering Inventory	-	-	-	-	-
Modified Erickson Scale of Communication Attitudes	-	-	-	-	-
Speech situation Checklist	"Arguing with parents" "Trying to get across your own point of view" "Giving directions" "Refuting a criticism"	-	-	-	-
A self-efficacy scale for adults stutterers	-	-	"Tell a joke in front of five people"	-	-
Pragmatic Stuttering Intervention for Adolescents and Adults	"Volunteers information during discussions" "Express point of view"	"Shares feelings" "Expresses feelings of disagreement"	"Describes events accurately and in a sufficient detail" "Describes personal experiences" "Express needs" "Expresses viewpoints"	-	-
The Wright & Ayre Stuttering Self-Rating Profile	-	-	-	-	-

NDI- negotiations, directions, and instructions; TMEL – theory of mind and emotion language; N – narrative; NLLIP – nonliteral language, use of indirection, and presupposition; GP – Gricean principles.

Table 3-5 (cont). Items that probe each PLC in the reactions/attitudes questionnaires for AWS.

Instrument	NDI	TMEL	N	NLLIP	GP
Clinical Use of Self-reports	-	-	-	-	-
The Overall Assessment of the Speaker's Experience of Stuttering	"How difficult is for you to communicate in the following general situations: Standing up for yourself verbally (e.g., defending your opinion, challenging someone who cuts in line in front of you)	-	"How difficult is for you to communicate in the following social situations: Telling stories or jokes"	-	-

NDI- negotiations, directions, and instructions; TMEL – theory of mind and emotion language; N – narrative; NLLIP – nonliteral language, use of indirection, and presupposition; GP – Gricean principles.

Stage 1 of the Content Validity analysis revealed that none of the eleven instruments assessed the seventeen PLC. The instruments reviewed contained items that assessed up to twelve PLC domains. During the construction of tables of contents, it was observed that the items used by the instruments were specific to the assessment of a PLC in a situation, person, place or conversation subject. The analysis of Table 3-4 and Table 3-5 revealed different contents that should be part of the new questionnaire, to be representative of the domain and to allow the assessment of PLC. It is important to note that some items do not cover the entire content of a specific PLC.

The analysis of the eleven instruments revealed items that were used to assess several PLC concerning the difficulty level, the engagement or the way used in a conversation to convey that pragmatic competency. Concerning the PLC *Nonverbal communication*, items related to eye contact, use of gestures and attention directing should be added; for the PLC *Speech characteristics and fluency* items related to pitch, and control of voice must be developed; responding and making different greetings (e.g., “hello”) is an item to be develop concerning PLC *Rituals, greeting, and goodbyes*. Appropriateness of turn-taking rules and to maintain the flow of conversation should be part of the new questionnaire to assess the PLC *Conversational turn-taking*. *Topic control and maintenance* referred to the appropriate strategies used to initiate (e.g., introducing), maintain (e.g., through nonverbal communication or appropriate/relevant contributions to the topic), joining and leaving a conversation. Concerning the PLC *Requests*, items to evaluate requests for information, assistance, directions or conversational repair/clarification (both asking and responding to) are important. Several items related to different *Speech Acts* are part of the pragmatic questionnaires, to assess asking, rejecting, giving information, ordering or answer in a social appropriate manner. The PLC *Interlocutor variability* are expressed in several items, mainly in the attitudes questionnaires specifically design to assess AWS. To asses this PLC, the new questionnaire should add items to express different types of familiarity (e.g., a family member or a stranger), interlocutor number (communication one-on-one or to a group), type of communication (e.g., through phone or face-to-face) or communication topic (e.g., work or feelings). *Negotiations, direction, and instructions* should be assessed in the questionnaires through items related to arguing, defending a point of view, response to conflicting views, refuting critics and give directions. Expression of different types of emotions (e.g., pleasure, disagreement, upset) should also be part on items of the new questionnaire in the assessment of *Theory of mind and emotion language*. Related to the PLC *Narrative* items to assess the description of events/stories and jokes appropriately must be develop. The response to nonliteral language should be also part of the

questionnaire to assess the PLC *Nonliteral language, use of indirection, and presupposition*. The PLCs *Discourse Attentiveness and empathy*, *Vocabulary*, *Comprehensibility*, *Syntax and Grammar* and *Gricean Principles* were not probe by item of the assessed questionnaires. To assess the totality of the pragmatic competencies, the items were generated based on the results draws from the content tables (reported above) and with the definitions proposed by Russel and Grizzle (2008).

The pre-final version of the *ALUSCA* questionnaire included an introduction and two parts. The first part (27 items) includes several questions regarding variables that characterize a conversation with an interlocutor. In part I the respondent classified the level of ease in communicate with a specific interlocutor, in a certain situation, with a certain topic of conversation and in a particular location. Part I presented four variables (People, location, topic of conversation and communicative situation); the first variable (People) present 9 characteristics related to the variable (e.g., communicate with a group of known people or communicate with an authority figure), the second variable (Location) present 5 characteristics (e.g., communicate at your job or communicate at home), the third variable (Topic of conversation) contain 6 characteristics (e.g., communicate about your work or communicate about your feelings) and the fourth variable (Communicative situation) include 7 characteristics (e.g., communicate on the telephone or communicate under time constrain).

On part 2 of the questionnaire, the subject classifies the level of ease in performing PLC on a situation classified with level of ease 1 or 2 (i.e., were considered hard) on part 1. In part 2 (64 items) the subject classifies the level of ease in each PLC item in the most difficult situation that is identified in Part 1 and also asked to indicate if the level of ease chosen was due to stuttering. Part 2 present a total of 64 items, distributed by 16 PLC (i.e., Nonverbal communication, Discourse attentiveness and empathy, Speech and voice, Rituals, greetings and goodbyes, Vocabulary, Comprehensibility, Turn-taking, Topic control and maintenance, Syntax/grammar, Requests, Use direct and indirect speech acts, Negotiations, directions and Procedures, Emotional language, Narrative, Nonliteral language and use of presupposition, Respect by Gricean principles) organized in three categories (i.e., Precursors, Basic exchanges and Extended literal and nonliteral language). Precursors had 26 items; Basic exchanges had 24 items and nonliteral language had 14 items. For example, a subject classifies a situation characterized by talking with an unknown person about job on the telephone as the most difficult on part I of the ALUSCA questionnaire; considering a situation with those characteristics, the subject characterizes the level of ease in use the pragmatic competencies described. Both parts used a Likert-

type scale from 1 (is never easy) to 5 (is always easy), with a neutral value (3). The total score was the average of the sum of the level of ease chosen in all ALUSCA part 2 items.

Table 3-6 presents the seven items that compose one of *ALUSCA*'s PLC (Requests) from part 2. The items that compose each PLC were preceded by an explanation of the language competency (rationale) and the procedures to assess the level of ease on each competency.

Table 3-6. Example of the items that compose the PLC Requests (part 2 of the ALUSCA questionnaire).

2.4 – Requests

Rationale: during a conversation with someone, we can make several requests to the person/people with whom we are talking.

Procedures: focus on the requests that you make when communicating with others. During a conversation with the **characteristics assessed in Part I of the questionnaire as *Difficult* rate the degree of ease in:**

	Never easy	Sometimes easy	Neutral	Often easy	Always easy	NA
1- Request for **help** (e.g., to solve a problem)?						
2- Ask for **something** (e.g., food in a restaurant)?						
3- Ask for **information** (e.g., where to go to pay something)?						
4- Ask for **directions** (e.g., on a road ask for directions to a certain place)?						
5- **Excuse** (e.g., say "sorry" due to an error that you have committed)?						
6- Ask for **clarification** (e.g., when you have a doubt)?						
7- **Answer** appropriately to clarification requests of the person you are talking to (e.g., using easier words)?						

The thinking aloud method (Goldman, 1971) used during the focus groups revealed that some of the *ALUSCA* items were complex and confusing. Specifically, AWS and controls suggested that: Difficult words should be avoided (e.g., pragmatic or skills); the most important parts of each questions should be highlighted, showing where the subject needs to focus; examples should be added in all PLC items to clarify what they are; definitions of specific words (e.g., sarcasm, metaphor or comprehensibility) should be included. The ambiguities revealed by the focus groups were eliminated and several items were rephrased, reworded and examples were added.

Based on the analysis and the opinions collected through the thinking aloud method on the expert panel meeting (stage 2 of the content validity analysis), the instructions were simplified and some *ALUSCA* items were modified. Particularly, examples were included in the introduction of part 2, examples of PLC speech acts were rephrased, items were divided and superfluous information was deleted. A question specifically related to stuttering (part 1 of the pre-final version) was deleted in the final version, based on experts' opinions that AWS may be unaware if the ease level chosen was due to stuttering.

The Bland and Altman (1986) modified analysis (Jesus *et al*. 2015, pp. 4–5) revealed that the experts were in disagreement in five questions of the instrument review questionnaire (see Figure 3-1) related to the PLC "Syntax and Grammar" (clarity and simplicity of instructions and clarity, simplicity and precision of the items), one question related to the PLC "Nonliteral language, use of indirection and emotion language" (relevance of the items) and three questions related to the PLC "Gricean principles" (relevance, clarity and simplicity of the items). Based on these results, simplicity and clarity were ensured for the items where disagreement was observed and the pre-final version of *ALUSCA* was developed.

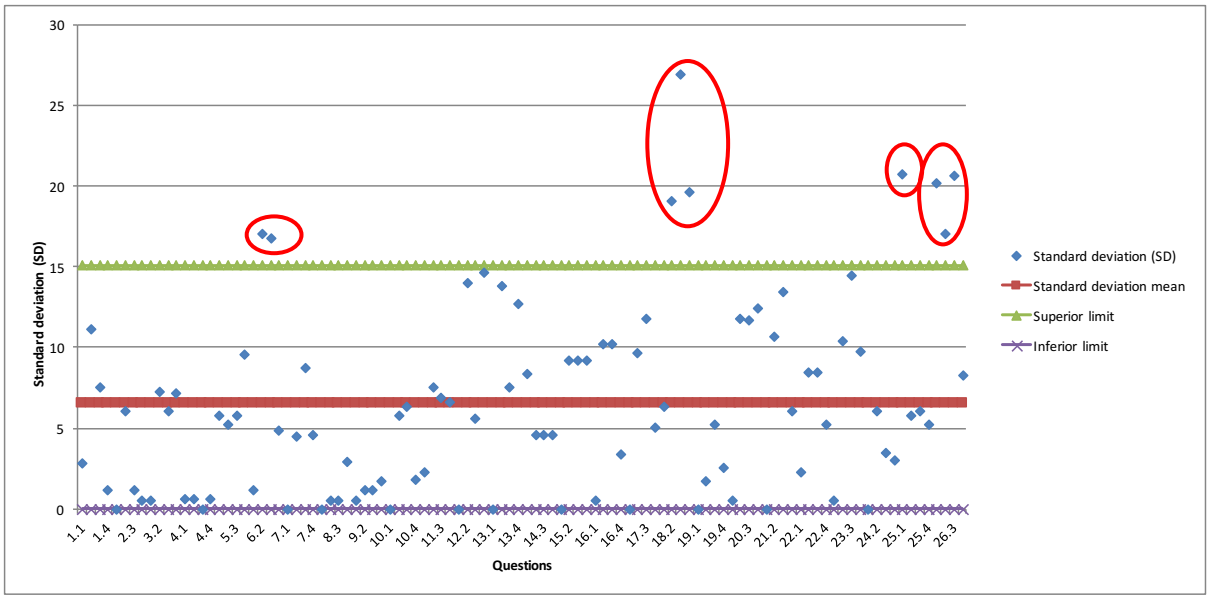


Figure 3-1 Modified Bland and Altman (1986) analysis (Jesus et al., 2015, pp. 4–5). Disagreements between experts are signalled with an ellipsis.

3.4.2 Pilot study

During the pilot study, five AWS and five controls (mean age: 27 years old; sex ratio of 1 female to 4 male) fulfil the paper and pencil questionnaire.

Item analysis revealed that only 0.72% of pairwise correlations were >0.9 on Pearson product moment correlation (i.e., items were quite similar). Descriptive statistics of the pilot study revealed that matched controls presented higher mean scores than AWS in all three categories and in the final score (see Table 3-7). Respondents did not reveal difficulties in filling in the *ALUSCA* questionnaire.

The final version of the *ALUSCA* questionnaire, developed after the pilot study, contained two parts, similarly to the pre-final version. Procedures and Likert-type scale were retained.

Table 3-7. Descriptive statistics of the pilot study.

	Mean ± SD		Minimum		Maximum	
	AWS	Controls	AWS	Controls	AWS	Controls
Category 1: Precursors/enablers	3.0 ± 0.8	3.7 ± 0.6	1.63	2.77	3.58	4.16
Category 2: Basic exchanges	3.0 ± 1.0	3.6 ± 0.8	1.18	2.65	3.66	4.31
Category 3: Extended literal discourse	2.9 ± 0.10	3.6 ±0.8	1.17	2.27	3.68	4.24
Total	3.0 ± 0.9	3.6 ± 0.7	1.32	2.56	3.60	4.24

4.3 Reliability analysis

Pilot study analysis revealed a Cronbach's alpha of 0.979 for the total score, 0.920 for category 1, 0.959 for category 2 and 0.947 for category 3. Data indicated that the *ALUSCA* presented good internal consistency (Bland & Altman, 1997). Test-retest analysis revealed a large correlation (r=0.947, p=0.000), which constitutes evidence of temporal stability of the questionnaire.

Data obtained for the construct validity process were also used to calculate internal consistency. Data revealed a Cronbach alpha of 0.973 for the *ALUSCA* total score, 0.936 for category 1, 0.936 for category 2 and 0.902 for category 3. Cronbach alpha values indicated a good internal consistency of the *ALUSCA* (Bland & Altman, 1997).

4.4 Construct validity analysis

To assess the hypothesis related to pragmatic difficulties in some AWS, 28 AWS and 28 controls (matched by age and gender) were recruited through the Portuguese Association of Stuttering (*Associação Portuguesa de Gagos*) and through local SLT with experience in stuttering intervention. The mean age was 32.35 years for both groups (range from 20-59 years old) and sex ratio of 9 females to 19 males. Table 3-8 presents descriptive statists for each category and for the total score of both groups.

Table 3-8. Descriptive statistics of the construct validity evidence study.

	Mean ± SD		Minimum		Maximum	
	AWS	Controls	AWS	Controls	AWS	Controls
Category 1: Precursors	3.1±0.6	3.5±0.7	1.63	2.77	3.58	4.16
Category 2: Basic exchanges	3.4±0.7	3.7±0.7	1.18	2.65	3.66	4.31
Category 3: Extended literal discourse	2.9±0.6	3.3±0.8	1.17	2.27	3.68	4.24
Total	3.1±0.6	3.5±0.7	1.33	2.56	3.60	4.24

AWS presented lower mean score in all categories and in the total score, when compared with controls. Independent sample *t*-test revealed statistical significant differences between AWS and controls on category 1 ($t(54)=-2.236$, $p=0.03$), in category 3 ($t(54)=-1.993$, $p=0.05$), and in the final score ($t(54)=-2.150$, $p=0.04$). In category 2, the *t*-test results ($t(54)=-1.604$, $p= 0.115$) revealed a non-significant difference between the two groups related to basic communication exchanges. An outlier was found in the boxplot of category 2 score. The outlier was deleted and independent sample *t*-test was performed without the outlier for category 2 and the final score. Values obtained were $t=(53)= -2.581$, $p=0.013$, for category 1, $t(53)=-2.222$, $p=0.031$, for category 2, $t(53)=-2.361$, $p=0.022$, for category 3 and $t(53)=-2.679$, $p=0.010$ for the final score, meaning that there were statically significant differences between AWS and controls in the second category from *ALUSCA* part .

PCA performed with Part 1 variables and Part 2 categories scores yielded two-factor solutions for both groups, with an explained variance of 78.5% for AWS and 81.5% for controls. The rotated solutions (varimax rotation) revealed that for AWS, the first component comprised the scores from all categories of part 2 and Communicative Situation score (from part 1), and the second component comprised People, Place and Subject scores from part 1. For controls, the first component comprised all part 2 scores and the second component comprised all part 1 scores. Each component comprises variables that are highly correlated and have a similar profile placed in the same component (Hair Jr et al., 2009). For controls, factor analysis revealed a solution that supports the underlying construct of the *ALUSCA* items. For AWS, factor analysis revealed the importance of speaking situations for the scores attributed in the 3 categories of part 2. Data revealed that the variables are correlated, i.e., a similar profile related to the attribution of scores is observed in the

Communicative Situation variable and in the 3 categories of *ALUSCA* part 2.

3.5 Discussion

3.5.1 Content validity analysis

The two-stage process (developmental stage and judgment stage) results allow us to say that the items of the *ALUSCA* questionnaire are representative and relevant for the content that the questionnaire intends to measure (Lynn, 1986). The exhaustive collection and analysis of questionnaires (specific to assess pragmatics and also attitudes of AWS) in terms of items that assessed the 17 PLC, allowed the authors of this paper to include a broad content related to the assessment of pragmatic competencies.

The consultation of a first expert panel (experts within the target population of the *ALUSCA* questionnaire) allowed the collection of relevant information, beyond the suggestions of the professional expert (Schilling et al., 2007).

Results obtained from the Bland and Altman (1986) modified analysis (Jesus *et al*., 2015 pp. 4–5) used with professional experts guided the modifications in the pre-final version of the *ALUSCA* in order to achieve more clarity, simplicity and precision of the non-agreed items. Using this method, the authors ensured that items were modified according to expert's opinion, to develop a final version of the instrument that was valid in content for all judges.

The use of a heterogeneous expert panel to assess content validity allowed the construction of an instrument that is more valid to the target population and to the professional audience that could use the instrument (Schilling *et al*., 2007).

3.5.2 Pilot study

Results from the pilot study were used to calculate pairwise correlations (item analysis), to observe differences between AWS and controls and to analyse difficulties referred by the respondents.

Item analysis results revealed that 0.72% of the pairwise correlations were quite similar. However, due to the framework used and the objectives of the *ALUSCA* (i.e., be exhaustive in the domains studied), all the items were retained in the final version.

The results indicated differences between the two groups at the level of ease in using PLC

domains (AWS presented lower mean results when compared with controls). These preliminary results suggest that pragmatic demands influenced the communication process of AWS.

The absence of difficulties expressed by the respondents lead the authors to retain all the items in the final version and conclude that the questionnaire was feasible to be used in a larger population (Leon et al., 2011).

3.5.3 Reliability analysis

Cronbach's alpha results indicated that the ALUSCA items presented good internal consistency, meaning that items are homogenous and that jointly measured the same underlying construct, as internal consistency value is higher than the cut-off threshold of 0.7 (Bland & Altman, 1997; Field, 2013; Kline, 2004).

Test-retest reliability results revealed strong correlation, provide assurance that the *ALUSCA* "measures the outcome the same way, in a stable client, each time is used" (Vaz, Falkmer, Passmore, Parsons, & Andreou, 2013, p. e73990). Strong correlation values on test-retest analysis means better reproducibility and implies better precision of measurements and better tracking of variations in measurements in research or clinical settings (Hopkins, 2000).

3.5.4 Construct validity analysis

Results from the validity analysis with a larger group testing the hypothesis that, in mean, AWS presented a significantly different level of ease in performing pragmatic competencies when compare with controls. Statistically significant differences were also previously revealed by instruments (Andrews & Cutler, 1974; Brutten & Shoemaker, 1974; Johnson et al., 1952; Ornstein & Manning, 1985; Riley et al., 2004; Tanner et al., 1999; Woolf, 1967; Wright & Ayre, 2000; Yaruss & Quesal, 2006) that were used to assess attitudes related to communication in AWS based on items that probe PLC domains (Russell & Grizzle, 2008). Specifically, those instruments also presented statically significant differences between AWS and controls, related to difficulties in specific speech situations (e.g., Brutten & Shoemaker, 1974) and attitudes related to communication in general (e.g., Andrews & Cutler, 1974).

Pragmatics continues to develop throughout the life span (Dewart & Summers, 2003). In

early ages, children learn to communicate appropriately in social situations, through interaction with different interlocutors (Hymes, 1967). Related to stuttering, pragmatics influences disfluency, as disfluencies increase when the child has to face specific pragmatic conditions, e.g., when interrupting, directing others in activities or responding to requests (Davis, 1940; Meyers & Freeman, 1985a, 1985b). Pragmatic development that occurs in order to master more complex pragmatic skills can create additional demands and challenges in a developing child, leading to competition for cerebral resources for language acquisition and fluent speech production. The competition for cerebral resources in an immature brain leaves fewer remaining resources available for fluent speech production, which can lead to normal disfluencies in speech production (Guitar, 2014). As several pragmatic aspects of language are developed after complex grammatical constructions (Cummings, 2014), the development of pragmatics occurs also in adulthood (Ninio & Snow, 1999). Related to AWS, Tanner, Belliveau and Seibert (1999) suggested that this population does not present a social (pragmatic) communication disorder (i.e., a diagnosis of communication disorder based on specific and defined criteria established by the American Psychiatric Association (2013)) but rather have difficulties in language use that come from the anxiety connected with communicative situations. Anxiety experienced by AWS "is secondary to stuttering and the result of having to cope with a serious communication problem" (Manning & Beck, 2013, p. 185). Alm (2014) showed in his review related to stuttering and anxiety, that many people with persistent stuttering developed high levels of anxiety associated within specific speech situations. Manning and Beck (2013) also showed that anxious personality disorder, which is characterised by social avoidance and hypersensitivity to potential criticism, is observed in AWS. Thus, we can argue that anxiety can lead to avoidance, that causes an important impact and negative consequences in several aspects of AWS live's (Beilby, Byrnes, & Yaruss, 2012). Cuthbert (2002) concurred that people with anxiety disorder avoid social, occupational and educational situations perceived as intimidating. The avoidance interferes with social interactions, relationships and with quality of life in general (Stein & Ken, 2000). The avoidance of speaking situations can be one of the arguments to explain differences in the level of ease reported by AWS. Avoidance of speaking situations are based on past experiences, cognitive/effective reactions or pragmatic demands, that can lead to loss of important opportunities to experiment and develop the use of language (Tanner et al., 1999) in a functional way, which reflects the effective use of pragmatic behaviours.

However, there would not be such a clear distinction between a social (pragmatic) communication disorder and pragmatic difficulties in AWS if the strengths and weaknesses

related to language use were identified (through the *ALUSCA* questionnaire, for example) and considered in an intervention plan.

The PCA revealed a two component solution from both groups. Related to AWS, the importance of speaking situations for the scores attributed in the three categories of part 2 is revealed in the results obtained. Cognitive reactions and, in particular social cognition, can lead to interferences in the ease level in performing speech situations, as "for persons who are concerned about stuttering it is likely that social situations often involve thoughts about possible scenarios, including what other may think if they stutter and alternative plans how to act" (Alm, 2014, p. 16). Manning (2010) also highlights the importance of speech situations, both for typical fluent speakers and for PWS. However, for those who stutter the variability is more noticeable based on fluency failures of past experiences that can cause reactions to fluency-disrupting stimuli, such as difficult speaking experiences). Thus, one can argue that the variability in stuttering symptoms observed in different speaking situations can influence the connection observed between level of ease in speaking situations and level of ease in PLC categories (as the present data from PCA indicated).

