



SLP Platform

IMPACT

Stuttering
Rating Scale

Copyright © 2020 by Lavi Institute

All rights reserved. No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission in writing from the copyright owner.

TABLE OF CONTENTS

About the Author.....	3
Acknowledgements	3
Chapter:	
1. Overview of the IMPACT Stuttering Rating Scale	4
IMPACT Description	
IMPACT Test Areas	
Testing Format	
Testing Time	
IMPACT Uses	
Features	
Rationale	
2. Administration and Scoring Procedures.....	7
Examiner Qualifications	
Testing Time	
Accessing Rating Scales	
Administration Instructions	
3. Standardization and Normative Information.....	10
4. Reliability and Validity.....	15
5. Highlights.....	23
References.....	24

About the Author

Adriana Lavi, PhD, CCC-SLP is a licensed speech-language pathologist and a pioneer in the development of speech and language video-based assessment tools. She is the creator and author of the Clinical Assessment of Pragmatics (CAPs), as well as Video Assessment Tools, an online assessment platform that features the Stuttering Video Assessment Tool, the IMPACT Social Communication Rating Scale, the IMPACT Stuttering Rating Scale, etc. Additionally, Dr. Lavi is the creator of the Video Learning Squad, an online therapy platform that features the Social Squad and Stutter Squad.

For over a decade, Dr. Lavi owned Go2Consult where she supervised 35+ speech-language pathologists and clinical fellows across Southern California. Dr. Lavi has also served as an Assistant Professor at the Department of Communicative Disorders at Loma Linda University, and is the founder of the Lavi Institute for Research and Professional Development. She earned a master's degree in speech-language pathology from California State University at Sacramento and a PhD degree in Rehabilitation Sciences with an emphasis in speech-language pathology from Loma Linda University. Dr. Lavi was one of three students selected by the Bureau of Educational and Cultural Affairs of the US Department of State from the country of Moldova to study in the US in 2000. She has lived through and understands the culture of poverty. Her professional career has always focused on service delivery for students from low-income backgrounds. Dr. Lavi is the proud mother of four young, highly energetic boys.

Consultant and Editor on Neurodiversity Topics

Tiffany Waddington, M.S. CCC-SLP, is a neurodivergent mother of 2 autistic kids. She has a passion for working with high school and transition aged students, with a focus on helping these students become as independent as possible.

During the school year, she works with high school and transition students in the Everett School District. She has worked previously with Aspiring Youth, running D&D games and social groups in multiple cities, as well as providing 1:1 mentoring with a variety of young adults.

Most recently, she has been working on a video series for neurodivergent friendships and is looking forward to running more sustainable groups to help teens and young adults thrive.

Acknowledgements

Sincere appreciation is extended to the families and children who generously offered their time and effort. Additional thanks goes to the expert panel and speech-language pathologists who field-tested the children.

Overview of the Rating Scale

IMPACT Stuttering Rating Scale Description

The *IMPACT Stuttering Rating Scale* is a norm-referenced stuttering rating scale for children and young adults ages 5 through 21 years old. It is composed of 30-35 test items, and has three separate forms to be completed by clinician, parent(s), and teacher(s). It is an accurate and reliable assessment tool that provides valid results on informal observations of fluency characteristics, social interactions, academic life, and home/after school life. Normative data of this test is based on a nationally representative sample of 1396 children and young adults in the United States.

The IMPACT Model

The IMPACT model was developed based on current literature and examination of real-world challenges faced by individuals who stutter such as school demands and social interactions. This model was designed to analyze the real-life authentic observations of teachers, parents, and clinicians. The IMPACT model uses a contextualized, whole student approach to see the impact and the outcome of stuttering on education and social interactions.

IMPACT Stuttering Rating Scale Areas

The test is composed of four areas: speech characteristics, social interactions, academic and home/after school life.

Testing Format

The *IMPACT Stuttering Rating Scale* is composed of 30-35 test items. The test uses a series of items that asks the rater to score on a 4-point scale (“never,” “sometimes,” “often,” and “typically”). The rating scale yields an overall percentile and standard score.

Administration Time

Administration time for the rating scale takes approximately 20-25 minutes.

IMPACT Stuttering Rating Scale Uses and Purpose

Clinicians, parents, and teachers can provide valuable information regarding a student's speech and fluency characteristics and how stuttering may impact the child in both the classroom and in the home environment. *The IMPACT Stuttering Rating Scale* should be used to evaluate children or young adults who have a suspected or previous diagnosis of stuttering. This tool will assist in the identification or continued diagnosis of stuttering. Additionally, this rating scale will help determine if there are any educational or personal impacts. The results of the *IMPACT Stuttering Rating Scale* provide clinicians information on fluency skills of children and young adults. By utilizing the *IMPACT Stuttering Rating Scale*, we are able to develop a better understanding as to how stuttering may impact students' academic performance and progress in school.

Code of Federal Regulations – Title 34: Education

34 C.F.R. §300.7 Child with a disability. (c) Definitions of disability terms. (11) Speech or language impairment means a communication disorder, such as stuttering, impaired articulation, a language impairment, or a voice impairment, that adversely affects a child's educational performance.

The Individual's with Disabilities Act (IDEA, 2004) states that when assessing a student for a speech or language impairment, we need to determine whether or not the impairment will negatively impact the child's educational performance. In order to determine whether an articulation or phonology impairment exists, we can collect a speech sample of the individual, and analyze intelligibility and the impact of the impairment on academic success.

Importance of Observations and Rationale for a Rating Scale

A speech and language evaluation should include systematic observations and a contextualized analysis that involves multiple observations across various environments and situations (Westby et al., 2003). According to IDEA (2004), such types of informal assessment must be used in conjunction with standardized assessments. Section. 300.532(b), 300.533 (a) (1) (I, ii, iii); 300.535(a)(1) of IDEA states that, "assessors must use a variety of different tools and strategies to gather relevant functional and developmental information about a child, including information provided by the parent, teacher, and information obtained from classroom-based assessments and observation." By using both formal and informal assessments, clinicians are able to capture a larger picture of a student's speech abilities. By observing a child's speech via informal observation, examinees (i.e., clinician, teacher, and parent) can observe the types of disfluencies a student makes, as well as the potential impact stuttering may have on a child's academic and social life. When we consider a formal fluency assessment, it may be difficult for clinicians to observe and gauge the impact of these errors on a student's everyday life. Parent and teacher input can be beneficial during a speech assessment because it allows for the observations to take place in an authentic setting. Additionally, the examiners are already familiar with the child and may know what to look for which, creates a true representation of the child's speech skills. *The IMPACT Stuttering Rating Scale* provides us with clinician, parent, and teacher observations and perspectives of a child's fluency ability. When given the guidelines of what to look for, parents will be able to provide numerous examples of their child's fluency ability or stuttering and the impact of stuttering. These disfluencies and the impact of these disfluencies may not be so easily observed during clinical assessment and observation. Furthermore, it can be important to obtain information on how a child engages with their family, friends, and peers during familiar tasks in order to gain ecologically and

culturally valid information on how a child functions and communicates on a day-to-day basis (Jackson, Pretti- Frontczak, Harjusola-Webb, Grisham-Brown, & Romani, 2009; Westby, Stevens, Dominguez, & Oetter, 1996).

During assessment and intervention planning, it is important to consider how stuttering may adversely affect educational performance and a child's social interactions. Previous research has suggested that stuttering can negatively impact a child's academic skills as well as their social and personal life. For example, students with stuttering may have difficulty with reading, and academic performance (Peterson, Pennington, Shriberg, & Boada, 2009; Bird, Bishop, Freeman, 1995; Nathan, Stackhouse, Goulandris, & Snowling, 2004; Anthony, Aghara, Dunkelberger, Anthony, Williams & Zhang, 2011). Additionally, these students may interact with their peers less due to fears of being made fun of or being bullied.

Administration and Scoring Procedures

The following testing guidelines represent specific administration and scoring procedures for the *IMPACT Stuttering Rating Scale*. These procedures are considered best professional practice required in any type of rating scale as described in the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, and NCME], 2014).

Examiner Qualifications

Professionals who are formally trained in the ethical administration, scoring, and interpretation of assessment tools and who hold appropriate educational and professional credentials may administer the *IMPACT Stuttering Rating Scale*. Qualified examiners include speech-language pathologists, clinical fellows and graduate students in speech-language pathology. It is a requirement to read and become familiar with the administration, recording, and scoring procedures before using this rating scale and asking parents and teachers to complete the rating scales.

Confidentiality Requirements

As described in Standard 6.7 of the Standards for Educational and Psychological Testing (AERA et al., 2014), it is the examiner's responsibility to protect the security of all testing material and ensure confidentiality of all testing results.

Eligibility for Testing

The *IMPACT Stuttering Rating Scale* is appropriate to use for individuals between the ages of 5-0 and 21-0 years of age. This rating scale is designed for individuals who are suspected of or who have been previously diagnosed with stuttering. The rating scale also addresses the potential impact that stuttering may have on a child.

Testing Time

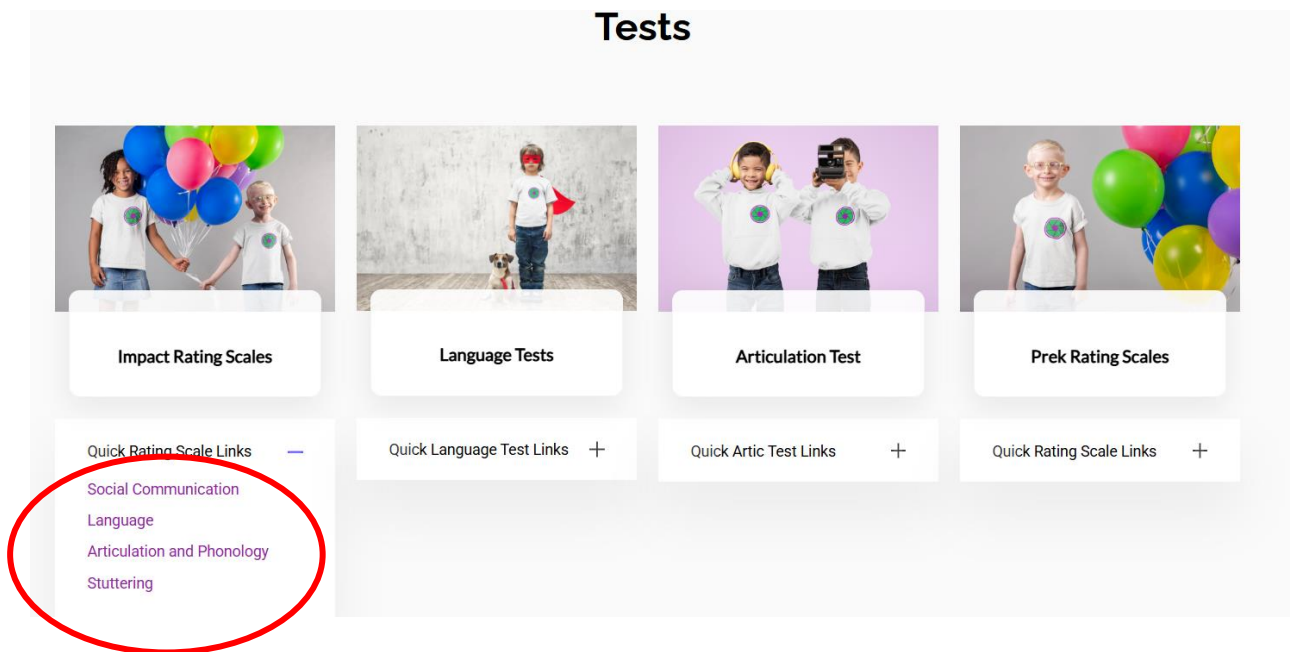
Administration of the clinician, teacher, and parent rating scale takes approximately 20-25 minutes respectively.