3.6 Conclusions

The present chapter reported the development and the determination of psychometric properties of the questionnaire Assessment of Language Use on Social Contexts for Adults (ALUSCA).

The development of the *ALUSCA* comprised a two-stage process to ensure that the "items (...) adequately measure a desired domain of content" (Grant & Davis, 1997, p. 269). Stage 1 comprised a domain identification through the analysis of pragmatic and stuttering related questionnaires and the development of *ALUSCA* items. Stage 2 included the judgment of the questionnaire items, which was carried by qualitative and quantitative analysis by focus groups and an expert panel. Content validity of *ALUSCA* was guaranteed.

The pilot study allowed item analysis and reliability analysis, showing that *ALUSCA* is a reliable instrument. The final version was obtained and a larger scale study with ALUSCA was considered to be feasible.

Evidences of construct validity, i.e., the establishment of the appropriateness of the inferences based on test results, were obtained through a hypothesized relationship method and factor analysis, as proposed by the American Educational Research Association (American Educational Research Association (AERA), 2014). It can be concluded that

ALUSCA presents evidences of construct validity, as shown through differences between AWS and controls in the hypothesized relationship method (statistical significant differences in the independent sample *t*-test) and also in the different response profiles revealed by the PCA for the two groups. The similar score profile evidenced by the three categories and the Speech Situation variable highlights the importance of pragmatic development through practice in speech situations (Tanner et al., 1999). Evidences found can lead to the conclusion that the theory supports the interpretation of test scores obtained (American Educational Research Association (AERA), 2014).

ALUSCA is a reliable and valid pragmatic instrument that was tested in AWS and controls; additionally, the tool could be also considered relevant as a procedure for adults with language problems or for those expressing difficulties with socialization.

Study limitations include the small sample size of the pilot study and the test-retest calculation that only have been performed with controls. Future research could include the assessment of avoidance in AWS, and correlating this with the level of ease in pragmatic competencies; the determination of impact ratings for the *ALUSCA* total scores was also important to analyse the influence of pragmatic on an AWS.

Chapter 4 - Public Opinion Survey of Human Attributes – Stuttering (*POSHA-S*)

4.1. Introduction

The Public Opinion Survey of Human Attributes – Stuttering (*POSHA-S*) a reliable and valid questionnaire, used internationally to measure knowledge and attitudes toward stuttering. The present chapter presents the translation and cross-cultural adaptation process of the *POSHA-S* to European Portuguese (EP) and also the results from a country-wide probability sampling collection of the Portuguese population to determine stuttering attitudes in an entire country.

4.2 Background

4.2.1 The Stuttering Stereotype and Attitudes toward Stuttering

Research studies conducted over several decades regarding attitudes toward PWS have shown that negative traits are attributed to them by the general public (Özdemir, St. Louis, & Topbaş, 2011a; St. Louis, Coskun, et al., 2005; Van Borsel, Verniers, & Bouvry, 1999). Specifically, people of different ages (Doody, Kalinowski, Armson, & Stuart, 1993; Evans, Healey, Kawai, & Rowland, 2008; Griffin & Leahy, 2007; Hartford & Leahy, 2007; Kirsch, 2006) from different professional groups (Crowe & Cooper, 1983; Crowe & Walton, 1981; Dorsey & Guenther, 2000; Lass et al., 1992, 1994; Ruscello, Lass, Schmitt, & Pannbacker, 1994; Yeakle & Cooper, 1986), and also SLTs (Cooper & Cooper, 1996; Lass, Ruscello, Pannbacker, Schmitt, & Everly-Myers, 1989; Yairi & Williams, 1970) and parents (Al-Khaledi et al., 2009; Crowe & Cooper, 1977) presented negative attitudes and perceptions toward stuttering and those who stutter. Anxiety, shyness, nervousness, introversion, unassertiveness are among those negative traits commonly attributed to PWS and referred to as the stuttering stereotype (MacKinnon, Hall, & MacIntyre, 2007; St. Louis & Roberts, 2010).

Stigma, regarded by Goffman (1963) as a manifestation of a “spoiled identity,” relates to the stuttering stereotype as the PWS is regarded as being defective in every aspect of life because of their “deviant labels others associate with a single characteristic that the person may possess” (Abdalla & St. Louis, 2012, p. 55). Stigma can come from others (i.e., public

stigma) or oneself (i.e., self-stigma). Public stigma refers to cognitive, affective and behavioural negative reactions and opinions of the public related to a group with a stigmatized condition (e.g., PWS), and may be noticeable in terms of stereotypes, exclusion or discrimination, with negative consequences for the stigmatized individuals (Boyle, 2013). Through the process of internalising those negative consequences, by a multidimensional process in progressive levels, PWS manifest self-stigma in the form of negative beliefs and negative emotional reactions (Boyle, 2013). These, in turn, lead to self-discrimination, lack of confidence or reduced self-esteem.

Stigma can be internalised among PWS; thus, research related to the different types of stigma (and stigma consequences) is needed to define the nature and impact related to PWS and to define therapeutic approaches to the public (e.g., public awareness campaigns) or to the person (Boyle, 2013; St. Louis, 2012c). With the objectives of assessing attitudes toward stuttering, to compare them in different countries/populations and to develop a science to improving such attitudes, the International Project on Attitudes Toward Human Attributes (IPATHA) was created (St. Louis, 2011). The IPATHA purposes required a standard survey instrument that could be translated and used internationally – the Public Opinion Survey on Human Attributes (*POSHA-S*) (St Louis 2011, 2012a, 2012b, 2015a, 2015b) and has been used since 2001 to collect information regarding attitudes toward stuttering in 42 different countries and in 26 different languages (circa March, 2016). Excluding experimental versions of the instrument and samples composed entirely of individuals who stutter, SLTs, or SLT students, 11383 respondents from 141 different samples comprise the public database. Because the database samples vary, based on such selection variables as country, age, level of education, profession, socio-economic status (SES), or languages known, the median of the 141 sample means is taken as the "average" for each POSHA–S rating. The rationale is that, compared to the mean of all 141 sample means, the median value minimizes influence of "outlier" samples, i.e., samples with extremely positive (high) or extremely negative (low) attitudes. Currently, the median of the "Overall Stuttering Score" (OSS, described below) is 17 for public samples.

4.2.2 European *POSHA-S* Studies

The attitudes toward PWS have been studied in different countries using the *POSHA-S* questionnaire. Most of the studies have employed various methods of convenience sampling. These involved investigators asking friends, acquaintances, family members, colleagues or students to fill out the questionnaire and to distribute to others. No systematic

research has been reported that has sought to collect a representative sample that would allow generalisation of findings to an entire country. Following is a review of European selected studies that measured attitudes using *POSHA-S* in samples across different regions of a country or sampled comprehensively enough to generate an impression of attitudes in a specific country. The majority of the studies collected sample through convenience sampling and the few that used probability sampling are highlighted.

In Italy (Tomaiuloi, Del Gado, Capparelli, & St. Louis, 2013), data were collected from 300 respondents from three different country regions using convenience sampling. More than half of the respondents reported not knowing anyone who stutters. In general, *POSHA–S Overall Stuttering Score*, the subscores and the various components were scored below the *POSHA-S* database median, indicating that the Italian attitudes were less accurate and less positive than average.

Four research studies (Przepiorka, Blachnio, St. Louis, & Woźniak, 2013; St. Louis, Przepiorka, et al., 2014; Węsierska, Węsierska, & St. Louis, 2013; Węsierska & St. Louis, 2014) have been reported on attitudes toward PWS in the Polish population. They include 268 respondents from the general public, 188 SLTs, 403 SLT students and 205 students from other majors, all collected via convenience sampling. Generally, the polish attitudes on different samples were quite consistent and differences (e.g., SLT presented higher scores on *POSHA-S* than SLT students) were considered minor and non significant (Węsierska et al., 2013).

The overall results suggest that the Polish samples manifested similar attitudes of those presented in other populations studied with *POSHA-S*, and, like other samples, with evidence of stigma and social exclusion toward PWS.

Public attitudes in Turkey were investigated in four different studies (Aydın, 2008; Çoşkun, 2006; Özdemir et al., 2011a; Özdemir, St. Louis, & Topbaş, 2011b; St. Louis et al., 2011; St. Louis, Andrade, Georgieva, & Troudt, 2005). One of the studies (Özdemir et al., 2011a) used a probability sample scheme to compare samples of sixth graders with the samples of parents, grandparents or adult relatives and adult neighbours and the other three studies used convenience samples. The convenience samples (Aydın, 2008; Çoşkun, 2006) held more positive beliefs and less positive self-reactions to stuttering and PWS when compared with the attitudes expressed on *POSHA-S* by probability samples. The comparison with the *POSHA-S* database revealed that the medians of the Turkish samples were lower than the median.

In Ireland (Daly & Leahy, 2014), a probability sampling scheme was used to collect 37 respondents. The dominant public attitude was a neutral impression regarding stuttering, although negative impressions related to PWS were also presented in this population (37% were of the opinion that PWS are nervous or excitable, 51% believed that PWS are shy and fearful, and 37% would feel pity for the person). The majority of the Irish population presented little or some knowledge about stuttering and some confusion about the cause of stuttering was observed. The comparison with the *POSHA-S* database revealed that the components, subscores and the *Overall Stuttering Score* of the Irish sample were scored above the median of the *POSHA-S* archive, with the exception of those related to the personality of PWS and with the knowledge about the etiology of stuttering (Daly & Leahy, 2014). This means that the majority of the Irish respondents held more positive than average attitudes.

Denmark, Bulgaria, Russia, Norway, United Kingdom, Germany, Bosnia-Herzegovina, Croatia, Spain, Malta, Belgium and The Netherlands also collected data (not published) related to public attitudes using *POSHA-S* questionnaire through convenience samples. In general, the scores obtained on the research studies from European countries were more accurate/more positive than the total score of the total samples from the *POSHA-S* archive (St. Louis, 2011).

Subsequent to data collection in the current study but highly relevant to it, St. Louis, Sønsterud, et al. (2016) reported a study of European country attitudes that involved combining data from eight different studies of public attitudes. It reveals that attitudes varied across countries, with the most positive attitudes being demonstrated in a combined sample from Norway and Sweden with an OSS of 34 (Nillson & Wetterling, 2013; St. Louis, Sønsterud, Carlo, Heitmann, & Kvenseth, 2014), followed next by the same OSS (23) from a sample from Bosnia-Herzegovina and a combined sample from Ireland and England (Daly & Leahy, 2014; Tyrrell, 2009; Węsierska & St. Louis, 2014), then by a sample from Germany (OSS=15) (Theiling, 2013), and finally with the least positive attitudes from Italy (OSS = −3) (Tomaiuloi et al., 2013). It was significant that for three of the country samples, i.e., Norway, Bosnia-Herzegovina, and Italy, three different regions of the countries had been sampled in the same investigations, permitting regional comparisons within the countries. In each case, differences within countries were small, even though differences among countries were large. In the case of Italy, for example, 100 adults were sampled from several provinces in the north, in the central region, and in the south, wherein OSSs were 0, 12, and −6, respectively. All of these studies used convenience sampling except a portion of the Swedish sample which employed probability sampling.

In searching for predictors in this compilation of studies of European attitudes, the authors were limited by the differences in sampling procedure (e.g., different strategies for convenience sampling), sample size, and other factors. Accordingly, they used a strategy of comparing rank orders of numerous demographic variables in the five countries (or areas) with rank orders for OSS. This procedure would not identify the effect of a given variable (e.g., sex of the respondents) on all the attitudes. Instead, it would provide an estimate of whether that variable had parallel or nonparallel effects in the different countries or areas. It is noteworthy that the demographic variables of sex, age, and education had less effect on public stuttering attitudes than one's citizenship. Thus, national identity, or unique factors related to being a citizen of a particular country, was hypothesized to be a variable worthy of further investigation. For example, it might seem that the most negative Italian attitudes might somehow be related to less positive attitudes that have been observed in southern Europe and Turkey, but the samples in Bosnia-Herzegovina, where more positive OSSs were observed, were roughly in the same latitude. Furthermore, recent unpublished POSHA–S data from France and Spain generated OSSs closer to or above the overall POSHA–S database mean, 19 and 16, respectively (M. Eboli, personal communication, April, 2014; L. Leprovost, personal communication November, 2015).

Given that no systematic studies of public attitudes toward stuttering in any country using the POSHA–S have been published, the current research focused on Portugal and its potential geographical and cultural influences that might affect Portuguese public attitudes. Would Portuguese attitudes be more negative than the POSHA–S median values, as was the case in Italy, or be close to the median, as was the case in France and Spain (with which Portugal shares a border)? Regardless of the answers to this question, the need exists for a study designed to sample public attitudes toward stuttering in such a way that would enable statistical inferences from the sample to the typical, literate, adult population of an entire country. To do so, the sample size would need to be large enough to provide the statistical power to identify potential demographic variables that influence public attitudes within the country.

A systematic, country-wide sample of public attitudes toward stuttering in Portugal would also be useful for stakeholders in the country. Although no systematic research study has been conducted related to public attitudes toward those who stutter, information distributed through pamphlets by the Portuguese Stuttering Association indicate that the public holds the stuttering stereotype with misconceptions about stuttering as well as a lack of information about the disorder (Associação Portuguesa de Gagos, 2014). Results from a carefully designed, probability sampling study could inform future public awareness

campaigns or cognitive-behavioural treatments to attempt to mitigate negative stuttering attitudes and also serve as a baseline against which potential long-term attitude changes could be compared.

4.2.3 Purpose of the Study

The purpose of the present research is to conduct a comprehensive and representative study to measure the knowledge, beliefs and attitudes related to stuttering and PWS throughout the country of Portugal using a translated version of the widely used questionnaire *POSHA-S*. Three main research questions guided the study:

- What are the knowledge, attitudes and beliefs regarding stuttering and toward PWS of the Portuguese population?
- How do the knowledge, attitudes and beliefs toward PWS of the Portuguese population compare to the attitudes of other samples worldwide?
- What are the demographic variables that may have influenced Portuguese stuttering attitudes?

4.3. Method

4.3.1 Attitude Measure

4.3.1.1. Instrument

The POSHA–S was used as the measure of public attitudes toward stuttering. The POSHA–S is an epidemiological survey instrument designed to measure public attitudes toward stuttering worldwide (St. Louis, 2012c). It has: a demographic section where it is requesting information related to age, sex, educational achievement, race, religion, income, past or current vocation, living arrangement, languages known, physical and mental health, speaking ability, and ability to learn; a section related to the comparison of stuttering with other stigmatizing and nonstigmatising attributes (“anchor” attributes) ranging from positive, i.e., intelligent, to neutral, i.e., left-handed and to negative, i.e., mental illness and obesity; and a detailed section related to aetiology, feelings, concerns, and attitudes related to stuttering.

4.3.1.2 Scoring

Questions of the demographic section and general section that require a 1-5 rating were converted to a scale from -100 to +100 (i.e., “1” = -100, “2” = -50, “3” = 0, “4” = +50 and “5” = +100), based on a standard attitude scale determined during the POSHA-S development (K. O. St. Louis, 2012c). The questions on the detailed section related to stuttering that require a “yes”, “no”, “not sure” response were first converted to a 1-3 scale (i.e., “no” = 1, “not sure” = 2 and “yes” = 3) and then to the -100 to +100 scale (e.g., St. Louis 2011, 2012c). The different POSHA-S scores were converted in the same standard attitude scale from -100 to +100 to allow statistical procedures and comparisons. Ratings on some items are inverted so that, uniformly, higher POSHA–S scores reflect more sensitive or accurate attitudes (consistent with recent literature findings) and lower scores reflect less sensitive or accurate attitudes.

Items ratings of the POSHA-S were clustered and averaged to calculate components, that were means of individual items that compose the POSHA-S questionnaire. Components were clustered and averaged into subscores, i.e., *Beliefs about people who stutter*, *Self reactions to people who stutter*, and Obesity/Mental Illness. *Beliefs about people who stutter* subscore reflect external opinions to the respondent and *Self Reactions to people who stutter* subscore involve internal opinions to the respondent. The internal and external opinion subscores were averaged to calculate the *Overall Stuttering Score* (St. Louis et al., 2011).

4.3.1.3 Psychometric properties

Psychometric and related properties of the POSHA–S have been carefully addressed and shown in numerous publications to be satisfactory and adequate, i.e., reliability (St. Louis, 2012c; St. Louis, Lubker, Yaruss, & Aliveto, 2009), construct and discriminant validity (Flynn & St. Louis, 2011; Louis et al., 2009), and internal consistency (Al-Khaledi et al., 2009; St. Louis, 2012c). The instrument has been shown to be user-friendly, since POSHA-S is easy to measure and scoring (Louis et al., 2009; St. Louis et al., 2008). It has been typically administered as a paper-and-pencil questionnaire, but an online administration versus paper-and-pencil administration were found to generate very similar results (St. Louis, 2012b).

4.3.1.4 Translations

The POSHA–S has been translated into 26 different languages (circa March 2016), documenting that the instrument can be successfully and efficiently translated to other languages in order to fulfil its purpose of being used internationally (St. Louis, 2015a, 2015b). For example, in the most comprehensive translation study, (St. Louis & Roberts, 2010) revealed that differences in stuttering attitudes in two officially bilingual countries (Canada and Cameroon), were large regardless of whether an experimental prototype of the POSHA–S was given in English or French (when respondents selected their stronger language for the survey). By contrast, differences between English and French POSHA–S were small, regardless of the country.

Based on the study developed on bilingual countries, (St. Louis & Roberts, 2010) recommended that translations should be carried out by a bilingual person (in English and the other language) with knowledge of speech and language therapy; a back-translation to English should be made by another bilingual person without familiarity with POSHA–S or the details of the study in question.

Because many survey and assessment instruments are used in different European languages, translation guidelines are more stringent than those recommended by St. Louis and Roberts (2010) in order to provide the sense and sensibility (i.e., the purpose of the measure, comprehensibility, content validity, replicability and suitability of the scales) of assessment tools in the original culture (Geisinger, 1994). Recommended procedures include (Beaton, Bombardier, Guillemin, & Ferraz, 1998; Guillemin, Bombardier, & Beaton, 1993): translation, synthesis of the translations, back-translation, committee review of the resultant translation, and a cognitive debriefing of the committee for final adjustments. POSHA–S was translated and cross-culturally adapted to European Portuguese (EP) in five recommended stages to address the differences in culture, cultural background, and language (Gaines, Runyan, & Meyers, 1991). International guidelines were used to conserve the sensibility of the assessment tool in the original culture (Geisinger, 1994), including the following procedures: forward translation, synthesis of the translations, back-translation, review of back translations, committee review and cognitive debriefing (Beaton et al., 1998; Guillemin et al., 1993; Wild et al., 2005). The different steps are described below.

4.3.1.4.1 Step 1: Forward translation

Before performing this initial step, the authors should prepare the translation process, obtain permission to use the instrument from the original developer and request an invitation to be involved (Wild et al., 2005).

The first stage of a cross-cultural adaptation must be the production of several translations by, at least, two independent translators. This leads to the detection of errors and divergent interpretations of ambiguous items in the original tool (Wild et al., 2005). The translators must be fluent in both languages (with the target language as their mother tongue), knowledgeable of the two cultures, and experts in the content measured by the instrument (Beaton et al., 1998; Gaines et al., 1991). One of the translators should be aware of the concepts of the questionnaire being translated, aiming for equivalence in a more clinical perspective; the other translator should not be sensitive to nor be informed of the concepts, and be more apt to detect different non-equivalent meaning from the original than the other translator (Beaton et al., 1998).

4.3.1.4.2 Step 2: Synthesis of translations or reconciliation

In a second stage the two forward translations should be reconciled into a single translation. The production of one common translation can be carried out by the two forward translations, by an independent native speaker or by a researcher who may have developed one forward translation and who will also be involved in the cognitive debriefing (Beaton et al., 1998; Wild et al., 2005).

4.3.1.4.3 Step 3: Back-translations

To help ensure the quality of the final version, it is necessary to back-translate the assessment tool, which means translating back from the target language into the source language (Guillemin et al., 1993). The same number of back-translations and translations, based on the synthesised translation (produced in the second stage) should be produced (Geisinger, 1994; Hutchinson, Bentzen, & König-Zahn, 1997; Wild et al., 2005). A quality back-translation should be made by fluent back-translators in the language, in order to address linguistic nuance and colloquial forms (these translators should have the source language as their mother tongue). The back-translators must be totally blind to the original version and, preferably, the translation should be carried out by a person who does not have any prior knowledge of the intent and concepts underlying the assessment tool, to

minimise bias and predisposed interpretations (Geisinger, 1994). Another technique to review the translation in a more effective way is to use a group of individuals meeting the same criteria as the test translator to review the quality of the adaptation. The individuals who are part of the panel review the items and make comments, then share the comments with one another and discuss discrepancies and differences of opinion (Gaines et al., 1991). After that, the translators need to consider the comments made by the panel of experts throughout a discussion process and arrive at a consensual version of the instrument, which will reflect the best judgment of the entire group (Gaines et al., 1991).

Geisinger (1994) suggests that this step can be substituted by a more effective technique; using a group of individuals similar to those used as translators, to review the translations made.

4.3.1.4.4 Step 4: Expert committee review

Wild et al. (2005, p. 100) considered that the comparison between the back translation and the source version "(…) was one of the most important components of the cross-cultural adaptation process (…), with the "(…) review of the back translation against the original being the key function".

To ensure that the instrument has been translated consistently and in order to highlight gross inconsistencies and conceptual errors, back-translations are compared with the original tool (Hutchinson et al., 1997). However, the agreement between back-translations and the original tool does not totally guarantee a satisfactory translation (Beaton et al., 1998).

To achieve cross-cultural equivalence between the final and the source version, an expert committee compares all the versions of the questionnaire (i.e., original instrument, translations and back-translations). Members of the committee must be multidisciplinary, i.e., professionals from different areas such as methodologists, health professionals, language professionals and translators (Beaton et al., 1998; Geisinger, 1994). The objective of the committee is to produce a pre-final version for field testing, based on the translations and back-translations obtained. The discrepancies should be solved using structured techniques. A decentering technique is one example of a structured technique, which "(…) considers the source and final versions equally important" (Guillemin et al., 1993, p. 1423) and both versions can be subject to modifications during the translation process.

It is also feasible that the committee modifies or eliminates irrelevant, inadequate or ambiguous items and generates others, considered more suitable for the new cultural target, while maintaining the general concept of the deleted items (Guillemin et al., 1993).

The committee needs to guarantee that the tool is fully comprehensible and that the introduction to the assessment tool and the instructions for the completion of the questionnaire are cautiously translated, to safeguard the replicability of the measure (Feinstein, 1987). The various strands of equivalence must be considered: semantic equivalence (i.e., equivalence in the meaning of words), idiomatic (i.e., equivalence in idioms and colloquialisms), experiential (i.e., equivalence in the target cultural context) and conceptual (i.e., equivalence of the concepts and experiences of the target culture); and their optimal transposition guaranteed (Beaton et al., 1998; Guillemin et al., 1993).