Test Materials

The *IMPACT Stuttering Rating Scale* consists of three observational rating scales, one for clinician, one for parent, and one for the teacher. All rating scales and scale converting software is available on the *SLP Platform* website at: www.slpplatform.com

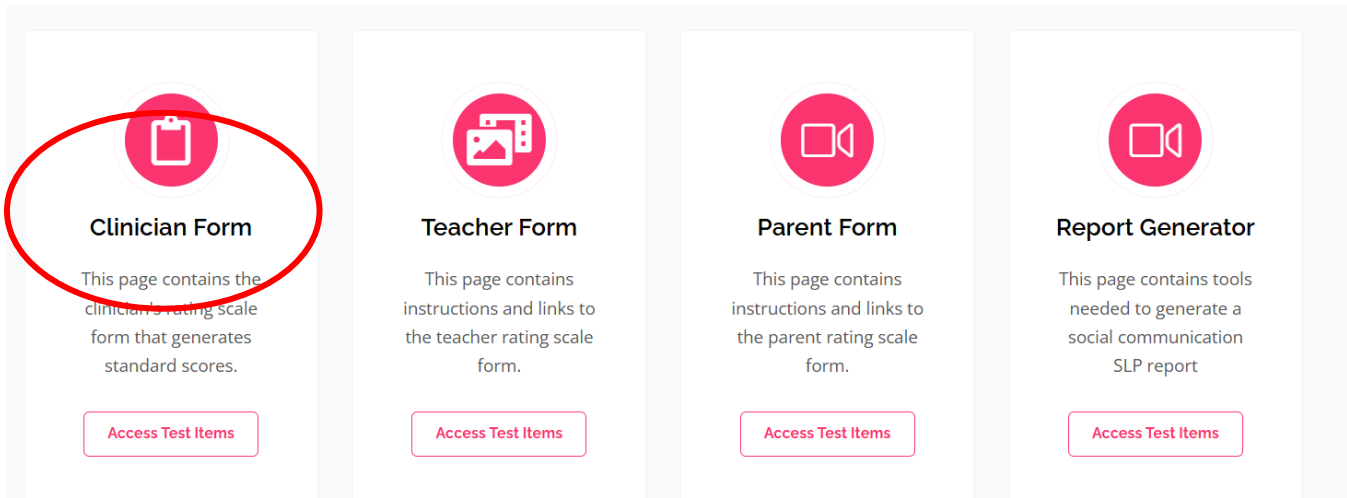
Accessing Clinician, Parent, and Teaching Rating Forms online

Begin by logging onto your account at www.slpplatform.com and select “Administer Tests”. Select the *IMPACT Stuttering Scale* as shown below,



Administration Instructions

Step 1/Clinician Form: Complete the Clinician Rating Scale. Please be sure to review the videos on the page to improve your understanding of what each test item is asking.



When you are finished filling out the form, click on the “Submit” button. The system will generate a scored protocol that contains standard scores and percentile ranks. Enter your own (the examiner’s) email address to receive a copy of the protocol and report by email.

Step 2/Teacher Form: Send an email/text message to the student’s teacher with the link to the “Teacher Rating Scale” that can be completed online. Explain to the teacher (a template of the email with the explanation is provided in step 2) that there are accompanying videos that he/she can watch that will provide examples of what each question is asking. After completing the rating scale, ask the teacher to type in your email address in the provided box (at the bottom of the form). Once the teacher completes the form, the system will generate and email you a scored protocol that contains standard scores and percentile ranks.

Step 3: Send an email/text message to the student’s parent(s) with the link to the “Parent Rating Scale” that can also be completed online. Explain to the parent (a template of the email with the explanation is provided in step 3) that there are accompanying videos that he/she can watch that will provide examples of what each question is asking. After completing the rating scale, ask the parent to type in your email address in the provided box (at the bottom of the form). Once the parent completes the form, the system will generate and email you a scored protocol that contains standard scores and percentile ranks.

Step 4: Use the optional report generator to assist you in writing the pragmatic language write-up portion of your evaluation.

Development, Standardization and Normative Information

This section describes the procedures followed in developing test items, implementing the pilot and normative study, and selecting the items for the final version of the test. This section also details the normative samples obtained to standardize and validate the IMPACT Stuttering Rating Scale. All test development and standardization project procedures were reviewed and approved by IntegReview IRB (now known as Advarra), a fully AAHRPP-accredited independent review board that provides ethical review for all phases of industry-sponsored and federally funded research in the U.S. Additionally, all test development and standardization methodology was based on best practices in research, and conducted in compliance with complex regulatory requirements, frameworks, and guidelines set forth by the Standards for Educational and Psychological Testing (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, and NCME], 2014).

Test Item Development

Selection of the test items began with an extensive review of research and theory related to stuttering, defining characteristics of successful communication, specific communication abilities and patterns required in the educational setting as well as an analysis of which communication behaviors are most predictive of stuttering and its impact on education for specific age groups. The literature reviewed consisted of research articles, textbooks, diagnostic tests and the diagnostic criteria for communication disorder (stuttering) from IDEA (2015). This analysis resulted in identification of 79 specific behaviors presumed to impact educational progress and to be indicative of stuttering. Next, the test items were reviewed and edited for clarity and face validity for use by clinicians, teachers and parents. The systematic review of the test items was completed by a panel of 9 experts in the area of speech pathology (specifically, stuttering). The panel also included 7 teachers and 11 parents of children who stutter. After receiving their feedback, some items were rewritten, dropped or rephrased.

The test was developed in three phases: pilot study, normative study, and national standardization. The procedures for each phase are detailed below.

Pilot Study

The pilot study was conducted to determine the appropriateness of questions and to review all test instructions. The pilot study included 64 children from the ages of 5:0 to 12:11. The sample was 19% Hispanic, 11% African American, 52% White, 5% Asian and 13% other ethnicities (60% males and 40% females). The pilot study included 75% typically developing children and 25% children with identified social communication disorder.

The rating scale responses were coded. These data were factor analyzed. From the results of this analysis, a scale of seven factors containing 40 items was produced. Cronbach's coefficient alphas were computed and results indicated the alphas were sufficiently large to provide support for the test reliability. The results of the pilot study were found to be effective for test item selection.