The expert committee should make sure that the final questionnaire is understood by a 12-years-old child, which is the general recommendation for questionnaires (Beaton et al., 1998).

4.3.1.4.5 Step 5: Cognitive debriefing or pre-testing

The fifth stage, also called cognitive debriefing or pre-testing, assesses the level of comprehensibility and cognitive equivalence, tests alternative wording and verifies understandability and interpretation of the translation (Wild et al., 2005). The cognitive debriefing is a probe of face validity, i.e., "the confirmation that questions are acceptable without arousing reluctance or hesitation" (Guillemin et al., 1993, p. 1424).

The assessment tool or questionnaire must be administered to a small sample of individuals (i.e., between 5 to 8 native speakers of the translated language, who represent the target population in terms of gender, age, education or/and diagnosis) and, after that, people are interviewed (e.g., using a think-aloud technique) in order to establish the level of comprehensibility of the instructions, the closing comments and questions and the cognitive equivalence of the translation. This may also permit the testing of translation alternatives and to find out if there are items that may be inappropriate or confusing (Wild et al., 2005). The translation should be revised taking into account the answers obtained. Wild et al., (2005, p. 102) points out that "items and response options may be reworded where respondents' comments justify such changes (…)".

In order to assess minor errors (e.g., spelling or grammatical) that could be missed during the cross-cultural adaptation process, the final version should be proof-read (Wild et al.,

2005). After the proof-reading step, the version of the assessment tool should then be administered to a representative sample of the population (Geisinger, 1994; Guillemin et al., 1993; Wild et al., 2005).

For the present study, the guidelines to reach equivalence between the cross-cultural adapted questionnaire and the original questionnaire (Beaton et al., 1998; Guillemin et al., 1993; Wild et al., 2005) presented above, were followed.

Two independent translators with EP as their mother tongue, following the recommendations of St. Louis and Roberts (2010), were involved during the first stage (forward translation). One of the translators was a bilingual person without familiarity with the questionnaire and the other one was a SLT with specific knowledge about stuttering, as recommended by Guillemin et al. (1993) and Beaton et al. (1998).

For the second step (synthesis of forward translations), the two translators produced a reconciled translation.

For the third step (production of back translations), two back-translations were carried out by two teachers from the Department of Languages and Cultures of the University of Aveiro, with the source language as their mother tongue (i.e., English) as recommended by Beaton et al. (1998). The two translators performed two independent back translations based on the common translation resulting from step 2.

The review committee (step 4) included one translator, one back translator and the project manager, in order to assess equivalence between the source and the translated version, to arrive at the pre final version.

The pre-final version produced was submitted to evaluation by a group of five persons, during the cognitive debriefing step (step 5) (Wild et al., 2005). The group, selected by convenience, was similar to the target population. After completing the POSHA-S EP pre final version, the five persons assessed instructions and items of each section of the questionnaire in terms of relevance, clarity, simplicity and accuracy, with a VAS (Crichton, 2001; Hasson & Arnetz, 2005). Bland and Altman (1986) modified method (Jesus et al., 2015, pp. 4–5) was used to assess the agreement between the five judges.

The proof-reading phase was executed by one of the translators and the project manager. Figure 4-1 presents the sequential process of the POSHA-S original into EP.

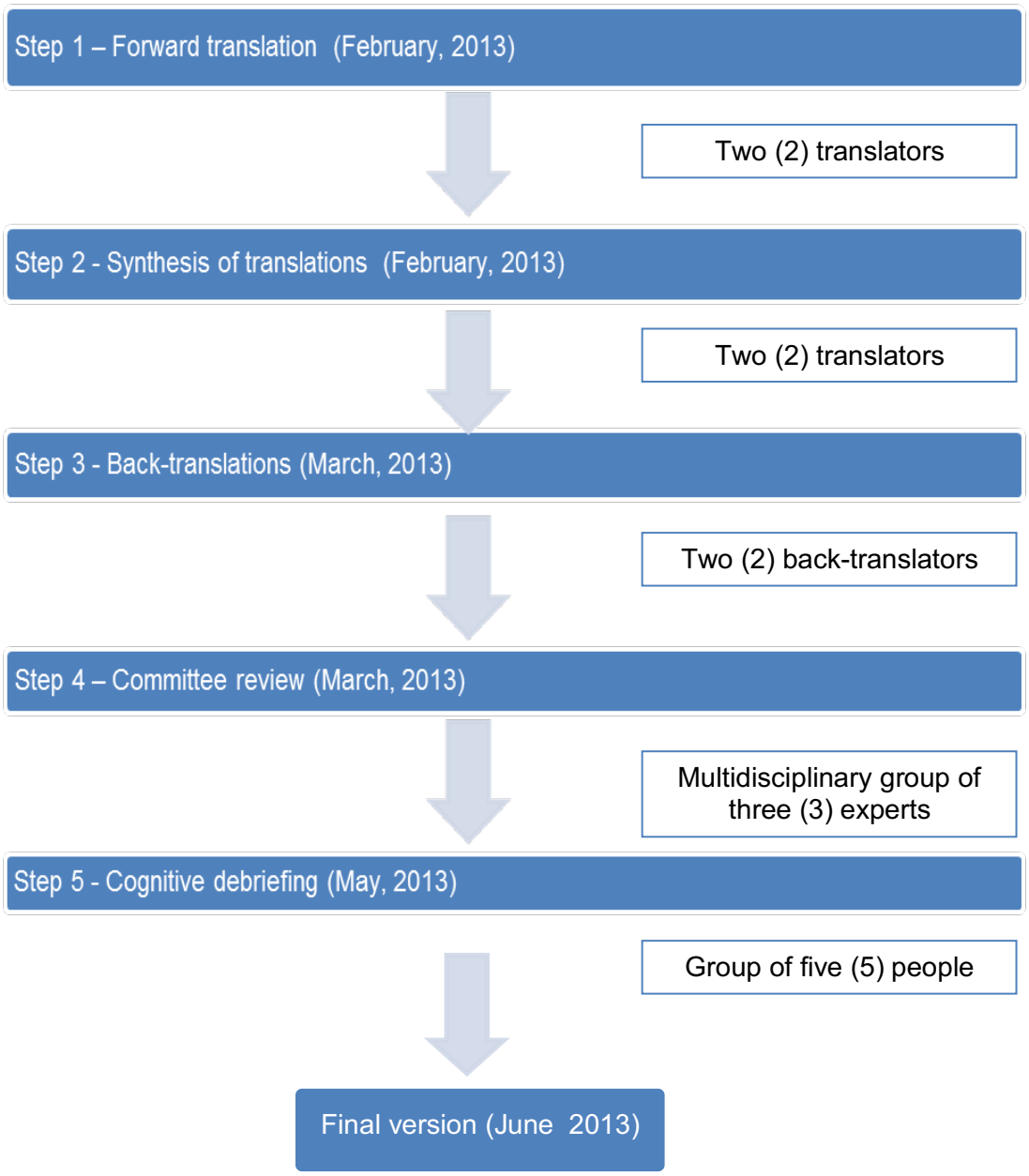


Figure 4-1 Cross-cultural translation adaptation process of the POSHA-S into EP.

4.3.2 Probability Sampling Scheme

A three-stage cluster probability sampling scheme, approved by an ethical committee and the National Data Protection Commission, was implemented for this study, with each stage including simple random sampling (Thompson, 2012). The scheme was designed to locate adults representing the population of Portugal as accurately as possible.

The first stage entailed compiling a list of all Portuguese mainland districts and Islands (a total of 20 districts) as well as a list of all administrative regions (*concelhos*) within them. All

308 administrative regions in the entire country were included, with a number assigned to each region. Using random sampling without replacement five administrative regions within each district were selected such that a total of 100 administrative regions were included in the sampling frame.

For the second stage, a list of all administrative subregions (*freguesias*) within each of the five administrative regions selected from each district was compiled resulting in a total of 1456 administrative subregions. Again, random sampling without replacement was used to choose one administrative subregion within each administrative region already chosen, yielding a total of 100 administrative subregions. The local administrative office (*Junta de Freguesia*) of each administrative subregion was the site where potential respondents would be selected because these local offices are accessible and used by the local population in Portugal for a wide variety of public services that are necessary and frequently used by adults, all in one place. Specifically, the local administrative office is the place where such activities as obtaining internet access, licenses (e.g., for animals or hunting), photocopy authentication, legal advice, social support services, tutoring (academic support), support for being unemployed, and – in some local offices – postal services are available. It is also the place where people attend various citizen meetings or workshops organized by the offices, register to vote, and cast ballots in elections. As such, a wide range of the population in terms of age, sex, occupation, education, income level, and health status would regularly go into the offices.

The third of the three-stage cluster probability sampling scheme comprised the actual respondent identification and questionnaire distribution. Staff at each local administrative office (representing each of the 100 administrative subregions) selected one male and one female within each of three age groups: 18–24 yr, 25–64 yr, and 65 yr or older to be included as potential respondents. In summary, the random sampling procedure would represent 100 administrative subregions (taken from five administrative regions, one from each of 20 districts) and would yield 600 potential respondents (300 males and 300 females). One third of the males and females would be in each of the three age ranges.

It was made a list of the local administrative offices in each of the 100 administrative regions through a web search and then contacted those that were selected randomly in the second stage. The first contact, carried out via telephone and/or email explained the purposes of the study, explained the random procedure to select suitable persons to fill out the POSHA–S questionnaires, and asked permission to distribute questionnaires. When a contact person of the local office declined to assist in the study, the next local office of the same

administrative subregion was contacted and followed the same procedure. Of the 100 local administrative offices contacted, the initial compliance rate was 70/100 or 70%. Proceeding to the next local administrative office after one declined generated a compliance rate of 100%.

The contact person, typically a staff employee, at each local office made the final selection of respondents (a male and female in each of three age groups). It was requested that the staff person ask the first six persons entering the office to complete the questionnaire who would complete the inclusion criteria, i.e., that they lived in the subregion, that they were able to read and write, and that they were male or female in one of the age ranges.

For statistical analysis, the administrative regions randomly chosen were clustered using the Nomenclature of Territorial Units for Statistics (NUTS level II, hereinafter referred to as "Region"), which is a hierarchical system that divides the country based on existing administrative units and population size (EUROSTAT, 2011). The 100 administrative regions were distributed through seven NUTS level II regions, i.e., five divisions on Portugal's mainland (North, Central, Lisbon, Alentejo, and Algarve) and two island regions (Madeira and Azores), as shown in Fig. 2. The region percentages of the NUTS II data are also represented in Fig. 4.1.

4.3.3 Data analysis

For the first purpose, documenting stuttering attitudes in Portugal, *POSHA–S* demographic, general and detailed stuttering items were analysed descriptively in terms of converted *POSHA–S* scores. For the second objective, predicting *POSHA-S* combined ratings (subscores and the Overall Stuttering Score), the demographic variables were targeted for their predictive potential for stuttering attitudes (i.e., age group, regions in country, completed school levels, working situation, sex, marital status, parenthood, religion, native language and number of languages). First, eight multivariate analysis of variance (MANOVA) procedures were run in IBM SPSS Statistics 22.00 to assess the overall influence of each of these predictive variables on the Beliefs subscore, the Self Reactions subscore, and the OSS with a Bonferroni correction such that alpha level for significance was set at $p \leq 0.00625$ (.05/8). Roy's largest root was chosen as the discriminant function variate because this statistic typically generates the most powerful results on MANOVA comparisons (Field, 2013). The second step involved identifying the predictor variables that revealed statistically significant impact using univariate ANOVAs for their influence on the stuttering subscores and the OSS. Again, a Bonferroni correction was implemented,

resulting in a criterion for significance of $p \leq 0.0168$ (.05/3). As a third and final step, Gabriel post-hoc pairwise comparisons were run between means in all variables for which the ANOVAs were significant. Gabriel post-hoc tests were chosen since they can accommodate differences in sample sizes (Field, 2013).

The third objective involved comparing the Portuguese attitudes to those from around the world. Mean ratings for each *POSHA–S* component, subscore, and the Overall Stuttering Score were compared with the lowest, highest, and median sample mean values of the *POSHA–S* database (developed in Excel), containing 141 samples representing 11383 public respondents from 42 countries and 26 languages (circa March, 2016). Mean values of the Portuguese sample above the median of the *POSHA–S* samples would reveal more accurate/positive attitudes than average and values below the median would represent more inaccurate/negative attitudes. Further, percentile ranks of the mean ratings for *POSHA–S* items, components, subscores and *OSS* for the Portuguese sample derived from all 141 samples in the database were also calculated. The percentage of ratings falling in the first quartile (0-25), interquartile range (25-75 percentile) and the fourth quartile (75-100) were calculated, to determine if the attitudes of the Portuguese sample were about average or more/less positive than those in the *POSHA–S* total sample.

4.4. Results

4.4.1 Translation process

For the EP translation the authors followed the six steps described previously (Beaton et al., 1998; Guillemin et al., 1993): translation, synthesis of the translations, back-translations, back translations review, committee review and cognitive debriefing. Concerning forward translation, the two translators made a question-by-question translation of POSHA-S into EP, with exception of one question in the demographics section relating to the Portuguese educational system; as there are differences between the Portuguese and American educational systems, the two translators made the necessary cultural adjustments. After the production of the two translations, both translators synthesised and agreed on a common final translation (Beaton et al., 1998; Wild et al., 2005). Disagreements were solved by discussing key issues until a consensus decision was attained. The result was a final reconciled version of POSHA-S in EP, prepared to be back translated into English, as recommended by Wild et al. (2005). The two back-translators had no knowledge of the purpose of the questionnaire and were totally blind to the original English POSHA-S version

(Geisinger, 1994; St. Louis & Roberts, 2010). The two back translators used the synthesised version compiled resulting from step 2 to translate independently into English.

The final version of POSHA-S EP, based on translations and back-translations, resulted from an expert committee meeting, involving three persons from different disciplinary areas (Beaton et al., 1998; Geisinger, 1994): One of the translators (a clinician), one of the back translators (a teacher at the Department of Languages and Cultures of the University of Aveiro and a language specialist) and the project manager (teacher at School of Health Sciences and with experience in methodology).

Based on the recommendations of Beaton et al. (1998), the common translation and the back translations were compared with the source version. The discrepancies were on a table (Table 4-1): The first four columns presented the specific item where the discrepancies were found in each version of the questionnaire and the final column presented the solution of the committee. Fifteen discrepancies were found and a solution was agreed upon for all of them. Throughout the expert committee meeting, semantic, experiential and conceptual equivalence of the questionnaire were discussed and ensured. Accordingly, the majority of the items of the pre-final translation were equivalent in terms of meaning, content addressed and concepts used (Beaton et al., 1998; Guillemin et al., 1993). Exceptionally, the committee members agreed that the item relating to race (“My race is?”) was not relevant for Portuguese culture and could trigger cultural filters; for that reason, the item was deleted from the pre-final version.

Table 4-1. Example of a summary report.

Source version	Back translation 1	Back translation 2	Synthesis of translations	Resolution
“we ask you to give”	Express	Express	Exprima	“Dê a sua opinião…”
“will help us…”	Allow	Allow	Permita	“(…) que nos irá ajudar a melhor interpretar…”

The modified Bland-Altman plot (Figure 4.1) for more than two judges allowed the visual observation of the relation between the mean and the standard deviation of the rating attributed by each judge in each question, during cognitive debriefing. The central line in the Bland and Altman plot represents the mean of the standard deviation. The upper limit was calculated as the square root of the division by n-1 of the average variance multiplied by the 95th percentile chi-square distribution with n-1 degrees of freedom (n is the number

of judges). The lower limit corresponds to x-axis. The responses are in agreement if the standard deviation of the judges' responses lies between the upper and lower limit (Jesus et al., 2015, pp. 4–5).

After the cognitive debriefing, a modified Bland and Altman plot was used to visualise the agreement/disagreement levels between the 5 judges (Figure 4.2).

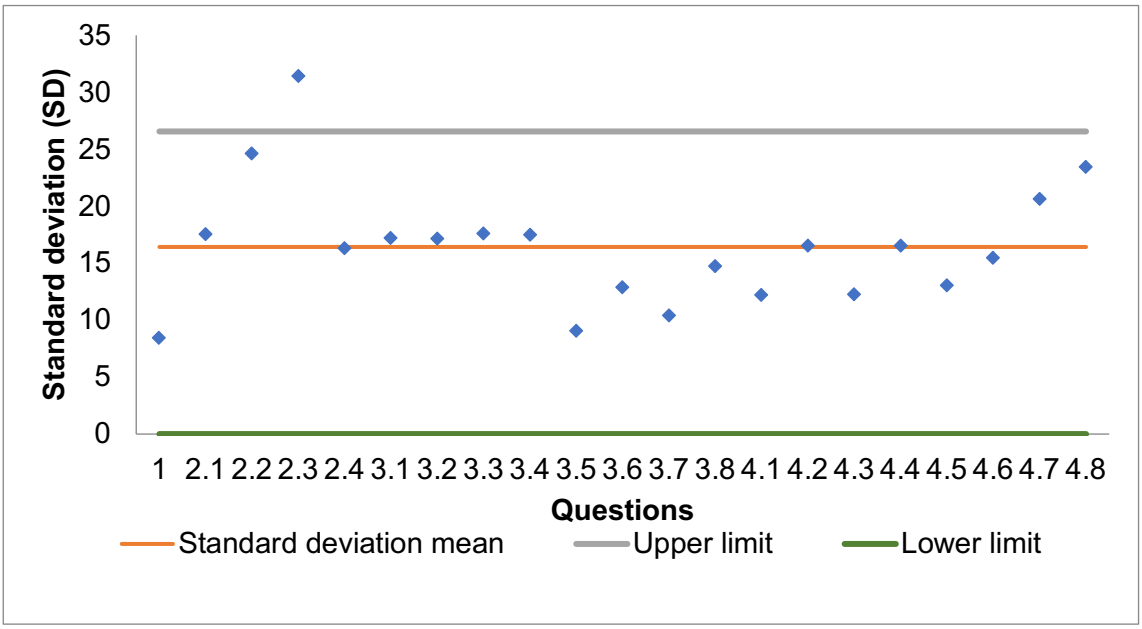


Figure 4-2 Modified Bland and Altman plot.

As can be observed in Figure 4-2, the standard deviation of question 2.2 exceeds the upper limit, which means that the 5 judges disagreed relating to the simplicity of the instructions. Such a finding meant that the instructions were not written with the simplest words to convey the meaning. Based on these results, simplicity in sentence construction of *POSHA-S* instructions was added to the translated questionnaire.

4.4.2 Respondents

From a total of 600 persons contacted through the three-stage sampling procedure described above, a total of 311 questionnaires were returned, generating a return rate of 51.8%. (Babbie, 2006) suggests that a response rate of at least 50% is considered acceptable for analysis in a social research postal survey.

Table 4-2 present the information of the respondents organized by demographic categories (i.e., absolute and relative frequencies of age group, gender, NUTS level II, school level completed, working situation, marital status, parenthood, religion, native language and number of languages spoken. The collected sample (according to NUTS level II) was roughly in line with the country's population (according to 2011 Census). Table 4-3 presents the respondents data related to whether the respondent knew a PWS, self-identification or no persons known with various attributes, relative income, self-rating of health and abilities, life priorities and completion time.

Relative income, self-rating of health and abilities and life priorities are presented on a -100 to +100 scale, with the percentiles related to the database shown in parentheses. The NUTS level II percentages are represented in Figure 4-3.

Table 4-2. Descriptive demographic variables information with numbers and percentages.

		Portuguese population sample: Number (%)
Age group	18-24 yr	88 (28.3%)
	25-64 yr	124 (39.9%)
	≥65 yr	90 (28.9%)
	No answer	9 (2.9%)
Gender	Male	150 (48.2%)
	Female	157 (50.5%)
	No answer	4 (1.3%)
NUTS level II	North	62 (19.9%)
	Center	94 (30.2%)
	Lisbon	38 (12.2%)
	Alentejo	68 (21.9%)
	Algarve	16 (5.1%)
	Azores	11 (3.5%)
	Madeira	16 (5.1%)
	No answer	6 (1.9%)
School level completed	Primary school (1st cycle)	71 (22.8%)
	Middle school (2nd cycle)	19 (6.1%)
	Middle school (3rd cycle)	43 (13.8%)
	High school	89 (28.6%)
	Trade/military/technical school	30 (9.6%)
	4-year university degree	35 (11.3%)
	Master or similar degree	10 (3.2%)
	Doctoral degree	1 (0.3%)
	No answer	13 (4.2%)
Working situation	Student	35 (11.3%)
	Working	140 (45.0%)
	Unemployed or not working	42 (13.5%)
	Retired	87 (28.0%)
	Student worker	2 (0.6%)
	No answer	5 (1.6%)
Marital status	Married	185 (59.5%)
	Not married	112 (36.0%)
	No answer	14 (4.5%)

Table 4-2 (cont). Descriptive demographic variables information with numbers and percentages.

		Portuguese population sample: Number (%)
Parenthood	Yes	165 (53.1%)
	No	117 (37.6%)
	No answer	29 (9.3%)
Religion	Christian	229 (73.6%)
	Catholic	204 (65.6%)
	Denomination unspecified	23 (7.4%)
	Evangelical	2 (0.6%)
	Muslim	1 (0.3%)
	Agnostic	4 (1.3%)
	Atheist	3 (1.0%)
	None	6 (1.9%)
Native language: Number (%)	Portuguese	303 (97.4%)
	Other	1 (0.3%)
	No answer	7 (2.3%)
Number of languages: Number (%)	1	137 (44.1%)
	2	83 (26.7%)
	3	47 (15.1%)
	4	39 (12.5%)
	No answer	5 (1.6%)

Table 4-3. Additional variables based on self ratings with numbers and percentages or mean – 100 to +100 ratings and percentiles relative to the *POSHA-S* database.

		Portuguese population sample: Number (%)
Persons known who have a stuttering disorder: Number (%)	Nobody	80 (25.7%)
	Acquaintance	172 (55.3%)
	Close friend	31 (10.0%)
	Relative	33 (10.6%)
	Me	5 (1.6%)
	Other	9 (2.9%)
Self-identification: Number (%)	Stuttering	5 (1.6%)
	Mentally ill	2 (0.6%)
	Obese	18 (5.8%)
	Left handed	13 (4.2%)
	Intelligent	51 (16.4%)
No persons known: Number (%)	Stuttering	80 (25.7%)
	Mentally ill	113 (36.3%)
	Obese	45 (14.5%)
	Left handed	47 (15.1%)
	Intelligent	23 (7.4%)
Relative income score: Mean (Percentile)	Income: Family/Friends	-16 (6)
	Income: Countrymen	23 (11)
	Income: Composite*	-22 (7)
Self-rating of health and abilities: Mean (Percentile)	Physical health	15 (0)
	Mental health	40 (6)
	Ability to learn	36 (11)
	Ability to speak	44 (11)
Life priorities: Mean (Percentile)	Be Safe/Secure	79 (40)
	Be Free	65 (54)
	Spend Time Alone	13 (7)
	Attend Social Events	14 (43)
	Imagine New Things	43 (65)
	Help Less Fortunate	60 (75)
	Have Exciting Experiences	-30 (25)
	Practice My Religion	13 (46)
	Earn Money	66 (71)
	Do Job/Duty	78 (59)
	Get Things Done	76 (55)
	Solve Big Problem	73 (64)
Completion time	Mean	14.2 minutes

*Composite income score is weighted more heavily for countrymen than family and friends.