Normative Study

Following the pilot study, a normative study was conducted to establish norms for IMPACT Stuttering Rating Scale by testing typically developing children representative of the general U.S. population. A clinical group was included for validation purposes. Additional goals of the normative study included investigation of optimal weighted scoring system/criteria as well as optimal test administration time. The study reviewed administrative and scoring procedures preliminary to national standardization. The test content was evaluated both qualitatively and quantitatively for bias.

The normative study included 86 children from the ages of 5:0 to 15:11. The sample was 11% Hispanic, 10% African American, 56% White, 8% Asian and 17% other ethnicities (60% males and 40% females). The pilot study included 88% typically developing children and 12% children who stutter (clinical group). The mean for the outcome variables were compared between the clinical and the typically developing groups of examinees using Kruskal Wallis analysis of variance (ANOVA). Further comparisons in mean scores between the groups were examined using Mann-Whitney U test. The level of significance was set at $p \leq 0.05$. Further comparisons using Mann-Whitney U test showed that there was a significant difference among all the study groups ($p < 0.001$).

Based on the feedback of all examinees, some test items were modified, while others were removed altogether. The test directions and scoring procedures were fine-tuned. Suggestions of the field test examiners were thoroughly reviewed prior to the national standardization.

National Standardization

One of the ways we can tell if an assessment is a strong test, is if it includes adequate norms. Norm-referenced testing is a method of evaluation where an individual's scores on a specific test are compared to scores of a group of test-takers (e.g., age norms) (AERA, APA, and NCME, 2014). Previous research has suggested that utilizing a normative sample can be beneficial in the identification of a disability. Additionally, research has suggested that the inclusion of children with disabilities in the normative sample may negatively impact the test's ability to differentiate between children with disorders and

children who are typically developing (Peña, Spaulding, & Plante, 2006). When reviewing a test’s normative sample, it is important to consider size, gender, race and ethnicity, age, geographic location, and whether individuals with disabilities were included in the normative sample.

The normative data for the *IMPACT Stuttering Rating Scale* is based on the performance of 1396 examinees across 11 age groups (shown in Table 4.1) from 17 states across the United States of America (Arizona, California, Colorado, Nevada, Idaho, Illinois, Iowa, Kansas, Ohio, Minnesota, Florida, New York, Pennsylvania, Florida, South Carolina, Texas, Washington).

Age Group	Age	N	%
1	5-0 to 5-11	135	9.5
2	6-0 to 6-11	126	9
3	7-0 to 7-11	134	9.5
4	8-0 to 8-11	120	9
5	9-0 to 9-11	119	8.5
6	10-0 to 10-11	126	9
7	11-0 to 11-11	131	9
8	12-0 to 12-11	118	8.5
9	13-0 to 13-11	125	9
10	14-0 to 14-11	120	9
11	15-0 to 21-0	142	10
Total Sample		1396	100%

The data was collected throughout the 2016-2020 school years by 34 state licensed speech-language pathologists (SLPs). The SLPs were recruited through Go2Consult Speech and Language Services, a speech-language pathology services and nonpublic agency certified by the CA Department of Education in conjunction with the Lavi Institute, an ASHA approved CE provider. All standardization project procedures were reviewed and approved by IntegReview IRB (now known as Advarra), a fully AAHRPP-accredited independent review board that provides ethical review for all phases of industry-sponsored and federally funded research in the U.S. To ensure representation of the national population, the *IMPACT Stuttering Rating Scale* standardization sample was selected to match the US Census data reported in the ProQuest Statistical Abstract of the United States (ProQuest, 2017). The sample was stratified within each age group by the following criteria: gender, race or ethnic group, and geographic region. The demographic table below (Table 4.2) specifies the distributions of these characteristics and shows that the normative sample is nationally representative.

Table 4.2			
Demographics of the Normative Sample vs. US Population			
Normative Sample Size = 1403			
Demographic	<i>N</i> Normative Sample	% Normative Sample	% US Population
Gender			
Male	713	51%	49%
Female	683	49%	51%
Total	1395	100%	100%
Race			
White	882	63%	77%
Black	194	14%	13%
Asian	70	5%	4%
Other	69	5%	6%
Hispanic	181	13%	12%
Total	1396	100%	100%
Clinical Groups			
	none	none	none
US Regions			
Northeast	208	15%	16%
Midwest	293	21%	22%
South	490	35%	38%
West	405	29%	24%
Total	1396	100%	100%
Parents' Educational Level			
Four years of college or more	420	30	31%
Some college	391	28	27%
High school graduate	405	29	30%
Less than high school graduate	179	13	12%
Total	1396	100%	100%

Criteria for inclusion in the normative sample

A strong assessment is one that provides results that will benefit the individual being tested or society as a whole (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education [AERA, APA, and NCME], 2014). One way we can tell if an assessment is strong, is if it includes adequate norms. Previous research has suggested that utilizing a normative sample can aid in the identification of a disability. Research has also suggested that the inclusion of children with disabilities may negatively impact the test's ability to differentiate between

children with disorders and children who are typically developing (Peña, Spaulding, & Plante, 2006). Since the purpose of the *IMPACT Stuttering Rating Scale* is to help to identify students who stutter, it was critical to exclude students from the normative sample who have diagnoses that are known to influence fluency of speech (Peña, Spaulding, & Plante, 2006). Students who had previously been diagnosed with articulation, phonological impairments, or motor planning deficits were not included in the normative sample. Further, students were excluded from the normative sample if they were diagnosed with autism spectrum disorder, intellectual disability, hearing loss, neurological disorders, or genetic syndromes. In order for students to be included in the normative sample for this assessment tool, students must have met criteria of fluent and typically developing speech, and show no evidence of stuttering. Thus, the normative sample for the *IMPACT Stuttering Rating Scale* provides an appropriate comparison group (i.e., a group without any known disorders that might affect fluency of speech) against which to compare students with the suspected disorder.

The *IMPACT Stuttering Rating Scale* is designed for students who are native speakers of English and/or are English language learners (ELL) who have demonstrated a proficiency in English based on state testing scores and school district language evaluations. Additionally, students who were native English speakers and also spoke a second language were included in this sample.

Norm-referenced testing is a method of evaluation where an individual's scores on a specific test are compared to scores of a group of test-takers (e.g., age norms) (AERA, APA, and NCME, 2014). Clinicians can compare clinician, teacher, and parent ratings on the *IMPACT Stuttering Rating Scale* to this normative sample to determine whether a student is scoring within normal limits or, if their scores are indicative of stuttering. Administration, scoring, and interpretation of the *IMPACT Stuttering Rating Scale* must be followed in order to make comparisons to normative data. This manual provides instructions to guide examiners in the administration, scoring, and interpretation of the rating scale.

Validity and Reliability

This section of the *IMPACT Stuttering Rating Scale* manual provides information on the psychometric characteristics of validity and reliability. Validity helps establish how well a test measures what it is supposed to measure and reliability represents the consistency with which an assessment tool measures a certain ability or skill. The first half of this chapter will evaluate content, construct, criterion, and clinical validity of the *IMPACT Stuttering Rating Scale*. The second half of the chapter will review the consistency and stability of the *IMPACT Stuttering Rating Scale* scores, in addition to test retest and inter-rater reliability.

Validity

Validity is essential when considering the strength of a test. Content validity refers to whether the test provides the clinician with accurate information on the ability being tested. Specifically, content validity measures whether or not the test actually assesses what it's supposed to. According to McCauley and Strand (2008), there should be a rationalization of the methods used to choose content, expert evaluation of the test's content, and an item analysis.

Content-oriented evidence of validation addresses the relationship between a student's learning standards and the test content. Specifically, content-sampling issues look at whether cognitive demands of a test are reflective of the student's learning standard level. Additionally, content sampling may address whether the test avoids inclusion of features irrelevant to what the test item is intended to target.

Single-cut Scores

It is common to use single cut scores (e.g., -1.5 standard deviations) to identify disorders, however, there is evidence that advises against using this practice (Spaulding, Plante, & Farinella, 2006). When using single cut scores (e.g., -1.5 SD, -2.5 SD, etc.), we may under identify students with impairments on tests for which the best-cut score is higher and over identify students' impairments on tests for which the best-cut score is lower. Additionally, using single cut scores may go against IDEA's (2004) mandate, which states assessments must be valid for the purpose for which they are used.

Inclusion/Exclusion Criteria for the Discriminant Analysis and the Group Differences Study

Typically developing participants were selected based on the following criteria: 1) exhibited hearing sensitivity within normal limits; 2) presented with age-appropriate speech and language skills; 3) successfully completed each school year with no academic failures; and 4) attended public school and placed in general education classrooms.

Inclusion criteria for the stuttering group was: 1) having a current diagnosis of stuttering (based on medical records and/or school-based special education eligibility criteria); 2) currently attending a local public school, and enrolled in the general education classroom; and 3) exhibited hearing sensitivity within normal limits.

Sensitivity and Specificity

Table 5.1 shows the cut scores needed to identify stuttering within each age range. Additionally, this table demonstrates the sensitivity and specificity information that indicates the accuracy of identification at these cut scores. Sensitivity and specificity are diagnostic validity statistics that explain how well a test performs. Vance and Plante (1994) set forth the standard that for an assessment to be considered clinically beneficial, it should reach at least 80% sensitivity and specificity.

Thus, strong sensitivity and specificity (i.e., 80% or stronger) is needed to support the use of a test in its identification of the presence of a disorder or impairment. Sensitivity measures how well the assessment will accurately identify those who truly have a disorder (Dollaghan, 2007). If sensitivity is high, this indicates that the test is highly likely to identify the speech sound disorder, or, there is a low chance of “false positives.” Specificity measures the degree to which the assessment will accurately identify those who do not have a disorder, or how well the test will identify those who are “typically developing” (Dollaghan, 2007).

Table 5.1 IMPACT Stuttering Rating Scale sensitivity, specificity and likelihood ratios**Clinician Rating Scale**

Age group	Cut score	Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio
5:0-5:11	77	.81	.90	4.39	.09
6:0-6:11	77	.80	.91	4.81	.08
7:0-7:11	77	.84	.89	5.91	.09
8:0-8:11	78	.94	.84	4.45	.09
9:0-9:11	78	.93	.91	4.10	.78
10:0-10:11	77	.91	.89	4.56	.12
11:0-11:11	77	.89	.85	4.89	.11
12:0-12:11	78	.92	.86	4.45	.09
13:0-13:11	77	.89	.88	4.67	.11
14:0-14:11	78	.90	.83	4.78	.16
15:0-15:11	78	.94	.87	4.10	.19
16:0-21:0	77	.94	.92	5.34	.12

Note: Age groups 16:0-21:0 are reported together as there were no age-related changes detected after the age of 16. Total N=1930; typically developing group n=1396; clinical group=534

Teacher Rating Scale

Age group	Cut score	Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio
5:0-5:11	77	.89	.82	4.65	.12
6:0-6:11	77	.90	.83	4.11	.18
7:0-7:11	77	.92	.85	3.88	.09
8:0-8:11	78	.89	.88	4.34	.15
9:0-9:11	77	.87	.81	4.15	.09
10:0-10:11	77	.89	.96	4.23	.14
11:0-11:11	78	.86	.91	4.45	.24
12:0-12:11	78	.94	.88	5.20	.18
13:0-13:11	77	.93	.85	5.11	.09
14:0-14:11	78	.80	.87	5.45	.11
15:0-15:11	77	.89	.86	4.39	.77
16:0-21:0	77	.93	.89	5.40	.13

Note: Age groups 16:0-21:0 are reported together as there were no age-related changes detected after the age of 16. Total N=1930; typically developing group n=1396; clinical group=534

Table 5.1 IMPACT Stuttering Rating Scale sensitivity, specificity and likelihood ratios**Parent Rating Scale**