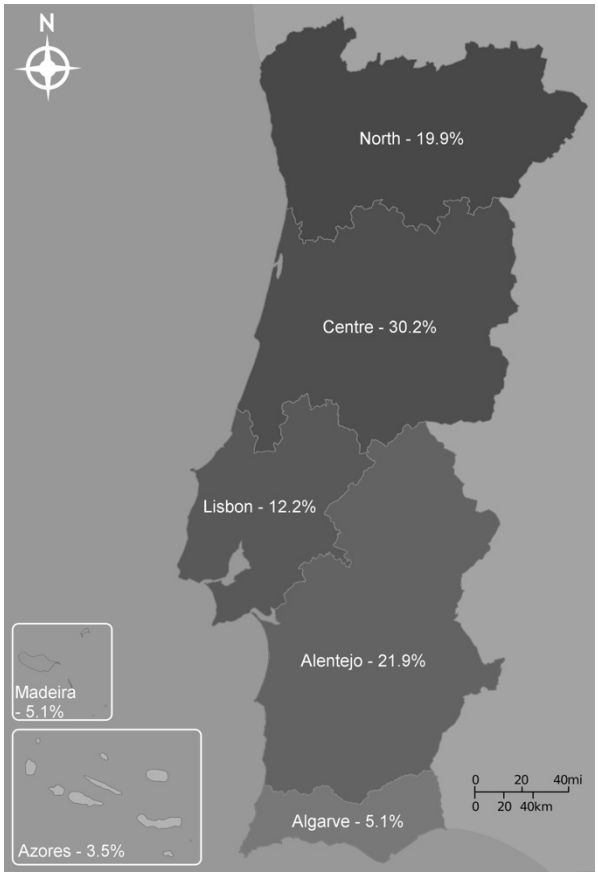


Figure 4-3 NUTS level II regions and percentage of respondents.

Considering the demographic variables presented on Table 4-2, the large majority of the respondents were from the mainland (89.7%), specifically from the centre of the country (30.2%). The majority of the Portuguese respondents who completed questionnaires were in the [25-64] age group (39.9%). Respondents were about evenly split female to male, i.e., 50.5% to 48.2%. The largest percentage of respondents had completed high school (28.6%) and were working (45.0%). The largest percentage of the sample were married (59.5%), had children (53.1%) and spoke one language (44.1%), i.e., European Portuguese. The vast majority were Catholic (65.6%).

Data from Table 4-3 revealed that the relative income and self-rating of health and abilities were scored below average of the *POSHA–S* total database samples. The majority of life priorities results were rated higher than average in comparison with the total database samples.

More than half (55.3%) of the respondents reported having an acquaintance who stutters. Five people (1.6%) identified themselves as individuals who stutter and 25.7% did not know anyone who stuttered. Very few of the respondents viewed themselves as a PWS (1.61%)

or mentally ill (0.6%). Somewhat surprisingly, only 16.4% of the Portuguese public regarded themselves as intelligent. More than one-third (36%) of the respondents reported knowing no one that was mentally ill, which represented the least familiarity with the five attributes.

4.4.3 Attitudes toward PWS of the Portuguese Population and comparison with the *POSHA-S* database

Table 4-4 provides all of the mean ratings of the Portuguese sample scored in the POSHA–S (items, components, subscores, and the OSS), along with the median database values and percentiles.

In comparison to previously analysed samples, the 60 sample ratings are distributed as follows: 1st quartile or least positive attitudes, 7%; interquartile range or average attitudes, 73%; and 4th quartile of most positive attitudes, 20%. The mean percentile value was 54. Taken together, the Portuguese attitudes obtained through the probability sampling procedure were mostly average but with a trend for slightly more positive than average ratings. The comparisons can be seen clearly in Fig. 4-4 where mean values parallel the median database value quite closely but reflect more positive attitudes for the Potential and Social Distance/Sympathy components (which were in the 4th quartile) and less positive than average attitudes for the Help From component (which falls into the 1st quartile). Portuguese respondents also indicated knowing considerably less than the median database value for the combined Amount Known about obesity and mental illness (1st quartile). Overall, however, Fig. 4-4 shows graphically that the Portuguese respondents, chosen according to a careful probability sampling procedure, generated mean stuttering attitudes ratings that were similar to the medians from the POSHA–S database.

Selected items that reflected more positive than average attitudes (i.e., in the 4th quartile) were: Agreeing that PWS can hold any job they want, rejecting that they should hide their stuttering; not feeling impatient while a conversational partner is stuttering; not being concerned if their doctor, a sibling, or the respondents themselves stuttered; wanting to stutter themselves. More negative than average attitudes (i.e., in the 1st quartile) were reported for three items: Accepting that a medical doctor should help a PWS and rejecting that others who stutter should do so and not agreeing that stuttering is caused by genetic inheritance.

Table 4-4. Comparison of mean POSHA–S ratings of Portuguese respondents, the median of 141 sample means from the POSHA–S database, and percentile ranks of the Portuguese means relative to the database sample means.

***POSHA–S* VARIABLE**	**Portuguese Public**	***POSHA–S* Database Median**	***POSHA–S* Database Percentile**
Overall Stuttering Score	***19***	***17***	65%ile
***Beliefs* about persons who stutter**	**34**	**32**	63%ile
Traits/personality	**18**	**18**	61%ile
Have self to blame[a]	84	80	70%ile
Nervous or excitable[a]	-24	0	31%ile
Shy or fearful[a]	-6	-23	60%ile
Stuttering should be helped by…	**4**	**16**	34%ile
Speech and language therapist	94	93	63%ile
Others who stutter	-21	-2	24%ile
Medical doctor	-60	-31	20%ile
Stuttering is caused by…	**37**	**32**	69%ile
Genetic inheritance	3	17	21%ile
Learning or habits[a]	47	20	68%ile
A very frightening event[a]	-12	-3	35%ile
An act of God[a]	65	64	56%ile
A virus or disease[a]	43	39	53%ile
Ghosts, demons, spirits[a]	76	87	26%ile
Potential	**75**	**64**	83%ile
Can make friends	93	92	59%ile
Can lead normal lives	91	89	66%ile
Can do any job they want	72	45	86%ile
Should have jobs requiring good judgment	42	40	44%ile
***Self Reactions* to people who stutter**	**5**	**1**	53%ile
Accommodating/helping	**43**	**40**	52%ile
Try to act like the person was talking normally	84	81	58%ile
Person like me	-20	-26	50%ile
Fill in the person's words[a]	20	25	47%ile
Tell the person to "slow down" or "relax"[a]	-1	0	40%ile
Make joke about stuttering[a]	88	88	60%ile
Should try to hide their stuttering[a]	89	72	80%ile

Table 4-4 (cont). Comparison of mean POSHA–S ratings of Portuguese respondents, the median of 141 sample means from the POSHA–S database, and percentile ranks of the Portuguese means relative to the database sample means.

***POSHA–S* VARIABLE**	**Portuguese Public**	***POSHA–S* Database Median**	***POSHA–S* Database Percentile**
Social distance/sympathy	**28**	**9**	77%ile
Feel comfortable or relaxed	37	31	62%ile
Feel pity[a]	10	17	39%ile
Feel impatient (not want to wait while the person stutters)[a]	74	62	81%ile
Concern about my doctor[a]	66	40	84%ile
Concern about my neighbor[a]	73	73	61%ile
Concern about my my brother or sister[a]	28	-14	80%ile
Concern about me[a]	6	-42	78%ile
Impression of person who stutters	7	2	69%ile
Want to have stuttering	-52	-70	75%ile
Knowledge/experience	**-38**	**-34**	35%ile
Amount known about stuttering	-31	-31	47%ile
People who stutter known	-86	-86	36%ile
Personal experience (me, my family, friends)	3	12	46%ile
Knowledge source	**-14**	**-10**	31%ile
Television, radio, films	18	14	52%ile
Magazines, newspapers, books	-7	-10	44%ile
Internet	-27	-20	32%ile
School	-12	0	32%ile
Doctors, nurses, other specialists	-41	-33	26%ile
Obesity/Mental Illness	**-34**	**-35**	47%ile
Overall impression	**-8**	**-14**	82%ile
Obese	-14	-22	76%ile
Mentally ill	-2	-8	77%ile
Want to be	**-73**	**-84**	65%ile
Obese	-71	-83	65%ile
Mentally ill	-75	-83	60%ile
Amount known about	**-21**	**-5**	24%ile
Obese	-14	3	27%ile
Mentally ill	-27	-18	34%ile

[a]Indicates that ratings are reversed so more positive, accurate, or desirable ratings are higher.

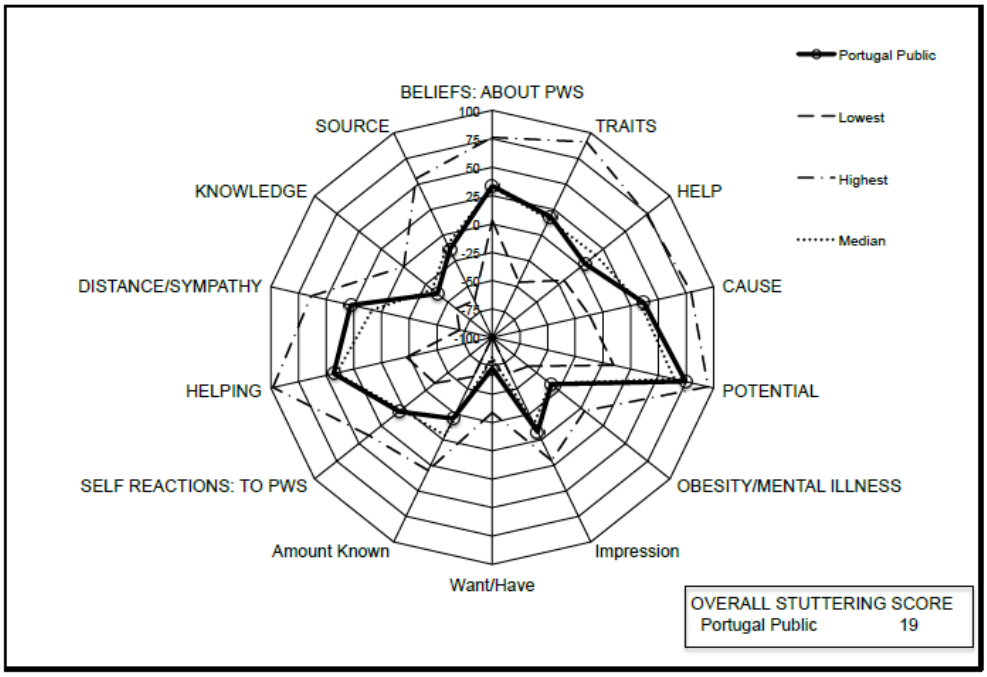


Figure 4-4 Comparison of POSHA–S results from the Portuguese population and the database samples.

4.4.4 *POSHA-S* Predictors Variables of Portuguese Stuttering Attitudes

The predictor demographic variables were analysed relative to subscores and the *Overall Stuttering Score* with multivariate analysis of variance (MANOVA) for the *Beliefs* and *Self Reactions* subscores and the *OSS*. Table 4-5 presented values for Beliefs, Self Reactions, and OSSs of the *POSHA-S* for subcategories in each of the eighth predictor variables. Table 4-6 provides detailed statistical MANOVA results for each variable.

Table 4-5. Mean values for POSHA–S stuttering subscores and Overall Stuttering Scores for predictor demographic variables in the Portuguese population.

		Beliefs about people who stutter	***Self Reactions to PWS***	***Overall Stuttering Score***
Age group	18-24 yr	35.6	9.0	22.3
	25-64 yr	36.5	8.9	22.7
	≥65 yr	28.4	-5.9	11.2
Gender	Male	33.8	5.7	19.7
	Female	33.5	5.0	19.3
NUTS level II	North	32.2	6.9	19.5
	Center	32.3	6.1	19.2
	Lisbon	31.9	-2.8	13.6
	Alentejo	33.2	1.5	17.3
	Algarve	36.8	-5.8	15.5
	Azores	36.6	27.7	32.1
	Madeira	46.7	12.3	29.5
School level completed	Primary school (1st cycle)	30.4	-4.3	13.1
	Middle school (2nd cycle)	36.9	6.8	21.9
	Middle school (3rd cycle)	30.0	7.7	18.8
	High school	35.7	9.9	22.8
	Trade/military/technical school	27.9	5.7	16.8
	4-year university degree	41.4	11.2	26.3
	Master or similar degree	31.0	9.4	20.2
	Doctoral degree	33.3	-21.7	5.8
Working situation	Student	38.0	5.8	21.9
	Working	32.6	10.6	21.6
	Unemployed or not working	38.6	6.4	22.5
	Retired	29.4	-5.02	12.2
	Student worker	30.2	-5.5	12.4
Marital status	Married	32.8	4.2	18.5
	Not married	35.1	5.9	20.5
Parenthood	Yes	33.4	3.4	18.4
	No	33.7	6.7	20.2
Religion	Christian	32.5	3.5	18.0
	Catholic	33.1	3.2	18.2
	Denomination unspecified	27.9	7.0	17.5
	Evangelical	14.6	-12.8	0.9
	Muslim	33.3	24.2	28.8
	Agnostic	35.4	23.4	29.4
	Atheist	47.2	21.3	34.3
	None	25.7	-14.0	5.9
Native language: Number (%)	Portuguese	33.0	4.7	18.8
	Other	50.0	0	25.0
Number of languages: Number (%)	1	30.3	0.2	15.3
	2	37.8	2.9	20.4
	3	30.8	16.2	23.5
	4	35.1	11.2	23.2

Table 4-6. MANOVA results and univariate ANOVAs of POSHA–S stuttering subscores and Overall Stuttering Scores (OSSs) for predictor demographic variables in the Portuguese population. Significant differences are shown by shading.

	MANOVA	***Univariate ANOVA***		
Predictor Variable		***Beliefs about people who stutter***	***Self Reactions to people who stutter***	***Overall Stuttering Score***
Age group	Θ=0.97, F(2,299)=14.52, p<0.001, partial η^2=0.089	F(2, 299)= 4.476, *p*= 0.012, partial η^2=0.029	F(2, 299)= 12.108, *p*< .001, partial η^2=0.075	F(2, 299)= 13.853, *p*< .001, partial η^2=0.085
Sex	Θ=0.000, F(2,304)=0.027, p=0.973, partial η^2=0.000	F(1, 305)= 0.009, *p*= 0.925, partial η^2=0.000	F(1, 305)= 0.051, *p*= 0.822, partial η^2=0.000	F(1, 305)= 0.046, *p*= 0.830, partial η^2=0.000
Region	Θ=0.066, F(6,298)=3.281, p=0.004, partial η^2=0.062	F(6, 298)= 1.279, *p*= 0.266, partial η^2=0.025	F(6, 298)= 3.280, *p*= 0.004, partial η^2=0.062	F(6, 298)= 2.580, *p*= 0.019, partial η^2=0.049
School level completed	Θ=0.068, F(6,290)=3.308, p=0.004, partial η^2=0.064	F(6, 290)= 1.911, *p*= 0.079, partial η^2=0.038	F(6, 290)= 2.826, *p*= 0.011, partial η^2=0.055	F(6, 290)= 3.258, *p*= 0.004, partial η^2=0.063
Working situation	Θ=0.076, F(4,301)=5.721, p=0.000, partial η^2=0.071	F(4, 301)= 1.889, *p*= 0.112, partial η^2=0.024	F(4, 301)= 5.639, *p*< .001, partial η^2=0.070	F(4, 301)= 4.783, *p*= 0.001, partial η^2=0.060
Marital status	Θ=0.003, F(2,294)=0.462, p=0.630, partial η^2=0.003	F(1, 295)= 0.764, *p*= 0.383, partial η^2=0.003	F(1, 295)= 0.316, *p*= 0.574, partial η^2=0.001	F(1, 295)= 0.836, *p*= 0.361, partial η^2=0.003
Parental status	Θ=0.005, F(2,279)=0.629, p=0.534, partial η^2=0.004	F(1, 280)= 0.013, *p*= 0.909, partial η^2=0.000	F(1, 280)= 1.252, *p*= 0.264, partial η^2=0.004	F(1, 280)= 0.715, *p*= 0.398, partial η^2=0.003
Number of languages known	Θ=0.063, F(3,302)=6.313, p<0.001, partial η^2=0.059	F(3, 302)= 2.587, *p*= 0.053, partial η^2=0.025	F(3, 302)= 6.058, *p*= 0.001, partial η^2=0.057	F(3, 302)= 4.055, *p*= 0.008, partial η^2=0.039

Table 4-5 revealed, regarding the *OSS* of the *POSHA-S* (mean of two subscores), that the following population characteristics presents the highest means: age group [25-64], male population, living in Azores, 4-years university degree, student, not married, no parent, atheist, with other native language and who speaks three languages.

Results of MANOVA and univariate ANOVA analyses for each of the eight predictor variables revealed that five variables were statistically significant for OSSs (Table 4-6): age

(F(2,299)=13.853, p=0.004), region (F(6, 298)=2.580, p=0.019), school level completed (F(6, 290) = 3.258, p = 0.004), working situation (F(4, 301) = 4.783, p = 0.001), and number of languages known (F(3, 302) = 4.055, p = 0.008). Three variables did generate significant effects: Sex, marital status, and parental status. Considering the Self Reactions subscore, age, region, school level completed, working situation, and number of languages were significantly different. Only the age group variable significantly influenced the Beliefs subscore.

Gabriel post-hoc multiple comparisons revealed significant differences between the 18–24 and ≥65 age groups (M = 14.966, p < 0.001) and also between the 25–64 and ≥65 age groups (M = 14.831, p < 0.001) for Self-Reactions. For Beliefs, the 25–64 age group held significantly more positive attitudes than the ≥65 group (M = 8.068, p = 0.014). And for OSS, the 18–24 and 25–64 age groups differed significantly from the older age group (M = 11.065, p < 0.001, and M = 11.450, p < 0.001, respectively).

Post-hoc comparisons for the region variable revealed that respondents from the Azores island region held more positive Self Reactions in comparison with Central region (M = 21.58, p = 0.046), Lisbon (M = 30.46, p = 0.004), Alentejo (M = 26.16, p = 0.008), and Algarve (M = 33.45, p = 0.011). Self-Reactions of respondents who completed secondary school or a 4-yr university degree were also more positive than those with only a primary school education (M = −14.245, p = 0.006 and M = −15.493, p=0.041, respectively). The same was true for the OSS comparisons (M=−9.765, p=0.010 and M=−13. 222, p=0.005, respectively).

Working people held significantly higher Self Reactions subscores and OSS values when compared with retired people (M = 15.640, p = 0.000 and M = 9.380, p = 0.001, respectively). Students and unemployed people also had higher OSSs than retired people did (M = 9.671, p = 0.048 and M = 10.267, p = 0.016, respectively).

We carried out additional ANOVAs to identify factors that might explain the lower scores of elderly people. The amount known about stuttering reported by the ≥65-yr age group was lower (although not significantly so) than the amount known by the younger age groups. Nevertheless, Post-hoc comparisons within the ≥65 age group revealed that the amount known influenced negatively the Self Reactions subscore [F(4, 84) = 3.865, p = 0.006]. Accordingly, we submit that misinformation or lack of information can be a reason for significantly negative ratings of older adults compared with younger adults.

Post-hoc comparisons for the number of languages known was associated with significantly more positive attitudes for respondents who spoke three languages versus one language

(M = −15.96, p < 0.001) and also between two languages and three languages spoken (M = −13.30, p = 0.017) for the Self Reactions subscore. OSSs were significantly higher for people who knew three languages in comparison with people knew only one language (M = −8.23, p = 0.024).

4.5. Discussion

4.5.1 Translation process

The *POSHA-S* questionnaire has been previously translated into 26 different languages. With the exception of the Arabic translation (Al-Khaledi et al., 2009) and the English to French translation by St. Louis and Roberts (2010), none of the other cross-cultural adaptation processes were described in a published document.

Al-Khaledi et al. (2009) described the translation process briefly: one person performed the translation and this was assessed by independent linguists; the back translation was produced by another linguist. The authors described several modifications of the items in order to "(...) ensure the accuracy and applicability of the questions when translated into Arabic and therefore ensure maintenance of the reliability and validity properties of the survey" (Al-Khaledi et al., 2009, p. 50). St. Louis and Roberts (2010) also described the process: an experienced translator carried out the forward translation, which then was checked by two bilingual native French speakers and one of the authors of the paper. The translated questionnaire was then submitted to two professional English-to-French translators for comment; this revealed absence of translation problems.

The translations referred did not follow the guidelines recommended and described in the present chapter exhaustively. The EP cross-cultural adaptation performed included an in depth translation process, to ensure the sensibility and the cultural equivalence of the *POSHA-S* EP version, in order to maintain the psychometric properties of the original version (Beaton et al., 1998; Guillemin et al., 1993; Wild et al., 2005).

On the other hand, as the *POSHA-S* was originally developed "(...) avoiding idiom and slang expressions, professional jargon, or words and concepts whose meaning would likely be different in other cultures (...)" (St. Louis, 2012c, p. 140), the translation/cultural adaptation process from the original source to other languages will probably be more simple, due to the "(...) guidelines to foster accurate translations (...) followed by the Task Force that developed the questionnaire (St. Louis, 2012c, p. 140).

Nevertheless, the cross-cultural adaptation process described in this chapter is proposed as a template of guidelines to be follow in subsequent research works using the *POSHA-S*. Through the implementation of the template described, equivalence between the source and the translated target questionnaire will be achieved, allowing for the "(...) comparability of responses across populations divided by language or by culture" (Beaton, Bombardier, Guillemin, & Ferraz, 2000, p. 3190).

4.5.2 Attitudes toward PWS of the Portuguese Population

On the stuttering items of POSHA-S (see Table 4-4), the large majority of Portuguese respondents expressed positive scores, meaning that the attitudes toward PWS are positive and accurate. The Portuguese population believed that PWS should not try to hide their stuttering and that they should have jobs with responsibility (where they have to understand correctly and decide important things). The Portuguese population also rated positively that they would try to act like a stuttering person was talking normally, that they would not make a joke, and that they would not fill in the person's words. They indicated they would not feel impatient and would feel relaxed and comfortable when talking with someone who stutters and not feel pity. However, results have shown a negative score (-1), meaning that the Portuguese population presented an inaccurate attitude, considering that they would tell the PWS to "slow down". Regarding the Portuguese population sample's knowledge of stuttering, the majority of scores are positive, indicating accurate knowledge about stuttering. The Portuguese population accurately agreed that PWS should not be blamed for their stuttering and that they can make friends, lead normal lives and do any job they want. The Portuguese population would have concerns or worry if a doctor, neighbour, sibling or themselves stuttered. Regarding etiology, this population presented an incorrect knowledge, believing that stuttering is caused by a very frightening event. However, and correctly, they also considered genetic inheritance as a cause of stuttering. SLTs were accurately considered people who should help PWS. Most knowledge originated from personal experience or television/radio. Some incorrect knowledge was also observed in this population: negative scores were observed on the items related to considering PWS as nervous/excitable and shy/fearful.

The attribution of negative traits (the Portuguese population considered that PWS are nervous and are shy or fearful) or the conveyance of inaccurate attitudes (the Portuguese sample would tell the person to "slow down") are in accord with the stuttering stereotype observed in different studies developed with and without the POSHA–S questionnaire (Al-

Khaledi et al., 2009; Cooper & Cooper, 1996; Daly & Leahy, 2014; MacKinnon et al., 2007; Özdemir et al., 2011b; St. Louis, 2005; St. Louis, Coskun, et al., 2005; St. Louis & Roberts, 2010; White & Collins, 1984). Despite the association of negative traits to PWS from the Portuguese population, it is encouraging to note that the large majority of scores were positive/accurate, meaning that this population held acceptable attitudes towards PWS.

Research relating to teachers' attitudes towards PWS indicated that lack of familiarity with PWS cannot explain some stereotypes (White & Collins, 1984). Although more than three-quarters (77.4%) of the Portuguese population states that it knows someone who stutters, 22.6% of the respondents reported knowing nobody who stutters; the social distance (the reverse of social contact) is a factor that has been shown to negatively influence attitudes toward PWS and can have a contribution to several stereotypes, confusion and negative attitudes observed in this population (Betz, Blood, & Blood, 2008; Flynn & St. Louis, 2011; Klassen, 2002).

Similar to other populations (e.g., the Irish and the Arabic samples), some public confusion relating to the etiology of stuttering was observed. The Portuguese population believe that stuttering is caused by a very frightening event, an hypothesis which is rejected by a considerable body of research (Ambrose, Cox, & Yairi, 1997; Bloodstein & Bernstein Ratner, 2008; Cox et al., 2005; Guitar, 2014).

The item used to assess the amount of knowledge relating to stuttering fell in the interquartile, which means that, at this time, the amount of knowledge of the Portuguese sample was on average comparable to that observed in the total respondents of the POSHA–S data sample (percentile 40); this fact can interfere and influence the "(…) attitudes toward the individual affected" (de Britto Pereira, Rossi, & Van Borsel, 2008, p. 24) making the attitudes more realistic and more positive toward stuttering (Crowe & Walton, 1981; Louis et al., 2009; Yeakle & Cooper, 1986).

The radial graph (Fig. 3) confirms that the measured attitudes toward stuttering of the Portuguese population were visibly very similar to the median tracing of all the sample means in the POSHA–S database. Their Beliefs about stuttering were 34, their Self Reactions, 5, and their OSS, 19. Their Obesity/Mental Illness ratings were −34. As noted, these ratings were average to slightly more positive than average. Consistent with the database sample ratings, fully 73% of the percentiles shown in Table 4-3 for each Portuguese POSHA–S rating are in the interquartile range, with 7% and 20%, respectively, in the 1st and 4th quartiles.

Public attitudes toward stuttering in Portugal, in comparison to other database samples around the world that used the POSHA–S, were also roughly "average." Importantly, however, in comparison to the most relevant study of European attitudes (St.Louis et al., 2016), Portuguese attitudes were substantially more positive than measured attitudes in Italy (OSS = −3), slightly more positive than those in Germany (OSS = 15), slightly lower than Bosnia & Herzegovina or the combined English/Irish samples (OSS = 23), and substantially lower than the combined Norway/Sweden samples (OSS = 34). Additionally, the present probability sample also confirmed that regions within the Portugal mainland held similar attitudes, confirming what has been showed in three different regions of Bosnia & Herzegovina, Italy, and Norway (St.Louis et al., 2016).

The authors of the European study acknowledged that variation in sampling schemes may have influenced their results; however, it is extremely unlikely that the large differences (i.e., between the Scandinavian countries and Italy) were due to sampling differences. Given that demographic variables such as sex, age, SES did not predict the rank orders of ratings among the five countries or areas, national identity was suggested as a likely predictor variable. The authors did not speculate on how national identity might affect attitudes, i.e., why the Italian attitudes were substantially less positive than the other four countries and why the Scandinavian attitudes were substantially more positive. The current study does not address the question either; yet, it further brings into focus the need to include a measure of national identity (e.g., the International Social Survey (Rusciano, 2003)) in public attitude studies in different countries or perhaps measures of identity within countries. The European Social Survey (ESS) is an academic survey conducted every two years since 2001 to measure attitudes, beliefs and behaviours across more than thirty European countries. The information gathered by ESS showed that the Portuguese population (as the entire sample of European countries) consider themselves, in relation to human values, as "self-transcending" (benevolence and universalism) (Silva, 2011; Torres & Brites, 2006). This can be interpreted as meaning that "Europeans see themselves as supporters of values that stress help for others and loyalty to friends (benevolence) and they attach importance to equal opportunities, respect for differences and the protection of nature (universalism)" (Torres & Brites, 2006, p. 204). As human values could be defined as "abstract principles that guide and justify attitudes, opinions and behaviour" (Vala, Cabral, & Ramos, 2003), it can be inferred that the positive attitudes of the Portuguese population toward those who stutter could be influenced by the human values of "self-transcending" with which this population is most identified. Additionally, the concept of welfare society (Santos, 1995) (relations networks based on familial and neighbourly relations) could also

contribute to the positive attitudes observed in this sample. In Portugal, the welfare society plays a significant role in the coverage of social risk (to compensate an incipient welfare state) and it is based on three principles: solidarity, reciprocity and emotional investment. The emotional investment is revealed in exchanges that involve positive attitudes and an understanding, sympathetic, empathetic, encouraging and supportive environment. The attitudes in which the welfare society is based could influence positively the attitudes toward minority groups (e.g., PWS), as it is observed in the results of the present research study.

Most studies that used the POSHA–S have employed convenience sampling methodology. Only one study in Turkey compared probability sampling of adults using a public school-based cluster sampling scheme with a previous study using convenience sampling in Eskisehir, Turkey (Özdemir et al., 2011a, 2011b). Whereas OSSs were similar, the profiles of the various components and subscores were markedly different using the different sampling methods, leading the authors to conclude that probability sampling was the preferred procedure. Comparing the Portuguese results to adults in the two different probability samples (Özdemir et al., 2011a), summary stuttering attitudes for Portugal (OSS = 19) were more positive than those for Turkey (OSS = −2 and OSS = 4). Of course, the Turkish studies were not attempting to generalize to an entire country, but demographic difference were likely important determinants for the differences observed from the current study.

School-based samples are likely effective and efficient strategies to generate probability samples in specified areas, but country differences in public versus private schools, socio-economic influences on school populations, sizes and grade level included in school, and other factors such as access to schools in an increasingly dangerous world limit their use. By contrast, the current study confirms that in countries such as Portugal where adults are obliged to visit various local governmental offices for a wide variety of necessary life functions, a local administrative office-based sampling model provides an attractive, relatively inexpensive strategy to carry out probability sampling of adults in an entire country where these or similar government offices exist.

4.5.3 *POSHA-S* Predictors Variables of Portuguese Stuttering Attitudes

Several studies have attempted to explain differences in attitudes toward stuttering related to various demographic variables. In a detailed review of potential predictors (or correlates) of POSHA–S-measured attitudes, St. Louis (St. Louis, 2015a, 2015b) considered such variables as familiarity with stuttering, sex, age, socio-economic status (mostly likely

manifested as education, occupation, and income), geography (continent or country), language of the questionnaire, and religion. Geography, education, familiarity with stuttering appeared to emerge the most frequently from published and unpublished POSHA–S studies.

The current study did document that respondents characterized by younger or middle age, university education, living in the island regions of Portugal, not being retired, and speaking at least three different languages were all associated with more positive attitudes. By contrast, respondents' gender, marital status, and parental status were not. It should be noted that sex of the respondent has been mentioned often as a predictor of stuttering attitudes, perhaps because one of the first studies to consider it found a significant difference (Burley & Rinaldi, 1986). In fact, the study sampled only 10 male and 10 female respondents. Like Burley and Rinaldi (1986), some studies have shown better attitudes for females than males, especially those with very large sample sizes (e.g., Arnold, Li, & Goltl, 2015; Li & Arnold, 2015). Such studies typically have very small effect sizes, raising questions of how the findings might best be interpreted. Careful studies using the POSHA–S have shown, as the current study did, that sex of the respondent did not make a difference in the attitudes of adults toward stuttering (St. Louis, 2012a; St. Louis, LeMasters, & Poormohammad, 2015). Supporting the lack of relationship between Portuguese parental status and attitudes, a recent study by St. Louis, Weidner, and Mancini (2016) revealed that parents of young children, parents of older children or adult children, and non-parents held very similar attitudes on the Appraisal of the Stuttering Environment (ASE), a clinical instrument that is very similar to the second experimental version of the POSHA–S (St. Louis, Kuhn, & Lytwak, 2015).

More positive stuttering attitudes were associated with achieving higher levels in education, and less positive attitudes were associated with the ≥65yr age group. Comparisons with the two aforementioned probability samples in Turkey using the POSHA–S (Özdemir et al., 2011a), might shed light on these findings. The mean Portuguese age for the 311 respondents was 46 yr compared to 12 yr for the Turkish children, 37 and 38 yr for the parents, 45 and 62 yr for the grandparents/adult relatives (with far more aunts and uncles and far fewer grandparents in the first sample), and 35 and 36 yr for the neighbors. Mean years of school completed in Portugal was 10.3 yr and in Eskisehir, Turkey, 5.0 and 5.0 yr (children), 7.1 and 7.3 yr (parents), 4.8 and 7.0 yr (grandparents/adult relatives), and 7.6 and 8.2 yr (neighbours). It would be reasonable to assume that the differences in education would explain the much less positive attitudes in Eskisehir, Turkey sample compared to Portugal sample. However, the lack of large differences in attitudes as a function of age in

the Turkish sample suggests that age, per se, might not be a universal predictor of stuttering attitudes. Our analyses suggested that the amount of knowledge of stuttering was an important factor in the older and retired Portuguese age group. Knowledge about stuttering in the age group ≥65 is lower (although not significantly so) than the amount of knowledge of the younger age groups; within the ≥65 age group, the amount of knowledge variable influences negatively the Self reactions to PWS subscore [F(4, 84)= 3.865, p= 0.006], since people who reported having no knowledge about stuttering presented a significantly lower score than people who reported to know "some" (M=-20.13, p= 0.015) or a "great deal" (M=-39.66, p= 0.021). Retired people reported to know "a little" about PWS, which is less than "some" knowledge presented in the groups of unemployed people, working people or students. Within the retired group, the amount of knowledge also negatively influenced the Self reactions to PWS subscore [F (4, 81)= 2.506, p= 0.048], since reporting no knowledge about PWS negatively influences the Self reactions subscore, when compared with retired people who reported to know a "great deal" (M=-36.83, p= 0.028). Since misinformation/lack of information could influence attitudes, perceptions and beliefs toward PWS, the variable relating to knowledge could be interpreted as the reason for the significantly negative effect of aging and retirement.

MANOVA results revealed that the variable Region presented significant differences, which means that living in different regions of the country affects the attitudes/opinions/feelings of the population toward PWS. However, pair-wise comparisons showed that few districts presented significant differences, which means that "(…) whatever influenced ratings for one region of each country appeared to have very similar influence (…)" in other regions (St.Louis et al., 2016, p. 120).

Achieving higher levels in education or speaking more than one language positively influenced the attitudes toward PWS. Better scores could be a product of "(…) broadening of one's perspective though education or travel (…)", which could be gained in terms of knowledge about stuttering during the school years (Özdemir et al., 2011a, p. 330). As comparisons across samples in the POSHA–S database have not been carried out concerning the number of languages known, at this point it was not possible to explain why multilingualism appears to be associated with more positive stuttering attitudes. Though only significant for one versus three languages known, the means in Table 4-5 suggest that Self Reactions were most different between one or two languages known versus three or four. St. Louis and Roberts (2010) compared Canadian and Cameroonian respondents taking an experimental version of the POSHA–S either in English or French. Importantly, both are official languages in Canada and Cameroon. A control group of monolingual USA

respondents taking the POSHA–S English held attitude ratings much more similar to the Canadians than the Cameroonians, irrespective of the latter groups' knowledge and use of other languages

The variables marital status, parental status and native language did not have a significant influence for the subscores and for the OSS.

These predictor findings suggest that future studies of public attitudes adults in circumscribed regions consider age, education, work status, and number of languages known as covariates. Additionally, this study adds support to previous research showing that respondents' sex, parental status, or marital status are not consistent predictors of better or worse attitudes.

4.6 Conclusions

The present study analysed the public attitudes of the Portuguese population toward PWS, through a representative sample and based on a translation/cultural adaptation of POSHA–S to European Portuguese.

The translation was based on an exhaustive process, to guarantee the sensibility of the translated tool, in order to allow comparisons between the attitudes of the general Portuguese public and the POSHA–S total sample.

It can be concluded that the majority of the Portuguese attitudes toward PWS were "consistent with recent research (…) and/or more likely to reflect the greatest understanding and knowledge" (St. Louis, 2012c, p. 131). The score of 73% of the POSHA–S EP items fell into the interquartile, which means that the scores were about average, when compared with the entire POSHA–S database. Data collected through the probability sampling scheme allows the generalization of the findings, as this sample, collected through a three-stage cluster probability sampling represents the opinions of an entire country (Armitage, Berry, & Matthews, 2002; Özdemir et al., 2011a). The attitudes revealed by this population were influenced by the demographic variables age group, region, completed school levels, working situations and number of languages, given that these variables influenced the subscores and the Overall Stuttering Score.

One of the limitations of this study is related with the return rate; whereas it was determined to be satisfactory (Babbie, 2006), nearly one half of the individuals handed a questionnaire by staff members at the local administrative offices did not return them. It is possible that

those who did not fill them out would have held attitudes different than those who did. If so, then the generalizability would be limited the “typical” individual who not only avails him or herself to the functions of the office but also those who would be likely to fill out a questionnaire that they did not expect to receive. This is similar to questions that could be raised about virtually any study. A follow-up mixed-method study with interviews of a number of potential respondents who did and did not complete the questionnaire would be one way to estimate the effect of such potential selection bias. The absence of national identity characterisation to sustain and explain findings related to the predictor variable impact could be also viewed as a limitation of the current research study.

Accordingly, future studies should address the impact of national identity of the Portuguese sample (through the inclusion of standard measuring of national identity) to explain the relevant predictor variables found in the current study. Further, a similar, though not necessarily identical probability sampling procedure should be carried out consistently in several different countries to determine the extent to which probability sampling can distinguish differences between countries as convenience sampling has. If Portugal, Italy, and a Scandinavian country could be included, along with a carefully selected measure of national (or regional) identity, the reasons behind the differences that have been observed between countries might be elucidated. In future studies with substantial sample sizes (in probability or convenience sampling) additional predictors of positive and negative stuttering attitudes should be sought (e.g., knowledge of and acquaintance with stuttering).

Chapter 5 - Final conclusions

5.1. Introduction

In this Thesis, the ICF framework adaptation to the study of stuttering has been used to develop assessment instruments designed to evaluate *impairment in body functions* and *environmental factors*.

The *Severity Assessment Based on Events of Stuttering* (*SABES*) was developed to assess *impairment in body functions*, i.e., the characteristics of the observable stuttering events (quantitative and qualitative), based on different speech samples collected in an audio-visual format and analysed with a freeware annotation software. The *SABES* has been shown to be a valid (content and construct) instrument, based on 92 speech samples (23 AWS $\times$ four samples) in terms of frequency, duration, associated behaviours and tension degree. It was also concluded that *SABES* present good internal consistency.

Besides the measurement of the surface or observable stuttering behaviours, other measures should also be implemented during the assessment process of an AWS. The *Assessment of Language Use in Social Contexts for Adults* (*ALUSCA*) self-report instrument, developed to measure pragmatic language competencies, could enhance the accuracy of the assessment process of an AWS. The *ALUSCA* was developed to assess the effects of *environmental factors*, specifically the use of language in a difficult communicative situation. Twenty-three AWS and 23 adults who do not stutter (matched by age and gender) assessed the level of ease in using several pragmatic language competencies (PLC) on a previously selected difficult communicative situation, based on four variables that characterise communication (interlocutors, location, topic of conversation and communicative situation). Content and construct validity, internal consistency and test-retest reliability were determined. Results presented evidences of content and construct validity and good levels of internal consistency and test-retest reliability.

SABES and *ALUSCA* are two independent instruments, i.e., they are not part of the same assessment battery, as each one presented a specific protocol and specific validation data. However, to implement a comprehensive assessment, considering most of the dimensions of stuttering disorders, it is advisable to use both assessment tools. Results of the *SABES* and *ALUSCA* application on the assessment process could be used to develop intervention objectives for the *impairment of body functions* (*SABES*) and *environmental factors* (*ALUSCA*). Additionally, personal reactions and the impact of stuttering in quality of life

should also be addressed on an assessment/intervention process (Coleman & Yaruss, 2014).

The impact of *environmental factors*, determined by the attitudes, knowledge and beliefs of the interlocutors toward stuttering and PWS were also studied. A valid and reliable questionnaire, the *Public Opinion Survey of Human Attributes – Stuttering* (*POSHA-S*) was translated and cross-culturally adapted to European Portuguese to assess public opinions related to five human attributes (be obese, be left handed, has a stuttering disorder, be mentally ill and be intelligent), with more detailed concerning stuttering and PWS. A probability sampling scheme (cluster probability sampling) was used to collect a country-wide representative sample of the Portuguese population. Results of the translated questionnaire application to 311 respondents showed that the general Portuguese public have some knowledge about stuttering and a neutral overall impression of stuttering and PWS. However, results have shown an inaccurate attitude of the Portuguese sample and some incorrect knowledge concerning PWS traits. The demographic variables that influence attitudes were also studied and it was concluded that age group, region, completed school levels, working situations and number of languages are the variables influenced the *POSHA-S* subscores and the *Overall Stuttering Score*. The comparison of the *POSHA-S* Portuguese scores with the total sample revealed that the majority of the scores fell into the interquartile, meaning that they were about average. Data collected through the probability sampling scheme could be use during the assessment and/or intervention process to contextualise a PWS about the opinions of the society in which he/she is inserted, in order to understand the beliefs and attitudes that can influence stuttering symptoms and the associated reactions.

5.2. Clinical implications

Through the assessment process, the clinician sets out to measure and evaluate communication behaviours, in order to determine a diagnosis, to identify the need for referral, to determine the frequency of the treatment and to make decisions about the treatment structure (Shipley & McAfee, 2016). It is crucial to collect valid and reliable information, to be able to integrate, interpret and make correct and informed decisions that are sustained by meaningful and useful data. During the assessment process, the clinician needs to face an important and intrinsic challenge, i.e., the variability of stuttering. As stuttering is an unpredictable phenomenon, it is, by nature, variable. Both PWS and clinicians acknowledge that this fluency disorder presents variations in frequency and

duration of stuttering moments, intensity, attitudes, reactions and global impact. Stuttering is influenced by linguistic factors (e.g., semantic or syntax demands), paralinguistic factors (e.g., speaking rate), communicative situations or emotional demands (Constantino, Leslie, Quesal, & Yaruss, 2016). Those factors influence stuttering in a positive or negative way, leading to different degrees of variability. Therefore, someone who stutter face variations in stuttering from day to day or situation to situation. For a person who stutters, dealing with this variability may result in several associated feelings, such as hope or excitement in moments of increased fluency or hopelessness when stuttering moments increase or when facing the unknown of the occurrence of a stuttering moment. The inherent variability of stuttering could be demanding for a clinician during the assessment process, since the main objective is to capture the entire experience of an individual with the disorder. However, as stuttering changes and undergoes influences of several factors that could lead to a greater or less variation in stuttering symptoms in different days, different contexts and with different interlocutors, assessing the true comprehensive characteristics of this disorder is difficult.

The ICF framework (used in the current Thesis) is a representation of the multidimensionality of stuttering, reflecting also the different elements that contribute to the entire experience of stuttering and, consequently, to their variability. The clinician needs to be aware of this high variability characteristic, since stuttering does not always occur in regular or cyclic patterns, in order to find methods and strategies to be as representative as possible in the measurements performed. Thus, the ICF model of stuttering could be an important framework for a clinician to become conscious of different factors that should be taken into account during the assessment to better understand the uniqueness of a person's stuttering experience.

The assessment of an AWS should not rely exclusively on collecting speech overt characteristics. With specific adjustments to the individual characteristics of the client, the clinician should collect several types of information. During initial meetings with a PWS, the clinician has the opportunity to ascertain the level of motivation for the necessary changes in the therapeutic process, to acknowledge the individual history of stuttering, to explain the therapeutic process or to clarify myths, showing to the client an understanding and a non-judgement of his/her own story. To obtain a complete view and appreciation of the multidimensionality and variability of stuttering (and the individual characteristics of stuttering, with an individual's unique background and life choices) data concerning surface (i.e., the characteristics of the observable stuttering moments) and intrinsic features (i.e., personal reactions that an AWS could present concerning the experience with the fluency disorder) should be both collected (Manning, 2010; Yaruss & Quesal, 2006).

In an attempt to capture a representative sample of a specific PWS's experience of stuttering, an assessment process could use a variety of evaluation methods, including open-questions on a case history, observation of clients in different contexts and the application of informal or formal instruments. The instruments used in the process must be (Polit & Beck, 2012; Shipley & McAfee, 2016) valid, to allow the collection of data that are real and meaningful (i.e., the contents expressed on an instrument are representative of a content domain and the instrument assess a theoretical construct) and reliable, to allow the collection of accurate information, (i.e., data collected should be similar when the instrument is used on repeated measures and interpreted by different judges). The assessment tools developed and described in Chapter 2 and Chapter 3 contribute to a more reliable and valid assessment process that exists currently. Both *SABES* and *ALUSCA* could will support the process of assessing and evaluating more accurately the variability of stuttering and the interactions of the factors influencing it.

The *SABES* is an instrument developed to characterise the surface features of stuttering. This new instrument could be incorporated in an assessment process, to assess frequency, duration, associated behaviours tension degree of each stuttering moment and naturalness (quantitative assessment), and to classify the type of stuttering (qualitative assessment). The procedures developed allow the determination of valid outcome measures that can be used to obtain a benchmark (in initial meetings) to assess the effectiveness of the therapeutic process implemented. This also intends to contribute to a more specific and accurate severity estimation. Repeated baseline measures of an AWS prior to treatment could reassure to the clinician that data collected are more specific and closer to the true indication of the problem (Manning, 2010). For Portugal, to the best or our knowledge, there was no valid instrument to assess surface features of AWS prior to *SABES*. In an international perspective, *SABES* will also contribute to the strengthening of the assessment process validity, with the use of several speech samples and an analysis processes that are not restricted to the use of auditory perception; however, the severity assessment protocol needs to be validated in other languages, in order to obtain severity levels to compare with.