Age group	Cut score	Sensitivity	Specificity	Positive likelihood ratio	Negative likelihood ratio
5:0-5:11	77	.86	.89	4.05	.23
6:0-6:11	77	.91	.89	4.16	.13
7:0-7:11	77	.91	.95	4.67	.28
8:0-8:11	77	.96	.86	4.11	.13
9:0-9:11	77	.83	.84	4.12	.24
10:0-10:11	77	.94	.82	4.44	.17
11:0-11:11	78	.89	.89	4.45	.23
12:0-12:11	78	.89	.93	4.12	.34
13:0-13:11	77	.96	.93	5.40	.21
14:0-14:11	77	.91	.92	4.08	.29
15:0-15:11	77	.89	.89	4.12	.25
16:0-21:0	77	.92	.89	4.27	.19

Note: Age groups 16:0-21:0 are reported together as there were no age-related changes detected after the age of 16. Total N=1930; typically developing group n=1396; clinical group=534

Content Validity

The validity of a test determines how well the test measures what it purports to measure. Validity can take various forms, both theoretical and empirical. This can often compare the instrument with other measures or criteria, which are known to be valid (Zumbo, 2014). For the content validity of the test, expert opinion was solicited. Twenty-two speech language pathologists (SLPs) reviewed the *IMPACT Stuttering Rating Scale*. All SLPs were licensed in the state of California, held the Clinical Certificate of Competence from the American Speech-Language-Hearing Association, and had at least 5 years of experience in assessment of children who stutter. Each of these experts was presented with a comprehensive overview of the rating scale descriptions, as well as rules for standardized administration and scoring. They all reviewed 6 full-length administrations. Following this, they were asked 30 questions related to the content of the rating scale and whether they believed the assessment tool to be an adequate measure of stuttering. For instance, their opinion was solicited regarding whether the questions and the raters' responses properly evaluated the impact of stuttering on educational performance and social interaction. The reviewers rated each rating scale on a decimal scale. All reviewers agreed that the *IMPACT Stuttering Rating Scale* is a valid informal observational measure to evaluate fluency and to determine the impact on educational performance and social interaction, in students who are between the ages of 5 and 21 years old. The mean ratings for the Clinician, Teacher, and Parent rating scales were 28.2±0.9, 29.2±0.9, 28.4±0.7, respectively.

Criterion Validity

In assessing criterion validity, a correlation analysis was not possible for the IMPACT Stuttering Rating Scale when compared to the current body of rating scales. The IMPACT Stuttering Rating Scale is unique in its content and design. This rating scale cannot be compared to the existing body of rating scales because of its unique focus which is not available within other rating scales.

Group Differences

Since a stuttering assessment tool is designed to identify those examinees who stutter, it would be expected that individuals identified as likely to exhibit stuttering would score lower than those who are typically developing. The mean for the outcome variables (Clinician, Teacher, and Parent ratings) were compared among the three clinical groups and the typically developing group of examinees using Kruskal Wallis analysis of variance (ANOVA). The level of significance was set at $p \leq 0.05$. Table 5.3 reviews the ANOVA, which reveals a significant difference between all three groups.

Table 5.3 Clinician, Teacher and Parent Rating Scale Comparison across Clinical and Typically-Developing groups (N=188)

	ST group (n=86)	TD group (n=102)	p-value*
Clinician	18 (4.2)	34 (0.8)	<.001
Teacher	20 (3.2)	29 (0.6)	<.001
Parent	21 (4.1)	28 (0.7)	<.001

Abbreviation: ST – stuttering, TD – typically developing:

*Kruskal-Wallis Analysis of Variance test

Standards for fairness

Standards of fairness are crucial to the validity and comparability of the interpretation of test scores (AERA, APA, and NCME, 2014). The identification and removal of construct-irrelevant barriers maximizes each test-taker's performance, allowing for skills to be compared to the normative sample for a valid interpretation. Test constructs and individuals or subgroups of those who the test is intended for must be clearly defined. In doing so, the test will be free of construct-irrelevant barriers as much as possible for the individuals and/or subgroups the test is intended for. It is also important that simple and clear instructions are provided.

Response Bias

A bias is defined as a tendency, inclination, or prejudice toward or against something or someone. For example, if you are interviewing for a new employer and asked to complete a personality questionnaire, you may answer the questions in a way that you think will impress the employer. These responses will of course impact the validity of the questionnaire.

Responses to questionnaires, tests, scales, and inventories may also be biased for a variety of reasons. Response bias may occur consciously or unconsciously, it may be malicious or cooperative, self-

enhancing or self-effacing (Furr, 2011). When response bias occurs, the reliability and validity of our measures is compromised. Diminished reliability and validity will in turn impact decisions we make regarding our students (Furr, 2011). Thus, psychometric damage may occur because of response bias.

Types of Response Biases

Acquiescence Bias ("Yea-Saying and Nay-Saying") refers to when an individual consistently agrees or disagrees with a statement without considering what the statement means (Danner & Rammstedt, 2016).

Extremity Bias refers to when an individual consistently over or underuses "extreme" response options, regardless of how the individual feels towards the statement (Wetzel, Lüdtkke, Zettler, & Bohnke, 2016).

Social desirability Bias refers to when an individual responds to a statement in a way that exaggerates his or her own positive qualities (Paulhus, 2002).

Malingering refers to when an individual attempts to exaggerate problems, or shortcomings (Rogers, 2008). *Random/careless responding* refers to when an individual responds to items with very little attention or care to the content of the items (Crede, 2010).

Guessing refers to when the individual is unaware of or unable to gage the correct answer regarding their own or someone else's ability, knowledge, skill, etc. (Foley, 2016).

In order to protect against biases, balanced scales are utilized. A balanced scale is a test or questionnaire that includes some items that are positively keyed and some items that are negatively keyed. For example, the *IMPACT Stuttering Rating Scale* items are rated on a 4-point scale ("never," "sometimes," "often," and "typically"). Now, imagine if we ask a teacher to answer the following two items regarding one of their students:

1. The student appears confident and eager to communicate when socializing with peers.
2. The student does not appear reserved and/or shy when socializing with peers.

Both of these items are positively keyed because a positive response indicates a stronger level of confidence in speech ability. To minimize the potential effects of acquiescence bias, the researcher may revise one of these items to be negatively keyed. For example:

1. The student appears reserved and/or shy when socializing with peers.
2. The student appears confident and eager to communicate when socializing with peers.