The use of several speech samples collected in audio and video support, analysed with an annotation software will contribute to the improvement on the accuracy of measurements. Despite the continuous effort to overcome difficulties with the severity assessment of stuttering, the clinician needs to be aware that severity estimation is a continuous process that goes on until the early stages of intervention. The clinician needs to become familiar

with the communication patterns of the client and his unique way of stutter to capture, in a valid manner, the true nature of the person's stuttering.

A way to capture the variability and uniqueness of stuttering, is the inclusion of information from self-assessment measures made by the AWS, both for severity of surface, cognitive and affective features and also environmental/contextual influences. Self-assessments are a powerful method to capture the perspective of the individual. The individual with a stuttering disorder is able to provide the most informative and accurate information concerning a specific construct. The *ALUSCA* is a self-report questionnaire, developed to assess the use of language (i.e., pragmatic competencies) of adults. During a common assessment process, it is not usual to include pragmatic competencies evaluation. However, it seems logical to think about the pragmatic influence on stuttering as linguist factors influence stuttering variability. Since different communicative contexts presented different demands and those influence the level of fluency in an individual, it seems rational that use language in speaking situations with an underlying variability that could be difficult for those who stutter. Additionally, and although the relationship and the impact of pragmatics on stuttering is not well stablished, some researchers highlight the influence of pragmatic demands on fluency and on psychological outcomes of PWS (Cummings, 2014; Tanner et al., 1999). Thus, the influence of environmental factors, specifically pragmatic language demands, should be included to determine their impact and importance and to provide a focus for therapy.

The exhaustive validity analysis lead to a questionnaire that is broad in the content and construct, and that can capture the respondent's perception on the use of language (e.g., nonverbal communication or initiation of a conversation topic) in a difficult communicative context. The *ALUSCA* is composed by two parts, in which the first one contextualises the focus of pragmatics and the features that characterise a communicative interaction. Thus, the person become familiarised about difficult concepts, to assist in the valid assessment of the ease level of pragmatic competencies (in the second part of the questionnaire). Similarly to *SABES*, the *ALUSCA* questionnaire could be included in the assessment process; specifically related to the questionnaire, it allows the determination of a profile of the individual's pragmatic competencies. The *ALUSCA* will help the clinicians outline a profile of the ease of pragmatic competencies, to obtain a baseline that will help in the decision making process, related to the inclusion of specific objectives of pragmatics in the intervention process.

Due to the response bias, lack of self-knowledge to assess a certain construct of all self-reports, clinicians must integrate their own observation of the person's language use, to verify the impact of stuttering on pragmatic, contributing to the collection of data that are more valid.

Assessment of someone who stutters in a broader perspective also includes the determination of the influence of the interlocutors, or, in a wider view, the society in which the person lives. Over the past years several research findings have shown that public stigma (i.e., negative attitudes from the public, concerning to prejudice and discrimination) can become internalised by individuals (self-stigma). A PWS experiencing internalised stigma presents associated reactions that could lead to different consequences for communicative participation, since avoidance of specific speaking situations and life choices are conditioned by stuttering. Because communication is a social phenomenon in which people speak to each other in different situations and about different topics, the society attitudes and reactions can contribute to variations in stuttering characteristics (i.e., different social attitudes can lead to differences in stuttering moments characteristics) and cause self-limitations and self-restriction attitudes for someone who stutters. Those affective factors, that are socially moderated, conditioned participation in everyday communication exchanges (from simple to more complex situations) and influence, in a broader perspective, the quality of life.

Public stigma is a reality and, for that reason, calls for actions are decisive to mitigate it. The information obtained through the application of the translated and cross-culturally adapted *POSHA-S* questionnaire with a probability sampling approach that allow the generalisation of findings, could be used to developed public campaigns to raise awareness and improve public attitudes toward stuttering. Concerning an individual who stutters and his unique assessment process, the information obtained from the respondents of *POSHA-S* EP version could be used to inform and to improve the understanding of their situation in a broader perspective. During the therapeutic process, the clinician might communicate the findings of the Portuguese attitudes study, to improve the knowledge and comprehension of the individual concerning public attitudes, to relativize them by moving away from focus excessively on the society thoughts. Specifically concerning the results obtained on the present work, Portuguese AWS have the advantage of knowing that the majority of listener's attitudes are positive, which could contribute to the decrease of self-imposed limitations on communication. For those individuals who stutter that are severely focused on the society thoughts and opinions toward stuttering, to give concrete information collected in a whole country could be a powerful begin of the desensitized process of other's communication

attitudes. The desensitisation process that can be achieved through this information exchange can improve the internalised view of stuttering nature, by putting the client's stuttering into a wider perspective. The generalisation process of the skills learned in therapy to extra-clinic contexts could also be improved by the knowledge of the society attitudes, since the person gains an advantage (and therefore makes the process easier) if he/she have prior knowledge about those attitudes and prepare, in advance, to an alternative way of coping with them.

5.3. Limitations of the present work

The development of the present research presents some limitations. Concerning the development of the *SABES* and *ALUSCA*, the small sample size of the pilot study studies could be pointed out as a limitation. The test-retest calculation of *ALUSCA* only has been performed with controls, which could also be interpreted as a limitation. Concerning the collected data process with *POSHA-S*, there was an absence of national identity characterisation to sustain and explain findings related to the predictor variable impact.

5.4. Future work

Future research with the *SABES* should include additional reliability analysis, specifically interjudge and intrajudge, to verify the equivalence of scoring. A study of the usability of the *Excel* spreadsheets developed should also be considered on a future research, to validate their usefulness on the application of *SABES* protocol to assess the characteristics of stuttering symptoms.

Future research with the *ALUSCA* questionnaire should include assessment of avoidance on AWS and the correlation with the level of ease in performing pragmatic competencies, as some authors (Tanner et al., 1999) have previously shown the avoidance of speaking situations that are common in some AWS (due to anxiety associated with communicative situations) could affect language experience and the pragmatic skills development. The determination of impact ratings to compare the ALUSCA total score of an AWS is also considered as future work with this questionnaire.

Results from the *SABES* and *ALUSCA* could also be correlated in future research studies, to analyse the relationship between the level of ease in use language and the severity of the observable stuttering behaviours.

Based on the limitations pointed to the research study using *POSHA-S*, future studies should consider the impact of national identity of the Portuguese sample to explain the predictor variables found.

Author's relevant publications

Valente, A.R.S., St. Louis, K. O., Leahy, M., Hall, A. & Jesus, L.M.T. (2017). A country-wide probability sample of public attitudes toward stuttering in Portugal. *Journal of Fluency Disorders*, *52*, 37–52.

Valente, A., Alvelos, H., Leahy, M. & Jesus, L.M.T. (2016). Construct Validity Evidence of the Assessment of Language Use in Adults (ALUSCA). *In* Proceedings of the American Speech-Language-Hearing Association Convention 2016 (ASHA Convention 2016), Philadelphia, USA, p. 208.

Valente, A.R.S., Jesus, L.M.T., Hall, A. & Leahy, M. (2016). Developing a New Stuttering Assessment Tool: Severity Assessment Based on Events of Stuttering (SABES). In Proceedings of the 30th World Congress of the International Association of Logopedics and Phoniatrics (IALP 2016), Dublin, Ireland, p. 149.

Valente, A.R.S., Jesus, L.M.T., Hall, A. & Leahy, M. (2015). Event- and interval-based measurement of stuttering: a review. *International Journal of Language and Communication Disorders,* 50 (1), 14-30. DOI: 10.1111/1460-6984.12113;

Valente, A.R.S., Jesus, L.M.T., Roberto, M.T., Leahy, M. & St. Louis, K.O. (2015). Cross-Cultural Adaptation of the Public Opinion Survey of Human Attributes (POSHA-S): The European Portuguese (EP) Translation and Adaptation Process as a Model. In *Stuttering: Risk Factors, Public Attitudes and Impact on Psychological Well-Being* (pp. 93-110). ISBN: 978-1-63483-695-1.

Valente, A.R.S., Jesus, L.M.T. Leahy, M. & Hall, A. (2015). Developing Assessment of Language Use in Social Context for Adults. In Proceedings of the 8th IFA World Congress on Fluency Disorders (IFA Congress 2015), Lisboa Portugal.

Jesus, L.M.T., Valente, A.R.S. & HALL, A. (2014). Is the Portuguese version of 'The North Wind and the Sun' passage phonetically balanced? *Journal of the International Phonetic Association*, 45, pp 1-11 doi:10.1017/S0025100314000255.

Valente, A.R.S., Jesus, L.M.T., Leahy, M. & K. St Louis (2014). Attitudes and knowledge of the Portuguese population about stuttering. In Proceedings of the 5th European Symposium on Fluency Disorders, Antwerp, Belgium.

Author's relevant oral presentations in international conferences

Valente, A.R.S., Jesus, L.M.T., Leahy, M. & Hall, A. (2015). Assessment of Language Use in Social Contexts for Adults: Establishing content validity. 8th World Congress on Fluency Disorders. Lisboa, 6-8 July.

Valente, A.R.S., Jesus, L.M.T., Leahy, M. & St. Louis, K. (2014). Attitudes and Knowledge of the Portuguese population about stuttering. European Symposium on Fluency Disorders. March 2014, Antwerp, Belgium.

Registered (Trademark and Copyright) Health Assessment Tools

Jesus, L.M.T. & Valente, A.R.S. (2015). Avaliação da Severidade Baseada em Eventos de Gaguez / Severity Assessment Based on Events of Stuttering (SABES). Universidade de Aveiro, Portugal. Deferimento pela IGAC em 21/12/2015 – No 4936/2015;

Jesus, L.M.T. & Valente, A.R.S., (2015). Avaliação do Uso da Linguagem em Contexto para Adultos / Assessment of Language Use in Social Context for Adults (ALUSCA). Universidade de Aveiro, Portugal. Deferimento pela IGAC em 21/12/2015 – No 4937/2015.

Appendix 1

Ethical committee approval

COMISSÃO DE ÉTICA

da **Unidade Investigação em Ciências da Saúde - Enfermagem** (UICISA-E)

da **Escola Superior de Enfermagem de Coimbra** (ESEnfC)

Parecer Nº P137-02/2013

Título do Projecto: Desenvolvimento de Instrumentos de Medida para a Avaliação de Adultos com Disfluência

Identificação do(s) Proponente(s)
Nome(s): Luís Miguel Teixeira de Jesus, Ana Rita dos Santos Valente
Filiação Institucional: Universidade de Aveiro e Instituto de Engenharia Electrónica e Telemática de Aveiro (IEETA)
Orientador: Luís Miguel Teixeira de Jesus

Relator(es): Maria Filomena Botelho

Parecer

O projecto pretende avaliar quanto à fiabilidade e validade instrumentos para analisar os comportamentos observáveis de pessoas com gaguez, bem como as reacções, atitudes e opiniões do público em geral relativamente à gaguez. O instrumento de avaliação desenvolvido foi adequado à realidade portuguesa e contempla o pressuposto teórico em relação à avaliação da incapacidade (comportamentos observáveis da gaguez) e aos factores ambientais (reacções, atitudes e opiniões do público relativamente à gaguez e dificuldade de comunicar em situações do dia-a-dia).

Trata-se de um estudo descritivo transversal. A população alvo será constituída por pessoas com gaguez (cerca de 150 pessoas) e população em geral (cerca de 90 pessoas).

Os critérios de inclusão e de exclusão estão claramente definidos. Existe garantia de confidencialidade.

Atendendo ao formato da investigação, a Comissão de Ética dá o seu parecer favorável.

O relator: Maria Filomena Botelho

Data: 13 / 03 / 2013 O Presidente da Comissão de Ética: _____

UNIDADE DE INVESTIGAÇÃO EM CIÊNCIAS DA SAÚDE

escola superior de enfermagem de coimbra

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR Portugal

Appendix 2

Consent Form

DECLARAÇÃO DE CONSENTIMENTO

Considerando a "Declaração de Helsínquia" da Associação Médica Mundial
(Helsínquia 1964; Tóquio 1975; Veneza 1983; Hong Kong 1989; Somerset West 1996; Edimburgo 2000; Washington 2002; Tóquio 2004; Seoul 2008)

Desenvolvimento de Instrumentos de Medida para a Avaliação de Adultos com Disfluência

Eu, abaixo-assinado,

_____, compreendi a explicação que me foi fornecida acerca do meu caso clínico e da investigação que se tenciona realizar, bem como do estudo em que serei incluído. Foi-me dada oportunidade de fazer as perguntas que julguei necessárias, e de todas obtive resposta satisfatória.

Tomei conhecimento de que, de acordo com as recomendações da Declaração de Helsínquia, a informação ou explicação que me foi prestada versou os objetivos, os métodos, os benefícios previstos, os riscos potenciais e o eventual desconforto. Além disso, foi-me afirmado que tenho o direito de recusar a todo o tempo a minha participação no estudo, sem que isso possa ter como efeito qualquer prejuízo na assistência que me é prestada.

Eu compreendo que os resultados do estudo podem ser publicados em revistas científicas, apresentados em conferências e usados noutras investigações, sem que haja qualquer quebra de confidencialidade. Portanto, dou autorização para a utilização dos dados para esses fins.

Por isso, consinto que me seja aplicado o método, o tratamento ou o inquérito proposto pelo investigador.

Data: ____ / _____ / ____

Assinatura do doente ou voluntário:

_____

O Investigador responsável:

Nome: Ana Rita dos Santos Valente

Assinatura:

References

Abdalla, F. A., & St. Louis, K. O. (2012). Arab school teachers' knowledge, beliefs and reactions regarding stuttering. *Journal of Fluency Disorders*, *37*, 54–69.

Al-Khaledi, M., Lincoln, M., McCabe, P., Packman, A., & Alshatti, T. (2009). The attitudes, knowledge and beliefs of Arab parents in Kuwait about stuttering. *Journal of Fluency Disorders*, *34*, 44–59.

Allen, G. (1975). Speech rhythm: Its relation to performance universals and articulatory timing. *Journal of Phonetics*, *3*, 75–86.

Alm, P. A. (2014). Stuttering in relation to anxiety, temperament, and personality: Review and analysis with focus on causality. *Journal of Fluency Disorders*, *40*, 5–21.

Ambrose, N. G., Cox, N. J., & Yairi, E. (1997). The genetic basis of persistence and recovery in stuttering. *Journal of Speech, Language and Hearing Research*, *40*, 567–580.

Ambrose, N. G., & Yairi, E. (1999). Normative Disfluency Data for Early Childhood Stuttering. *Journal of Speech, Language and Hearing Disorders*, *42*, 895–909.

American Educational Research Association (AERA). (2014). *The Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association (AERA).

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5)* (5th ed.). Washington, DC: American Psychiatric Publishing.

Andrews, G., & Cutler, J. (1974). Stuttering therapy: The relationship between changes in symptom level and attitudes. *Journal of Speech and Hearing Disorders*, *38*, 312–319.

Antipova, E. A., Purdy, S. C., Blakeley, M., & Williams, S. (2008). Effects of altered auditory feedback (AAF) on stuttering frequency during monologue speech production. *Journal of Fluency Disorders*, *33*, 274–290.

Armitage, P., Berry, G., & Matthews, J. (2002). *Statistical Methods in Medical Research* (4th ed.). Oxford: Blackwell Science.

Armson, J., & Kiefte, M. (2008). The effect of SpeechEasy on stuttering frequency, speech rate, and speech naturalness. *Journal of Fluency Disorders*, *33*, 120–134.

Armson, J., Kiefte, M., Mason, J., & De Croos, D. (2006). The effect of SpeechEasy on stuttering frequency in laboratory conditions. *Journal of Fluency Disorders*, *31*, 137–152.

Armson, J., & Stuart, A. (1998). Effect of Extended Exposure to Frequency-Altered Feedback on Stuttering During Reading and Monologue. *Journal of Speech, Language, and Hearing Research*, *41*, 479–490.

Arnold, H. S., Li, J., & Goltl, K. (2015). Beliefs of teachers versus non-teachers toward people who stutter. *Journal of Fluency Disorders*, *43*, 28–39.

Aron, M. L. (1967). The relationship between measurements of stuttering behaviour. *Journal of South African Logopedic Society*, *14*, 15–34.

ASHA. (1993). Definitions of communication disorders and variations [Relevant Paper]. Retrieved from Available from www.asha.org/policy

ASHA. (2014). Assessment of Fluency Disorders In the Context of the WHO ICF Framework. Retrieved from http://www.asha.org/Practice-Portal/Clinical-Topics/Childhood-Fluency-Disorders/Assessment-of-Fluency-Disorders-In-the-Context-of-the-WHO-ICF-Framework/

Asher, M. (2010). A different way to view probability densities. Retrieved April 1, 2017, from http://www.statisticsblog.com/2010/06/a-different-way-to-view-probability-densities/

Associação Portuguesa de Gagos. (2014). Associação Portuguesa de Gagos. Retrieved from http://gaguez-apg.com/

Aydın, Ç. (2008). *Kekemeliğe yönelik toplumsal farkındalığın belirlenmesi (Determining awareness toward stuttering). Unpublished Master's Thesis.* Anadolu University, Eskişehir, Turkey.

Babbie, E. (2006). *The practice of social research* (11th ed.). Belmont: Wadsworth

Publishing.
Baer, D. M., Wolf, M. M., & Rlsley, T. R. (1987). Some still-current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *20*, 313–327.
Beaton, D., Bombardier, C., Guillemin, F., & Ferraz, M. B. (1998). Recommendations for the Cross-Cultural Adaptation of Health Status Measures. *American Academy of Orthopedaedic Surgeons and Institute for Work & Health*.
Beaton, D., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the Process of Cross-Cultural Adaptation of Self-Report Measures. *Spine*, *25*(24), 3186–3191.
Beilby, J. M., Byrnes, M. L., & Yaruss, J. S. (2012). Acceptance and Commitment Therapy for adults who stutter: Psychosocial adjustment and speech fluency. *Journal of Fluency Disorders*, *37*, 289–299.
Bernstein, N. E. (1981). Are there constraints on childhood disfluency? *Journal of Fluency Disorders*, *6*, 341–350.
Bernstein Ratner, N. (1997). Stuttering: A psycholinguistic perspective. In G. Curlee, R.; Siegel (Ed.), *Nature and treatment of stuttering: New directions* (pp. 99–127). Needham, MA: Allyn & Bacon.
Betz, I. R., Blood, G. W., & Blood, I. M. (2008). University students' perceptions of pre-school and kindergarten children who stutter. *Journal of Communication Didorders*, *41*, 259–273.
Bland, J. M., & Altman, D. G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, *1*, 307–310.
Bland, J. M., & Altman, D. G. (1997). Cronbach's alpha. *BMJ (Clinical Research Ed.)*, *314*(7080), 572. http://doi.org/10.1136/bmj.314.7080.572
Bleek, B., Reuter, M., Yaruss, J., Cook, S., Faber, J., & Montaga, C. (2012). Relationships between personality characteristics of people who stutter and the impact of stuttering on everyday life. *Journal of Fluency Disorders*, *37*, 325–333.
Blood, G. W., & Blood, I. M. (2016). Long-term Consequences of Childhood Bullying in Adults who Stutter: Social Anxiety, Fear of Negative Evaluation, Self-esteem, and Satisfaction with Life. *Journal of Fluency Disorders*, *50*, 72–84.
Bloodstein, O. (2006). Some empirical observations about early stuttering: A possible link to language development. *Journal of Communication Disorders*, *39*(3), 185–191.
Bloodstein, O., & Bernstein Ratner, N. (2008). *A Handbook On Stuttering* (6th ed.). Clifton Park: Delmar.
Blumgart, E., Tran, Y., & Craig, A. (2010). Social anxiety disorder in adults who stutter. *Depression and Anxiety*, *27*(7), 687–692.
Boey, R. A., Wuyts, F. L., Van de Heyning, P. H., De Bodt, M. S., & Heylen, L. (2007). Characteristics of stuttering-like disfluencies in Dutch-speaking children. *Journal of Fluency Disorders*, *32*(4), 310–329.
Bothe, A. K. (2004). *Evidence-Based Treatment of Stuttering Empirical Bases and Clinical Applications*. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.
Boyle, M. P. (2013). Assessment of Stigma Associated With Stuttering: Development and Evaluation of the Self-Stigma of Stuttering Scale (4S). *Journal of Speech, Language and Hearing Disorders*, *56*, 1517–1529.
Boyle, M. P., & Blood, G. W. (2015). Stigma and stuttering: Conceptualizations, applications, and coping. In K. O. St. Louis (Ed.), *Stuttering meets stereotype, stigma, and discrimination: An overview of attitude research.* (pp. 43–70). Morgantown, WV: West Virginia University Press.
Boyle, M. P., Dioguardi, L., & Pate, J. E. (2016). A comparison of three strategies for reducing the public stigma associated with stuttering. *Journal of Fluency Disorders*, *50*, 44–58.
Brandsma, J. W., Lakerveld-Heyl, K., Van Ravensberg, C. D., & Heerkens, Y. F. (1995). Reflection on the definition of impairment and disability as defined by the World Health

Organization. *Disability and Rehabilitation*, *17*(3–4), 119–127.
Brundage, S., Bothe, A., Lengeling, A. K., & Evans, J. (2006). Comparing judgments of stuttering made by students, clinicians, and highly experienced judges. *Journal of Fluency Disorders*, *31*, 271–283.
Brutten, G., & Shoemaker, D. (1974). Speech situation checklist. Carbondale, IL: Speech Clinic: Southern Illinois University.
Bryant, K., & Packman, A. (1999). The Lidcombe Behavioral Data Language: A new way of talking about stuttering. In S. McLeod & L. McAllister (Eds.), *1999 Speech Pathology Australia National Conference* (pp. 191–196). Melbourne, Australia: Speech Pathology Australia.
Burley, P. M., & Rinaldi, W. (1986). Effects of sex of listener and of stutterer on ratings of stuttering speakers. *Journal of Fluency Disorders*, *17*, 329–333.
Butler, C. (2013). Identity and stammering: Negotiating hesitation, side-stepping repetition: And sometimes avoiding deviation. *Sociology of Health & Illness*, *35*, 1113–1127.
Campbell, J., & Hill, D. (1987). Systematic disfluency analysis: Accountability for differential evaluation and treatment, Miniseminar presented to the Annual Convention of the American Speech-Language-Hearing Association. New Orleans, LA.
Campbell, J., Hill, D., & Driscoll, M. (1991). Systematic Disfluency Analysis: Using SDA to determine stuttering severity. Poster paper presented to the Annual Convention of the American Speech-Language-Hearing Association. Anaheim, CA: ASHA.
Chang, S. E., Erickson, K. I., Ambrose, N. G., Hasegawa-Johnson, M. A., & Ludlow, C. L. (2008). Brain Anatomy Differences in Childhood Stuttering. *Neuroimage*, *39*(3), 1333–1344.
Coleman, C., & Yaruss, J. S. (2014). A comprehensive view of stuttering: Implications for assessment and treatment. *SIG 16 Perspectives on School-Based Issues*, *15*(2), 75–80.
Constantino, C. D., Leslie, P., Quesal, R. W., & Yaruss, J. S. (2016). A preliminary investigation of daily variability of stuttering in adults. *Journal of Communication Disorders*, *60*, 39–50.
Conture, E. G. (1990a). Childhood stuttering: What is it and who does it? *ASHA Reports*, *18*, 2–14.
Conture, E. G. (1990b). *Stuttering* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
Cooper, C. S., & Cooper, E. B. (1985). *Cooper personalized fluency control therapy revised*. USA: DLM Teaching Resources.
Cooper, E. B., & Cooper, C. S. (1996). Clinician attitudes towards stuttering: Two decades of change. *Journal of Fluency Disorders*, *21*(2), 119–135.
Cordes, A. K., Ingham, R., Frank, P., & Ingham, J. (1992). Time-Interval Analysis of Interjudge and Intrajudge Agreement for Stuttering Event Judgment. *Journal of Speech and Hearing Research*, *35*, 483–494.
Çoşkun, M. (2006). *Kekemelik ve hızlı-bozuk konus¸ma hakkında farkındalıgın˘ betimlenmesi (Describing awareness about stuttering and cluttering)*. Anadolu University, Eskis¸ ehir, Turkey.
Costello, J. M., & Ingham, R. J. (1984). Assessment strategies for stuttering. In R. F. Curlee (Ed.), *Nature and treatment of stuttering: New directions*. San Diego: College-Hill Press.
Cox, N., Roe, C., Suresh, R., Cook, E., Lundstrom, C., Garsten, M., … Yairi, E. (2005). Chromosomal signals for genes underlying stuttering. *Oxford Disfluency Conference*. Oxford University, Oxford, United Kingdom.
Coyle, M. M., & Mallard, A. R. (1979). Word-by-word analysis of observer agreement utilizing audio and audiovisual techniques. *Journal of Fluency Disorders*, *4*, 23–28.
Craig, A., Blumgart, E., & Tran, Y. (2009). The impact of stuttering on the quality of life in adults who stutter. *Journal of Fluency Disorders*, *34*(2), 61–71. http://doi.org/10.1016/j.jfludis.2009.05.002