Now, the first item is keyed positively and the second item is keyed negatively. The revised scale, which represents a balanced scale, helps control acquiescence bias by including one item that is positively keyed and one that is negatively keyed. If the teacher responded highly on both items, the teacher may be viewed as an acquiescent responder (i.e., the teacher is simply agreeing to items without regard for the content). If the teacher responds high on the first item, and responds low on the second item, we know that the teacher is reading each test item carefully and responding appropriately.

For a balanced scale to be useful, it must be scored appropriately, meaning the key must accommodate the fact that there are both positively and negatively keyed items. To achieve this, the rating scale must keep track of the negatively keyed items and "reverse the score." Scores are only reversed for negatively keyed items. For example, on the negatively keyed item above, if the teacher scored a 1 ("never") the score should be converted to a 4 ("typically") and if the teacher scored a 2 ("sometimes") the score should be converted to a 3 ("often"). Similarly, the researcher recodes responses of 4 ("typically") to 1

(“never”) and 3 (“often”) to 2 (“sometimes”). Balanced scales help researchers differentiate between acquiescent responders and valid responders. Therefore, test users can be confident that the individual reporting is a reliable and valid source.

Inter-rater Reliability

Inter-rater reliability measures the extent to which consistency is demonstrated between different raters with regard to their scoring of examinees on the same instrument (Osborne, 2008). For the *IMPACT Stuttering Rating Scale*, inter-rater reliability was evaluated by examining the consistency with which the raters are able to follow the test scoring procedures. Two clinicians, two teachers, and two caregivers simultaneously rated students. The results of the scorings were correlated. The coefficients were averaged using the z-transformation method. The resulting correlations for the subtests are listed in Table 5.5.

<i>Rating Scale</i>	<i>Reliability</i>
<i>Clinician (N=20)</i>	<i>.84</i>
<i>Teacher (N=20)</i>	<i>.91</i>

Test-Retest Reliability

This is a factor determined by the variation between scores or different evaluative measurements of the same subject taking the same test during a given period of time. If the test is a strong instrument, this variation would be expected to be low (Osborne, 2008). The *IMPACT Stuttering Rating Scale* was completed with 47 randomly selected examinees, ages 5-0 through 21-0 over two rating periods. The interval between the two periods ranged from 16 to 20 days. To reduce recall bias, the examiners did not inform the raters at the time of the first rating session that they would be rating again. All subsequent ratings were completed by the same examiners who administered the test the first time. The test-retest coefficients for the three rating scales were all greater than .80 indicating strong test-retest reliability for the *IMPACT Stuttering Rating Scale*. The results are listed in Table 5.6

Age Groups	N	1st Test		2nd Test		Correlation Coefficient
		Mean	SD	Mean	SD	
1,2, & 3	21					
Clinician		32	2	33	1	0.92
Teacher		28	2	29	1	0.96
Parent		28	1	28	1	0.91
4,5, & 6	21					
Clinician		34	1	34	1	0.93
Teacher		29	1	29	1	0.89
Parent		28	2	29	1	0.93
7, 8, 9, 10 & 11	21					

Clinician		34	1	34	1	0.94
Teacher		29	1	30	1	0.88
Parent		29	1	29	1	0.84

Internal Consistency

Internal consistency ensures that all items within the scale are measuring the same construct (i.e., social communication behavior as it relates to educational performance), and that they are related to each other and consistently contribute to the overall score, thereby providing a reliable and accurate representation of the attribute being measured. Table 5.7 shows the results for each of the samples.

Table 5.7						
Internal Consistency						
	Clinician		Teacher		Parent	
Age Groups	n	Alpha	n	Alpha	n	Alpha
5:0-6:11	54	.93	49	.90	62	.89
7:0-8:11	46	.91	55	.93	51	.94
9:0-10:11	67	.90	63	.98	80	.93
11:0-11:11	64	.93	69	.92	68	.92
12:0-15:11	71	.89	58	.93	52	.94
16:0-21:0	62	.94	64	.89	61	.96

Highlights of the IMPACT Stuttering Rating Scale

The results of the *IMPACT Stuttering Rating Scale* provide information on a student's fluency of speech, and how stuttering may impact children and adolescents' success in school and social situations. Data obtained from the *IMPACT Stuttering Rating Scale* is useful in determining eligibility criteria for a students who stutter.

Strong Psychometric Properties

The *IMPACT Stuttering Rating Scale* was normed on a nationwide standardization sample of 1396 examinees. The sample was stratified to match the most recent U.S. Census data on gender, race/ethnicity, and region. Please refer to Chapter 3 for more information of the standardization process.

The *IMPACT Stuttering Rating Scale* areas have strong sensitivity and specificity (above 80%), high internal consistency, and test-retest reliabilities. Criterion-related validity studies were conducted during standardization, with over 200 participants. Please refer to Chapter 4 for more information on the summary results of the reliability and validity studies.

Ease and Efficiency of Administration and Scoring

The *IMPACT Stuttering Rating Scale* consists of three observational rating scales, one for clinician, one for parent, and one for the teacher. All *IMPACT* rating scales and scale converting software is available on the SLP Platform website. Rating scale item clarification videos are also provided on this website. Additionally, an instructional email with a link to the website and rating form is prepared for your convenience to send to teacher and parents. Please review Chapter 2 for more information on the easy and effective administration process.