Cream, A., Onslow, M., Packman, A., & Llewellyn, G. (2003). Protection from harm: the experience of adults after therapy with prolonged-speech. *International Journal of Language and Communication Disorders*, *38*, 379–395.
Crichton, N. (2001). Visual Analogue Scale (VAS). *Journal of Clinical Nursing*, *10*, 697–706.
Crowe, T. A., & Cooper, E. B. (1977). Parental attitudes toward and knowledge of stuttering. *Journal of Communication Didorders*, *10*(4), 343–357.
Crowe, T. A., & Cooper, E. B. (1983). Vocational rehabilitation counselors' attitudes toward stuttering. *Journal of Fluency Disorders*, *8*(1), 13–27.
Crowe, T. A., & Walton, J. H. (1981). Teacher attitudes toward stuttering. *Journal of Fluency Disorders*, *6*, 163–174.
Crystal, D. (1987). Towards a "bucket" theory of language disability: Taking account of interaction between linguistic levels. *Clinical Linguistics and Phonetics*, *1*, 7–22.
Cullinan, W., Prather, E., & Williams, D. (1963). Comparison of procedures for scaling severity of stuttering. *Journal of Speech and Hearing Research*, *6*, 187–194.
Cummings, L. (2014). Pragmatic Disorders. In *Perspectives in Pragmatics, Philosophy & Psychology*. Dordrecht: Springer.
Curlee, R. F. (1981). Observer agreement on disfluency and stuttering. *Journal of Speech and Hearing Research*, *24*, 595–600.
Curlee, R. F. (1993). Evaluating treatment efficacy for adults: Assessment of stuttering disability. *Journal of Fluency Disorders*, *18*, 319–332.
Curran, M. F., & Hood, S. B. (1977). Listener Ratings of Severity for Specific Disfluency Types in Children. *Journal of Fluency Disorders*, *2*, 87–97.
Cuthbert, B. N. (2002). Social Anxiety Disorder: Trends and Translational Research. *Biological Psychiatry*, *153*, 1367–1368.
Daly, C., & Leahy, M. (2014). Public Attitudes Towards Stuttering: an Epidemiological Study of an Irish Popultion. *European Symposium on Fluency Disorders*. Antwerp, Belgium.
Davis, D. M. (1940). The relation of repetitions in the speech of young children to certain measures of language maturity and situational factors: Parts II & III. *Journal of Speech Disorders*, *5*, 235–246.
de Britto Pereira, M. M., Rossi, J. P., & Van Borsel, J. (2008). Public awareness and knowledge of stuttering in Rio de Janeiro. *Journal of Fluency Disorders*, *33*, 24–31.
Dewart, H., & Summers, S. (2003). *The pragmatic profile of everyday communication skills - adults*.
Doody, I., Kalinowski, J., Armson, J., & Stuart, A. (1993). Stereotypes of stutterers and nonstutterers in three rural communities in Newfoundland. *Journal of Fluency Disorders*, *18*, 363–373.
Dorsey, M., & Guenther, R. K. (2000). Attitudes of professors and students toward college students who stutter. *Journal of Fluency Disorders*, *25*, 77–83.
Einarsdóttir, J., & Ingham, R. (2005). Have disfluency-type measures contributed to the understanding and treatment of developmental stuttering? *American Journal of Speech-Language Pathology*, *14*, 231–243.
Emerick, L. L. (1960). Extensional definition and attitude toward stuttering. *Journal of Speech and Hearing Research*, *3*, 181–186.
Enderly, P., & John, A. (2015). *Therapy Outcome measures for rehabilitation professionals* (3rd ed.). Guildford: J & R Press Ltd.
EUROSTAT. (2011). Regions in the European Union: Nomenclature of territorial units for statistics NUTS 2010/EU-27. Luxembourg: European Union.
Evans, D., Healey, E. C., Kawai, N., & Rowland, S. (2008). Middle School students' perceptions of a peer who stutter. *Journal of Fluency Disorders*, *33*(3), 203–219.
Eve, C. L., Onslow, M., Andrews, C., & Adams, R. (1995). Clinical Measurement of Early Stuttering Severity: The Reliability of a 10-point Scale. *Australian Journal of Human Communication Disorders*, *23*(2), 26–39.
Feinstein, A. R. (1987). The theory and evaluation of sensibility. In A. R. Feinstein (Ed.),

Clinimetrics (pp. 141–166). New Haven: Yale University Press.
Fey, M. (1986). *Language Intervention with Young Children*. San Diego: College-Hill Press.
Field, A. (2013). *Discovering statistics using IBM SPSS Statistics: and sex and drugs and rock "n" roll* (Sage publi). London.
Flipsen Jr, P., Hammer, J. B., & Yost, K. M. (2005). Measuring Severity of Involvement in Speech Delay: Segmental and Whole-Word Measures. *American Journal of Speech-Language Pathology*, *14*, 298–312.
Flynn, T. W., & St. Louis, K. O. (2011). Changing adolescent attitudes toward stuttering. *Journal of Fluency Disorders*, *15*, 59–273.
Foundas, A. L., Bollich, A. M., Corey, D. M., Hurley, M., & Heilman, K. M. (2001). Anomalous anatomy of speech-language areas in adults with persistent developmental stuttering. *Neurology*, *57*(2), 207–215.
Gaines, N. D., Runyan, C. M., & Meyers, S. C. (1991). A comparison of young stutterers' fluent versus stuttered utterances on measures of length and complexity. *Journal of Speech and Hearing Research*, *34*, 37–42.
Geisinger, K. (1994). Cross-Cultural Normative Assessment: Translation and Adaptation Issues Influencing the Normative Interpretation of Assessment Instruments. *Psychological Assessment*, *6*(4), 304–312.
Gillam, R. B., Logan, K. J., & Pearson, N. A. (2009). *Test of Childhood Stuttering*. Austin, Texas: Pro-Ed Inc.
Goffman, E. (1963). *Stigma: Notes on management of a spoiled identity*. Englewood Cliff, NJ: Prentice-Hall.
Goldman, L. (1971). *Using tests in counseling* (2nd ed.). Pacific Palisades, CA: Goodyear Publishing.
Gordis, L. (2014). *Epidemiology* (5th ed.). Philadelphia, PA: Elsevier Saunders.
Grant, J. S., & Davis, L. L. (1997). Selection and Use of Content Experts for Instruments Development . *Research in Nursing & Health*, *20*, 269–274.
Gregory, H. H. (1986). *Stuttering: Differential evaluation and therapy*. Austin, TX: Pro-Ed.
Gregory, H. H. (2003). *Stuttering therapy: Rationale and procedures. Boston, MA: Allyn & Bacon.* Boston, MA: Allyn & Bacon.
Gregory, H. H., Campbell, J. H., & Hill, D. G. (2003). Differential evaluation of stuttering problems. In H. H. Gregory (Ed.), *Stuttering therapy: Rationale and procedures* (pp. 80–141). Boston: Allyn & Bacon.
Gregory, H. H., & Hill, D. (1993). Differetial evaluation-differential therapy for stuttering children. In R. F. Curlee (Ed.), *Stuttering and related disorders of fluency* (pp. 23–44). New York: Thieme Medical Publishers.
Griffin, M., & Leahy, M. M. (2007). Young children's (3-5 years old) perceptions of a dysfluent speaker. In J. Au-Yeung & M. M. Leahy (Eds.), *Fifth World Congress on Fluency Disorders* (pp. 223–229). International Fluency Association.
Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: literature review and proposed guidelines. *Journal of Clinical Epidemiology*, *46*(12), 1417–1432.
Guitar, B. (2014). *Stuttering: An Integrated Approach to Its Nature and Treatment* (4th ed.). Baltimore: Lippincott Williams & Wilkins.
Hair Jr, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis* (7th ed.). Upper Saddle River: Prentice Hall.
Ham, R. E. (1989). What are we measuring? *Journal of Fluency Disorders*, *14*, 231–243.
Hargrave, S., Kalinowski, J., Stuart, A., Armson, J., & Jones, K. (1994). Effect of Frequency-Altered Feedback on Stuttering Frequency at Normal and Fast Speech Rates. *Journal of Speech and Hearing Research*, *37*, 1313–1319.
Hartford, E., & Leahy, M. M. (2007). The perceptions of the primary school children of a person who stutters. In J. Au-Yeung & M. M. Leahy (Eds.), *Fifth World Congress on Fluency Disorders* (pp. 223–229). International Fluency Association.

Hasson, D., & Arnetz, B. B. (2005). Validation and Findings Comparing VAS vs. Likert Scales for Psychosocial Measurements. *International Electronic Journal of Health Education*, *8*, 178–192.
Healey, E. C. (1991). Assessment review: Stuttering Severity Instrument for Children and Adults (revised). *Journal of Fluency Disorders*, *12*, 167–173.
Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (5th ed.). Boston: Houghton Mifflin.
Hopkins, W. G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, *30*, 1–15.
Hopkins, W. G. (2002). A scale of magnitudes for effect statistics.
Horn, L. R., & Ward, G. (2006). *The Handbook of Pragmatics*. Clackwell Publishing.
Howell, P. (2004). Assessment of some contemporary theories of stuttering that apply to spontaneous speech. *Contemporary Issues in Communication Science and Disorders.*, *31*, 123–140.
Howell, P., & Au-Yeung, J. (1995). Syntactic determinants of stuttering in the spontaneous speech of normally fluent and stuttering children. *Journal of Fluency Disorders*, *20*, 317–330.
Howell, P., Soukup-Ascencao, T., Davis, S., & Rusbridge, S. (2011). Comparison of alternative methods for obtaining severity scores of the speech of people who stutter. *Clinical Linguistics and Phonetics*, *25*(5), 368–378.
Huinck, W., & Rietveld, T. (2007). The Validity of a Simple Outcome Measure to Assess Stuttering Therapy. *Folia Phoniatrica et Logopaedica*, *59*, 91–99.
Hutchinson, A., Bentzen, N., & König-Zahn, C. (1997). *Cross Cultural Health Outcome Assessment: a user's guide.* European Research Group on Health Outcomes.
Hymes, D. (1967). Models of the interaction of language and social setting. *Journal of Social Issues*, *23*(2), 8–28.
Ingham, R., & Cordes, A. K. (1992). Interclinic differences in stuttering event counts. *Journal of Fluency Disorders*, *17*, 171–176.
Ingham, R. J., Gow, M., & Costello, J. M. (1985). Stuttering and speech naturalness: some additional data. *Journal of Speech and Hearing Disorders*, *50*, 217–219.
Iverach, L., O'Brian, S., Jones, M., Block, S., Lincoln, M., Harrison, E., … Menzies, R. G. (2009). Prevalence of anxiety disorders among adults seeking speech therapy for stuttering. *Journal of Anxiety Disorders*, *23*, 928–934.
Iverach, L., & Rapee, R. M. (2014). Social anxiety disorder and stuttering: Current status and future directions. *Journal of Fluency Disorders*, *40*, 69–82.
Jesus, L. M. T., Valente, A. R. S., & Hall, A. (2015). Is the Portuguese version of the passage "The North Wind and the Sun" phonetically balanced? *Journal of the International Phonetic Association*, *45*, 1–11.
Johnson, W. (1961). Measurement of oral reading and speaking rate and disfluency of adult male and female stutterers and nonstutterers. *Journal of Speech and Hearing Disorders, Monograph Supplement*, *7*, 1–20.
Johnson, W., Boehmler, R. M., Dahlstrom, W. G., Darley, F. L., Goodstein, L. D., Kools, J. A., … Young, M. A. (1959). *The onset of stuttering*. Minneapolis, MN: University of Minneapolis Press.
Johnson, W., Darley, F. L., & Spriestersbach, D. C. (1963). *Diagnostic methods in speech pathology*. New York: Harper.
Johnson, W., Darley, F., & Spriestersbach, D. (1952). *Form 14: Stutterers' Self-Ratings of Reactions to Speech Situations. Diagnostic Manual in Speech Corrections*. New York: Harper and Brothers.
Jones, M., Onlow, M., Harrison, E., & Packman, A. (2000). Treating Stuttering in Young Children: Predicting Treatment Time in the Lidcombe Program. *Journal of Speech, Language, and Hearing Research*, *43*, 1440–1450.
Jones, M., Onslow, M., Packman, A., & Gebski, V. (2006). Guidelines for Statistical Analysis

of Percentage of Syllables Stuttered Data. *Journal of Speech, Language, and Hearing Research*, *49*, 867–878.
Karimi, H., O'Brian, S., Onslow, M., & Jonesc, M. (2014). Absolute and Relative Reliability of Percentage of Syllables Stuttered and Severity Rating Scales. *Journal of Speech, Language, and Hearing Research*, *57*, 1284–1295.
Kelly, E. M., & Conture, E. G. (1988). Acoustic and perceptual correlates of adults stutterers' typical and imitated stutterings. *Journal of Fluency Disorders*, *13*, 233–252.
Kirsch, D. (2006). *Adolescent perceptions of their peers who stutter*. University of Oregon.
Klassen, T. R. (2002). Social distance in the negative stereotype of people who stutter. *Journal of Speech-Language Pathology and Audiology*, *26*, 90–99.
Kline, R. B. (2004). *Beyond significance testing: Reforming data analysis methods in behavioral research*. Washington, DC: American Psychological Association.
Kully, D., & Boberg, E. (1988). An investigatio of inter-clinic agreement in the identification of fluent and astuttered syllables. *Journal of Fluency Disorders*, *13*, 309–318.
Kundel, H. L., & Polansky, M. (2003). Measurement of observer agreement. *Radiology*, *228*(2), 303–308.
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.
Langevin, M., & Boberg, E. (1993). Results of an Intensive Stuttering Therapy Program. *Canadian Journal of Speech-Language Pathology and Audiology*, *17*(3 and 4), 158–166.
Langevin, M., Huinck, W. J., Kully, D., Peters, H. F. M., Lomheim, H., & Tellers, M. (2006). A cross-cultural, long-term outcome evaluation of the ISTAR Comprehensive Stuttering Program across Dutch and Canadian adults who stutter. *Journal of Fluency Disorders*, *31*, 229–256.
Langevin, M., Kully, D., Teshima, S., Hagler, P., & Narasimha, P. N. G. (2010). Five-year longitudinal treatment outcomes of the ISTAR Comprehensive Stuttering Program. *Journal of Fluency Disorders*, *35*, 123–140.
Lass, N. J., Ruscello, D. M., Pannbacker, M., Schmitt, J. F., & Everly-Myers, D. S. (1989). Speech-language pathologists' perceptions of child and adult female and male stutterers. *Journal of Fluency Disorders*, *14*, 127–134.
Lass, N. J., Ruscello, D. M., Pannbacker, M., Schmitt, J. F., Kiser, A. M., Mussa, A. M., & Lockhart, P. (1994). School administrators' perceptions of people who stutter. *Language, Speech, and Hearing Services in Schools*, *25*, 90–93.
Lass, N. J., Ruscello, D. M., Schmitt, J. F., Pannbacker, M. D., Orlando, M. B., Dean, K. A., … Bradshaw, K. H. (1992). Teachers' perceptions of stutterers. *Language, Speech, and Hearing Services in Schools*, *23*, 78–81.
Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The Role and Interpretation of Pilot Studies in Clinical Research. *Journal of Psychiatric Research*, *45*(5), 626–629.
Lewis, D., & Sherman, D. (1951). Measuring the severity of stuttering. *Journal of Speech Disorders*, *16*, 320–326.
Lewis, K. E. (1995). Do SSI-3 scores adequately reflect observations of stuttering behaviors? *American Journal of Speech-Language Pathology*, *4*, 46–59.
Li, J., & Arnold, H. (2015). Reactions of teachers versus non-teachers toward people who stutter. *Journal of Communication Disorders*, *56*, 8–18.
Lincoln, M., Packman, A., & Onslow, M. (2010). An Experimental Investigation of the Effect of Altered Auditory Feedback on the Conversational Speech of Adults Who Stutter. *Journal of Speech, Language, and Hearing Research*, *53*, 1122–1131.
Logan, K., & Conture, E. G. (1995). Length, grammatical complexity, and rate differences in stuttered and fluent conversational utterances of children who stutter. *Journal of Fluency Disorders*, *20*, 35–61.
Logan, K. J., & Haj-Tas, M. A. (2007). Effect of Sample Size on the Measurement of Stutter-Like Disfluencies. *Perspectives on Fluency and Fluency Disorders*, *17*, 3–6.

Louis, K. O., Reichel, I. K., Yaruss, J. S., & Lukber, B. B. (2009). Construct and concurrent validity of a prototype questionnaire to survey public attitudes toward stuttering. *Journal of Fluency Disorders*, *34*, 11–28.
Lund, N., & Duchan, J. F. (1993). *Assessing children's language in naturalistic contexts*. Englewood Cliff, NJ: Prentice Hall.
Lynn, M. (1986). Determination and Quantification of Content Validity. *Nursing Research*, *35*(6), 382–385.
Ma, D. (2015). The skewness of a probability distribution. Retrieved from https://probabilityandstats.wordpress.com/2015/05/09/the-skewness-of-a-probability-distribution/
MacDonald, J., & Haroldson, S. K. (1973). Stuttering and disfluency as two reliable and unambiguous response classes. *Journal of Speech and Hearing Research*, *16*, 691–699.
MacKinnon, S. P., Hall, S., & MacIntyre, P. D. (2007). Origins of the stuttering stereotype: Stereotype formation through anchoring-adjustment. *Journal of Fluency Disorders*, *32*, 297–309.
Manning, W. (2010). *Clinical Decision Making in Fluency Disorders* (3rd ed.). Clifton Park: Delmar.
Manning, W., & Beck, J. G. (2013). Personality dysfunction in adults who stutter: Another look. *Journal of Fluency Disorders*. Retrieved from http://dx.doi.org/10.1016/j.jfludis.2013.02.001
Martin, R. R., & Haroldson, S. K. (1981). Stuttering identification: Standard definition and moment of stuttering. *Journal of Speech and Hearing Research*, *24*, 59–63.
Martin, R. R., Haroldson, S. K., & Triden, K. A. (1984). Naturalness and Speech Stuttering. *Journal of Speech and Hearing Disorders*, *49*, 53–58.
Martin, R. R., Haroldson, S. K., & Woessner, G. L. (1988). Perceptual scaling of stuttering severity. *Journal of Fluency Disorders*, *13*, 27–47.
Martins, F., Vigário, M., & Frota, S. (2009). FreP - Frequências no Português. Version 2.0. Software. (IGAC 5969/2009).
McClean, M. D. (1990). Neuromotor aspects of stuttering: Levels of impairment and disability. In J. Cooper (Ed.), *Research needs in stuttering: Roadblocks and future directions* (pp. 64–71). Rockville, MD: American Speech-Language-Hearing Association.
McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*, *22*(3), 276–282.
Menzies, R. G., O'Brian, S., Onslow, M., Packman, A., St Clare, T., & Block, S. (2008). An Experimental Clinical Trial of a Cognitive-Behavior Therapy Package for Chronic Stuttering. *Journal of Speech, Language, and Hearing Research*, *51*, 1451–1464.
Meyers, S. C. (1986). Qualitative and quantitative differences and patterns of variability in disfluencies emitted by preschool stutterers and nonstutterers during dyadic conversation. *Journal of Fluency Disorders1*, *11*, 293–306.
Meyers, S. C., & Freeman, F. J. (1985a). Interruptions as a variable in stuttering and disfluency. *Journal of Speech and Hearing Research*, *28*, 428–425.
Meyers, S. C., & Freeman, F. J. (1985b). Mother and child speech rate as a variable in stuttering and disfluency. *Journal of Speech and Hearing Research*, *28*, 436–444.
Moran, P., & Crawford, M. J. (2013). Assessing the severity of borderline personality disorder. *The British Journal of Psychiatry*, *203*, 163–164.
Mowrer, D. (1991). Assessment review: Stuttering Severity Instrument for Children and Adults (revised). *Journal of Fluency Disorders*, *13*, 311–314.
Naylor, R. V. (1953). A comparative study of methods os estimating the severity of stuttering. *Journal of Speech and Hearing Disorders*, *18*, 30–37.
Nillson, S., & Wetterling, T. (2013). *"Try to relax" - A survey study in Sweden among native Swedes and Arabic immigrants concerning views on stuttering. Unpublished master's*