References:

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*. Washington, DC: AERA.
- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Arlington, VA: Author.
- Anthony, J. L., Aghara, R. G., Dunkelberger, M. J., Anthony, T. I., Williams, J. M., & Zhang, Z. (2011). What factors place children with speech sound disorder at risk for reading problems? *American Journal of Speech-Language Pathology*, 20(2), 146–160. [https://doi.org/10.1044/1058-0360\(2011/10-0053\)](https://doi.org/10.1044/1058-0360(2011/10-0053))
- Bird, J., Bishop, D. V. M., & Freeman, N. H. (1995). Phonological awareness and literacy development in children with expressive phonological impairments. *Journal of Speech and Hearing Research*, 38, 446–462.
- Crede, M. (2010). Random responding as a threat to the validity of effect size estimates in correlational research. *Educational and Psychological Measurement*, 70, 596–612.
- Danner, D. & Rammstedt, B. (2016). Facets of acquiescence: Agreeing with negations is not the same as accepting inconsistency. *Journal of Research in Personality*, 65, 120–129. <https://doi.org/10.1016/j.jrp.2016.10.010>
- Dodd, B., Hua, Z., Crosbie, S., Holm, A., & Ozanne, A. (2006). *Diagnostic Evaluation of Articulation and Phonology* (U.S. ed.). San Antonio, TX: Pearson.
- Dollaghan, C. A. (2007). *The Handbook for Evidence-Based Practice in Communication Disorders*. Baltimore, MD: MD Brookes.
- Felsenfeld, S., Broen, A., & McGue, M. (1992). A 28-year follow-up of adults with a history of moderate phonological disorder: Linguistic and personality results. *Journal of Speech and Hearing Research*, 35(5), 1114–1125. <https://doi.org/10.1044/jshr.3505.1114>
- Felsenfeld, S., Broen, A., & McGue, M. (1994). A 28-year follow-up of adults with a history of moderate phonological disorder: Educational and occupational results. *Journal of Speech and Hearing Research*, 37(6), 1341–1353. <https://doi.org/10.1044/jshr.3706.1341>
- Foley, P. B. (2016). Getting Lucky: How Guessing Threatens the Validity of Performance Classifications. *Practical Assessment, Research & Evaluation*, 21.3, 1–23.
- Fudala, J. B. & Stegall, S. (2017). *Arizona Stuttering Scale - Fourth Edition (Arizona-4)*. Torrance, CA: Western Psychological Services.
- Furr, R. M. & Bacharach, V. R. (2008). *Psychometrics: An introduction*. Thousand Oaks, CA: Sage.
- Individuals with Disabilities Education Act. (2004). Section 300.8 child with a disability. Retrieved

from <https://sites.ed.gov/idea/regs/b/a/300.8>

- Jackson S., Pretti-Frontczak K., Harjusola-Webb S., et al. (2009) Response to intervention: implications for early childhood professions. *Language, Speech and Hearing Services in Schools* 40: 424–434.
- Leitão, S., & Fletcher, J. (2004). Literacy outcomes for students with speech impairment: Long-term follow-up. *International Journal of Language and Communication Disorders*, 39(2), 245–256. <https://doi.org/10.1080/13682820310001619478>
- McCauley, R. J., & Strand, E. A. (2008). A review of standardized tests of nonverbal oral and speech motor performance in children. *American Journal of Speech-Language Pathology*, 17(1), 81–91. [https://doi.org/10.1044/1058-0360\(2008/007\)](https://doi.org/10.1044/1058-0360(2008/007))
- McCormack, J., McLeod, S., McAllister, L., & Harrison, L. J. (2009). A systematic review of the association between childhood speech impairment and participation across the lifespan. *International Journal of Speech-Language Pathology*, 11(2), 155–170. <https://doi.org/10.1080/17549500802676859>
- McCormack, J., McLeod, S., McAllister, L., & Harrison, L. J. (2010). My speech problem, your listening problem, and my frustration: The experience of living with childhood speech impairment. *Language, Speech, and Hearing Services in Schools*, 41(4), 379–392. [https://doi.org/10.1044/0161-1461\(2009/08-0129\)](https://doi.org/10.1044/0161-1461(2009/08-0129))
- McLeod, S., & Baker, E. (2017). *Children’s speech: An evidence-based approach to assessment and intervention*. Boston, MA: Pearson.
- McLeod, S., Daniel, G. R., & Barr, J. (2013). When he’s around his brothers ... he’s not so quiet”: The private and public worlds of school-aged children with speech sound disorder. *Journal of Communication Disorders*, 46(1), 70–83. <https://doi.org/10.1016/j.jcomdis.2012.08.006>
- Mullen, R., & Schooling, T. (2010). The National Outcomes Measurement System for pediatric speech-language pathology. *Language, Speech, and Hearing Services in Schools*, 41, 44–60. [https://doi.org/10.1044/0161-1461\(2009/08-0051\)](https://doi.org/10.1044/0161-1461(2009/08-0051))
- Osborne, J. W. (Ed.). (2008). *Best practices in quantitative methods*. Los Angeles: Sage Publications.
- Paulhus, D.L. (2002). Socially desirable responding: The evolution of a construct. In H.I. Braun & D.N. Jackson (Eds.), *The role of constructs in psychological and educational measurement* (pp. 37–48). Mahwah, NJ: Erlbaum.
- Peña, E. D., Spaulding, T. J., & Plante, E. (2006). The composition of normative groups and diagnostic decision making: Shooting ourselves in the foot. *American Journal of Speech-Language Pathology*, 15(3), 247–254. [https://doi.org/10.1044/1058-0360\(2006/023\)](https://doi.org/10.1044/1058-0360(2006/023))
- Plante, E. and Vance, R. (1994) Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools* 25, 15–24.

- ProQuest. (2017). *The statistical abstract of the United States, 2017*. Lanham, MD: Bernan Press, an imprint of The Rowman & Littlefield Publishing Group, Incorporated.
- Spaulding, T.J., Plante, E., & Farinella, K.A. (2006). Eligibility criteria for language impairment: Is the low end of normal always appropriate? *Language, Speech, and Hearing Services in Schools*, 37(1), 61-72.
- Sweeting, H., & West, P. (2001). Being different: Correlates of the experience of teasing and bullying at age 11. *Research Papers in Education*, 16(3), 225–246.
<https://doi.org/10.1080/02671520110058679>
- Westby, C. E., Stevens-Dominguez, M., & Oetter, P. (1996). A performance/competence model of observational assessment. *Language, Speech, and Hearing in the Schools*, 27, 144–156.
- Wetzel, E., Lüdtke, O., Zettler, I., & Bohnke, J. R. (2016). The stability of extreme response style and acquiescence over 8 years. *Assessment*, 23, 279–291.
- Zumbo, B. D. & Chan, E. K. (2014). *Validity and Validation in Social, Behavioral, and Health Sciences*. Springer.