Thesis. Lund, Sweden: Lund University.
Ninio, A., & Snow, C. (1999). The development of pragmatics: Learning to use language appropriately. In K. T. Bhatia & W. C. Ritchie (Eds.), *Handbook of language acquisition* (pp. 347–383). New York: Academic Press.
Nippold, M., Schwarz, I., & Jescheniak, J. (1991). Narrative ability in school-age stuttering boys: A preliminary investigation. *Journal of Fluency Disorders*, *16*, 289–308.
Ntourou, K., Conture, E. G., & Lipsey, M. W. (2011). Language abilities of children who stutter: A meta-analytical review. *American Journal of Speech Language Pathology*, *20*(3), 163–179.
Nunnally, J. C., & Bernestein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill Inc.
O'Brian, S., Carey, B., Onslow, M., Packman, A., & Cream, A. (2010). The Camperdown program for stuttering: Tratment manual. Retrieved from http://sydney.edu.au/health_sciences/ asrc/docs/camperdown_manual.pdf
O'Brian, S., Onslow, M., Cream, A., & Packman, A. (2003). The Camperdown Program: outcomes of a new prolonged-speech treatment model. *Journal of Speech, Language, and Hearing Research*, *46*, 933–946.
O'Brian, S., Packman, A., & Onslow, M. (2004). Self-Rating of Stuttering Severity as a Clinical Tool. *American Journal of Speech-Language Pathology*, *13*, 219–226.
O'Brian, S., Packman, A., & Onslow, M. (2008). Telehealth Delivery of the Camperdown Program for Adults Who Stutter: A Phase I Trial. *Journal of Speech, Language, and Hearing Research*, *51*, 184–195.
O'Brian, S., Packman, A., Onslow, M., & O'Brian, N. (2004). Measurement of stuttering in adults: Comparison of stuttering rate and severity rating methods. *Journal of Speech, Language and Hearing Research*, *47*, 1081–1087.
O'Donnell, J. J., Armson, J., & Kiefte, M. (2008). The effectiveness of SpeechEasy during situations of daily living. *Journal of Fluency Disorders*, *33*, 99–119.
Onslow, M. (2016). *Stuttering and its treatment: eleven lectures*. Sidney.
Onslow, M. (2017). *Stuttering and its treatment: eleven lectures*. Sidney.
Onslow, M., Adams, R., & Ingham, R. (1992). Reliability of Speech Naturalness Ratings of Stuttered Speech During Treatment. *Journal of Speech and Hearing Research*, *35*, 994–1001.
Onslow, M., Costa, L., & Rue, S. (1990). Direct early intervention with stuttering: Some preliminary data. *Journal of Speech and Hearing Disorders*, *55*, 405–416.
Onslow, M., Harrison, E., Jones, M., & Packman, S. (2002). Beyond-clinic speech measures during the Lidcombe Program of early stuttering intervention. *Acquiring Knowledge in Speech, Language, and Hearing*, *4*, 82–85.
Ornstein, A., & Manning, W. (1985). Self-efficacy scaling by adult stutterers. *Journal of Communication Disorders*, *18*, 313–320.
Owens, R. E. (2011). *Language Development: An Introduction* (8th ed.). Boston: Pearson.
Özdemir, R. S., St. Louis, K. O., & Topbaş, S. (2011a). Public attitudes toward stuttering in Turkey: Probability versus convenience sampling. *Journal of Fluency Disorders*, *36*, 262–267.
Özdemir, R. S., St. Louis, K. O., & Topbaş, S. (2011b). Stuttering attitudes among Turkish family generations and neighbors from representative samples. *Journal of Fluency Disorders*, *36*, 318–333.
Packman, A., & Onslow, M. (1998). The behavioral data language of stuttering. In A. K. C. & R. J. Ingham (Ed.), *Treatment efficacy for stuttering: A search for empirical bases* (pp. 27–50). San Diego: CA: Singular.
Packman, A., Onslow, M., & Bryant, K. (2000). The lexicon of stuttering. In K. L. Baker, L. Rustin, & F. Cook (Eds.), *Proceedings of the Fifth Oxford Dysfluency Conference* (pp. 53–60). Leicester, U.K.: K. L. Baker.
Packman, A., Onslow, M., O'Brian, S., & Huber, A. (2004). Measuring the Outcomes of

Behavioral Stuttering Treatments Across Situations: What Are Our Data Telling Us? In *Evidence-Based Treatment of Stuttering: Empirical Bases and Clinical Applications* (pp. 245–258). Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Packman, A., Onslow, M., & van Doom, J. (1994). Prolonged Speech and Modification of Stuttering: Perceptual, Acoustic, and Electroglottographic Data. *Journal of Speech and Hearing Research*, *37*, 724–737.

Paul, R., & Norbury, C. (2012). *Language disorders from infancy through adolescence: listening, speaking, reading, writing, and communicating* (4th ed.). St. Louis, MO: Elsevier Mosby.

Perry, A., Morris, M., Unsworth, C., Duckett, S., Skeat, J., Dodd, K., … Reilly, K. (2004). Therapy outcome measure for allied health practitioners in Australia: the AusTOMs. *International Journal for Quality in Health Care*, *16*(4), 285–291.

Plexico, L. W., Manning, W. H., & Levitt, H. (2009). Coping responses by adults who stutter: Part I: Protecting the self and others. *Journal of Fluency Disorders*, *34*, 87–107.

Polit, D. F., & Beck, C. T. (2012). *Nursing research : generating and assessing evidence for nursing practice* (9th ed.). Philadelphia: Wolters Kluwer Health/Lippincott Williams & Wilkins.

Postma, A., & Kolk, H. (1993). The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. *Journal of Speech and Hearing Research*, *36*, 472–487.

Prins, D. (1991). Theories of stuttering as event and disorder: Implications for speech production processes. In H. F. M. Peters, W. Hulstijn, & C. W. Starkweather (Eds.), *Speech motor control and stuttering* (pp. 571–580). Amsterdam: Elsevier Sciences Publishers.

Przepiorka, A., Blachnio, A., St. Louis, K. O., & Woźniak, T. (2013). Public attitudes toward stuttering in Poland. *International Journal of Language and Communication Disorders*, *48*, 703–714.

Riley, G. D. (1972). *Stuttering severity instrument*. Austin, Texas: PRO-ED.

Riley, G. D. (1980). *Stuttering Severity Instrument (rev. ed.)*. Austin, Texas: Pro-Ed.

Riley, G. D. (1994). *The Stuttering Severity Instrument for Adults and Children (SSI-3) (3rd ed.)*. Austin, Texas: PRO-ED.

Riley, G. D. (2009). *The Stuttering Severity Instrument for Adults and Children (SSI-4) (4th ed.)* (4th ed.). Austin, Texas: PRO-ED.

Riley, J., Riley, G. D., & Maguire, G. (2004). Subjective screening of stuttering severity, locus of control and avoidance: Research edition. *Journal of Fluency Disorders*, *29*, 51–62.

Roberts, P. M., Meltzer, A., & Wilding, J. (2009). Disfluencies in non-stuttering adults across sample lengths and topics. *Journal of Communication Disorders*, *42*(6), 414–427.

Rousseau, S., Onslow, M., Packman, A., & Jones, M. (2008). Comparisons of Audio and Audiovisual Measures of Stuttering Frequency and Severity in Preschool-Age Children. *American Journal of Speech-Language Pathology*, *17*, 173–178.

Ruscello, D. M., Lass, N. J., Schmitt, J. F., & Pannbacker, M. D. (1994). Special educators' perceptions of stutterers. *Journal of Fluency Disorders*, *19*, 125–132.

Rusciano, F. L. (2003). The construction of national identity—A 23-nation study. *Political Research Quarterly*, *56*, 361–366.

Russell, R. L., & Grizzle, K. L. (2008). Assessing Child and Adolescent Pragmatic Language Competencies: Toward Evidence-Based Assessments. *Clinical Child and Family Psychology Review*, *11*, 59–73.

Ryan, B. (2001). *Programmed therapy for stuttering children and adults* (2nd ed.). Springfield, IL: Charles C. Thomas.

Salkind, N. J. (2007). *Encyclopedia of Measurement and Statistics*. SAGE Publications, Inc.

Santos, B. S. S. (1995). Sociedade-providência ou autoritrismo social? . *Revista Crítica de Ciências Sociais*, *42*, i–vii.

Sawyer, J., & Yairi, E. (2006). The Effect of Sample Size on the Assessment of Stuttering Severity. *American Journal of Speech-Language Pathology*, *15*, 36–44.
Schilling, L. S., Dixon, J. K., Knafl, K. A., Grey, M., Ives, B., & Lynn, M. R. (2007). Determining content validity of a self-report instrument for adolescents using a heterogeneous expert panel. *Nursing Research*, *56*(5), 361–366.
Schloss, P. L., Freeman, C. A., & Smith, M. A. (1987). Influence of assertiveness training on the stuttering rates exhibited by three young adults. *Journal of Fluency Disorders*, *12*, 333–353.
Shapiro, D. (2011). *Stuttering Intervention: A Collaborative Journey to Fluency Freedom* (2nd ed.). Pro-Ed Inc.
Shapiro, D. A. (1999). *Stuttering intervention: a colaborative journey to fluency freedom*. Austin, Texas: Pro-ed.
Shipley, K. G., & McAfee, J. G. (2016). Assessment in speech-language pathology: a resource manual (5th ed.). Boston: Delmar Cengage Learning.
Silva, R. (2011). *Valores e Felicidade no Século XXI: Um Retrato Sociológico dos Portugueses em comparação europeia [Values and Hapiness on XXI century: A Sociological Portrait of the Portuguese in an European comparison]*. *Departamento de Sociologia*. Instituto Universitário de Lisboa, Lisboa.
Silverman, S., & Ratner, N. (1997). Syntactic complexity, fluency, and accuracy of sentence imitation in adolescence. *Journal of Speech, Language, and Hearing Research*, *40*, 95–106.
Sommer, M., Koch, M. A., Paulus, W., Weiller, C., & Büchel, C. (2002). Disconnection of speech-relevant brain areas in persistent developmental stuttering. *The Lancet*, *360*, 380–383.
Spencer, E., Packman, A., Onslow, M., & Ferguson, A. (2005). A preliminary investigation of the impact of stuttering on language use. *Clinical Linguistics and Phonetics*, *19*(3), 191–201.
Spencer, E., Packman, A., Onslow, M., & Ferguson, A. (2009). The effect of stuttering on communication: A preliminary investigation. *Clinical Linguistics and Phonetics*, *23*(7), 473–488.
St. Louis, K. O. (2005). A global project to measure public attitudes of stuttering. *The Asha Leader*, *10*(12–13), 22.
St. Louis, K. O. (2011). International project on attitudes toward human attributes (IPATHA). Retrieved from http://www.stutteringattitudes.com/
St. Louis, K. O. (2012a). Male versus female attitudes toward stuttering. *Journal of Communication Didorders*, *45*, 246–253.
St. Louis, K. O. (2012b). POSHA-S public attitudes toward stuttering: Online versus paper surveys . *Canadian Journal of Speech-Language Pathology and Audiology*, *36*(2), 116–123.
St. Louis, K. O. (2012c). Research and development on a public attitud instrument for stuttering. *Journal of Communication Didorders*, *46*, 129–146.
St. Louis, K. O. (2015a). Epidemiology of public attitudes toward stuttering. In K. O. St. Louis (Ed.), *Stuttering meets stereotype, stigma, and discrimination: An overview of attitude research* (pp. 7–42). Morgantown, WV: West Virginia University Press.
St. Louis, K. O. (2015b). Worldwide perspectives on public and professional attitudes toward fluency disorders. Invited keynote address at the 8th World Congress of the International Fluency Association. Lisbon, Portugal.
St. Louis, K. O., Andrade, C., Georgieva, D., & Troudt, F. (2005). Experiência e relato pessoal sobre pesquisa de cooperação internacional - Brasil, Bulgária e Turquia - que avalia as atitudes em relação à gagueira [Experience and personal report about an international cooperation research - Brazil, Bulgaria and Turckey . *Pró-Fono Revisa de Atualização Científica*, *17*, n. *3*, 413–416.
St. Louis, K. O., Coskun, M., Ozdemir, S., Topbas, S., Goranova, E., & Filatova, Y. (2005).

Experience and personal report about an international cooperation research—Brazil, Bulgaria and Turkey—Attitudes Toward Stuttering. *Pró-Fono Revista de Atualização Cientifica*, *17*, 413–416.
St. Louis, K. O., Filatova, Y., Çoşkun, M., Topbaş, S., Özdemir, S., Georgieva, D., … George, R. D. (2011). Public attitudes toward cluttering and stuttering in four countries. In E. L. Simon (Ed.), *Psychology of stereotypes* (pp. 81–113). Hauppauge, NY: Nova Science Publishers.
St. Louis, K. O., Kuhn, C. D., & Lytwak, L. (2015). The Appraisal of the Stuttering Environment (ASE): A new clinical tool to measure stuttering attitudes the client's environment. In K. O. St. Louis (Ed.), *Stuttering meets stereotype, stigma, and discrimination: An overview of attitude research* (pp. 255–273). Morgantown, WV: West Virginia University Press.
St. Louis, K. O., LeMasters, S. N., & Poormohammad, A. (2015). Public attitudes toward males versus females who stutter: United States and Iran. In K. O. St. Louis (Ed.), *Stuttering meets stereotype, stigma, and discrimination: An overview of attitude research* (pp. 165–177). Morgantown, WV: West Virginia University Press.
St. Louis, K. O., Lubker, B. B., Yaruss, J. S., & Aliveto, E. F. (2009). Development of a prototype questionnaire to survey public attitudes toward stuttering: Reliability of the second prototype. *Contemporary Issues in Communication Science and Disorders*, *36*, 101–107.
St. Louis, K. O., Lukber, B. B., Yaruss, J. S., Adkins, T. A., & Pill, J. C. (2008). Development of a prototype questionnaire to survey public attitudes toward stuttering: Principles and methodologies in the first prototype. *The Internet Journal of Epidemiology*, *5(2)*.
St. Louis, K. O., Przepiorka, A. M., Beste-Guldborgc, A., Williams, M. J., Blachnio, A., Guendouzi, J., … Ware, M. B. (2014). Stuttering attitudes of students: Professional, intracultural, and international comparisons. *Journal of Fluency Disorders*, *39*, 34–50.
St. Louis, K. O., & Roberts, P. M. (2010). Measuring attitudes toward stuttering: English-to-French translations in Canada and Cameroon. *Journal of Communication Disorders*, *43*, 361–377.
St. Louis, K. O., Sønsterud, H., Carlo, E. J., Heitmann, R. R., & Kvenseth, H. (2014). Public attitudes toward—and identification of—cluttering and stuttering in Norway and Puerto Rico. *Journal of Fluency Disorders*, *42*, 21–34.
St.Louis, K. O., Sønsterud, H., Junuzović-Žunić, L., Tomaiuoli, D., DelGado, F., Caparelli, E., … We ̨sierska, M. (2016). Public Attitudes Toward Stuttering in Europe: Within-Country and Between-Country Comparisons. *Journal of Communication Disorders*. http://doi.org/http://dx.doi.org/doi:10.1016/j.jcomdis.2016.05.010
Starkweather, C. W. (1987). *Fluency and stuttering*. Englewood Cliffs, NJ: Prentice–Hall.
Starkweather, C. W., Gottwald, S. R., & Halfond, M. M. (1990). *Stuttering prevention: A clinical method*. Englewood Cliffs, NJ: Prentice-Hall.
Stein, M. B., & Ken, Y. M. (2000). Disability and quality of life in social phobia: Epidemiological findings. *American Journal of Psychiatry*, *157*, 1606–1613.
Stetson, R. H. (1951). *Motor phonetics* (2nd ed.). Amsterdam: North-Holland.
Stuart, A., & Kalinowski, J. (2004). The Perception of Speech Naturalness of Post-Therapeutic and Altered Auditory Feedback Speech of Adults with Mild and Severe Stuttering. *Folia Phoniatrica et Logopaedica*, *56*(6), 347–357.
Tacchetti, M. (2013). *User Guide for ELAN Linguistic Annotator version 4.1.0*. Nijmegen, The Netherland: The Language Archive, MPI for Psycholinguistics.
Tanner, D. C., Belliveau, W., & Seibert, G. (1999). *Pragmatic Stuttering Intervention for Adolescents and Adults*. Oceanside, CA: Academic Communication Associates.
Teesson, K., Packman, A., & Onslow, M. (2003). The Lidcombe Behavioral Data Language of stuttering. *Journal of Speech Language and Hearing Research*, *46*(4), 1009–1015.
Theiling, M. (2013). *Öffentliche Wahrnehmung von Stotternden in Deutschland –Erprobung und Validierung der deutschen Version des POSHA-S [Public perception of people*

who stutter in Germany – Testing and validation of the German version of the POSHA–S.] Unpublished Master's th. RWTH Aachen University, Aachen, Germany.
Thompson, S. K. (2012). *Sampling* (3rd ed.). New York: Wiley.
Tomaiuloi, D., Del Gado, F., Capparelli, E., & St. Louis, K. O. (2013). Stuttering attitudes in Italy. *29th World Congress of the International Association of Logopedics and Phoniatrics*. Turin, Italy.
Torres, A., & Brites, R. (2006). European attitudes and values: The perspective of gender in a transverse analysis. *Portuguese Journal of Social Science Volume*, *5*(3), 179–214.
Tran, Y., Blumgart, E., & Craig, A. (2011). Subjective distress associated with chronic stuttering. *Journal of Fluency Disorders*, *36*, 17–26.
Tumanova, V., Conture, E. G., Lambert, E. W., & Walden, T. A. (2014). Speech disfluencies of preschool-age children who do and do not stutter. *Journal of Communication Disorders*, *49*, 24–41.
Tyrrell, L. (2009). *A comparison of the attitudes, knowledge and beliefs about stuttering in British university students living in the UK and Syrian university students living in Syria. MSc Dissertation*. University of Reading.
Unger, J. P., Glückb, C. W., & Cholewaa, J. (2012). Immediate effects of AAF devices on the characteristics of stuttering: A clinical analysis. *Journal of Fluency Disorders*, *37*, 122–134.
Vala, J., Cabral, M. V, & Ramos, A. (2003). *Valores sociais: mudanças e contrastes em Portugal e na Europa*. Lisboa: Imprensa de Ciências Sociais.
Valente, A. R. S., Jesus, L. M. T., Hall, A., & Leahy, M. (2014). Event- and interval-based measurement of stuttering: a review. *International Journal of Language and Communication Disorders*, *50*(1), 14–30.
Van Borsel, J., & Eeckhout, H. (2008). The speech naturalness of people who stutter speaking under delayed auditory feedback as perceived by different groups of listeners. *Journal of Fluency Disorders*, *33*, 241–251.
Van Borsel, J., Verniers, I., & Bouvry, S. (1999). Public awareness of stuttering. *Folia Phoniatrica Logopaedica*, *51*, 124–132.
Van Riper, C. (1982). *The Nature of Stuttering* (2nd ed.). Englewood Cliffs NJ: Prentice Hall.
Vaz, S., Falkmer, T., Passmore, A. E., Parsons, R., & Andreou, P. (2013). The Case for Using the Repeatability Coefficient When Calculating Test–Retest Reliability. *PLoS ONE*, *8*(9), e73990.
Venkatagiri, H. S. (2005). Recent advances in the treatment of stuttering: A theoretical perspective. *Journal of Communication Disorders*, *38*(5), 375–393.
Ward, D. (2008). *Stuttering and Cluttering: Frameworks for understanding and treatment*. USA: Taylor & Francis e-Library.
Weiss, A. (2004). Why We Should Consider Pragmatics When Planning Treatment for Children Who Stutter. *Language, Speech, and Hearing Services in Schools*, *35*, 34–45.
Weiss, A., & Zebrowski, P. (1994). The narrative productions of children who stutter: A preliminary view. *Journal of Fluency Disorders*, *19*, 39–63.
Węsierska, K., Węsierska, M., & St. Louis, K. O. (2013). A comparison of attitudes toward stuttering between Polish SLTPs and SLP students using POSHA-S. *Poster Presented at the Stuttering Attitudes Research Symposium*. Morgantown, WV.
Węsierska, M., & St. Louis, K. O. (2014). Comparison of sttitudes towards stuttering among Polish and English university students. *Chowanna*, *42*, 263–284.
White, P. A., & Collins, S. R. C. (1984). Stereotype formation by inference: a possible explanation for the "stutterer" stereotype. *Journal of Speech and Hearing Research*, *27*, 567–570.
WHO. (1980). *International classification of impairments, disabilities, and handicaps: A manual of classification relating to the consequences of disease*. Geneva: World Health Organization.

WHO. (2001). *The International classification of functioning, disability, and health*. Geneva: World Health Organization.
Wiig, E. H., Semel, E., & Secord, W. A. (2013). *Clinical Evaluation of Language Fundamentals - Fifth Edition (CELF-5)*. Pearson.
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*, *8*(2), 94–104.
Wingate, M. (1976). *Stuttering: Therapy and treatment.* New York, NY: Irvington.
Wingate, M. (1988). *The structure of stuttering: A psycholinguistic analysis*. New York, NY: Springer- Verlag.
Woolf, G. (1967). The assessment of stuttering as struggle, avoidance, and expectancy. *British Journal of Disorders of Communication*, *2*, 158–171.
Wright, L., & Ayre, A. (2000). *WASSP : the Wright and Ayre stuttering self-rating profile*. Winslow: Bicester.
Yairi, E., & Ambrose, N. (1992). A longitudinal study of stuttering in children: A preliminary report. *Journal of Speech and Hearing Research*, *35*, 755–760.
Yairi, E., & Seery, C. H. (2014). *Stuttering: Foundations and Clinical Applications*. (2, Ed.). Boston: Pearson Communication Sciences & Disorders.
Yairi, E., & Williams, D. E. (1970). Speech clinicians' stereotypes of elementary-school boys who stutter. *Journal of Communication Disorders*, *3*, 161–170.
Yaruss, J. S. (1997). Clinical Measurement of Stuttering Behaviors. *Contemporary Issues in Communication Science and Disorders*, *24*, 33–44.
Yaruss, J. S. (1998a). Describing the Consequences of Disorders: Stuttering and the International Classification of Impairments, Disabilities, and Handicaps. *Journal of Speech, Language, and Hearing Research*, *41*, 249–257.
Yaruss, J. S. (1998b). Real-time analysis of speech fluency: Procedures and reliability training. *American Journal of Speech-Language Pathology*, *7*, 25–37.
Yaruss, J. S. (2007). Application of the ICF in Fluency Disorders. *Seminars in Speech and Language*, *28*, 312–322.
Yaruss, J. S., & Quesal, R. W. (2004a). Partnerships between clinicians, researchers, and people who stutter in the evaluation of stuttering treatment outcomes. In *Stammering Research* (pp. 1–15).
Yaruss, J. S., & Quesal, R. W. (2004b). Stuttering and the international classification of functioning, disability, and health (ICF). *Journal of Communication Disorders*, *37(1)*, 35–52.
Yaruss, J. S., & Quesal, R. W. (2006). Overall Assessment of the Speaker's Experience of Stuttering (OASES): Documenting multiple outcomes in stuttering treatment. *Journal of Fluency Disorders*, *31*, 90–115.
Yeakle, M. K., & Cooper, E. B. (1986). Teacher perceptions of stuttering. *Journal of Fluency Disorders*, *11*(4), 345–359.
Young, M. A. (1975). Observer agreement for marking moments of stuttering. *Journal of Speech and Hearing Research*, *29*, 484–489.
Zebrowski, P. M. (1991). Duration of the speech disfluencies of beginning stutterers. *Journal of Speech and Hearing Research*, *34*, 483–491.
Zebrowski, P. M. (1994). Duration of Sound Prolongation and Sound/Syllable Repetition in Children Who Stutter: Preliminary Observations. *Journal of Speech and Hearing Research*, *37*, 254–263.
Zebrowski, P. M., & Conture, E. G. (1989). Judgments of disfluency by mothers of stuttering and normally fluent children. *Journal of Speech and Hearing Research*, *32*, 625–634.