

Assessment of Egyptian children who stutter using the standardized Arabic form of the Test of Childhood Stuttering

Rania Abdou

Department of Otorhinolaryngology, Faculty of Medicine, Alexandria University, Chompollion Street, Azareeta, Egypt

Correspondence to Rania Abdou,
Department of Otorhinolaryngology,
Faculty of Medicine, Alexandria University,
Chompollion Street, Azareeta, Egypt
Tel: 01005688879
e-mail: rania_abdou@hotmail.com

Received 27 February 2015

Accepted 28 April 2015

The Egyptian Journal of Otolaryngology
2015, 31:180–187

Introduction

Stuttering is one of the most common speech disorders affecting children. The proper assessment of stuttering has been a point of interest for researchers over the years. The aim of this study was to adapt the Test of Childhood Stuttering (TOCS) to suit the Egyptian culture and to test the validity and reliability of application of this test for the assessment of Egyptian children who stutter, for proper management of this ailment.

Patients

This study included 130 children. The ages ranged between 4 years and 12 years 11 months among both boys and girls. The children were divided into two groups: group 1 included 30 children who stuttered; group 2 included 100 typically developing children as the control group, who did not suffer from any language or speech disorder.

Methods

A pilot study was conducted on eight children who stuttered, after Arabic translation of the TOCS, and a few modifications were made to the pictures of the picture book of the TOCS to suit the Egyptian culture. Reliability was assessed by the test–retest method, and test validity was established on the basis of content description validity, internal consistency validity, convergent validity, and known group comparisons.

Results

The Arabic version of the TOCS is a valid and reliable test for the evaluation of Arabic speaking children who stutter and to determine the severity of a child's stuttering. The results were highly significant and were capable of discriminating between children with normal disfluency and those who stutter.

Keywords:

Arabic language, assessment, stuttering

Egypt J Otolaryngol 31:180–187

© 2015 The Egyptian Oto - Rhino - Laryngological Society
1012-5574

Introduction

Stuttering is a multifactorial speech disorder characterized by repetition of sounds or syllables, sound prolongations, interjection, broken words, blocking of sounds, word substitutions, or excessive physical tension during speech production [1]. Other coexistent symptoms may include facial grimacing, tremors of muscles involved in speech, and eye blinks, as well as avoidance of words or situations that exacerbate stuttering episodes [2,3]. A total of 80–90% of developmental stuttering begins by 6 years of age and affects about 5% of children [4,5]. Spontaneous recovery occurs in about 75% of individuals [6].

Development of a comprehensive tool in the Arabic language is needed to tackle more detailed parts of the problem of stuttering. The Test of Childhood Stuttering (TOCS) [7] serves as an objective and standardized measure for identifying children who stutter, and it has four principal uses:

- (i) To identify children who stutter,
- (ii) To determine the severity of a child's stuttering,

- (iii) To document changes in a child's fluency functioning over time, and
- (iv) For research on childhood stuttering.

The basis of our study was to use a new scale that could elicit speech disfluency in children under certain situations such as time pressure, by using sentences that contain developmentally advanced syntactic structures, by utilizing a context that is similar to the types of conversational exchanges that the children are likely to encounter in school settings, and by producing narratives. The study is designed to test the validity and reliability of this scale.

Aim of the work

The aim of this study was to adapt the TOCS to suit the Egyptian culture and to assess its validity and reliability in assessing Egyptian children who stutter, for proper management of this ailment.

Participants

This study included 130 children attending the Unit of Phoniatrics, in the outpatient clinic of

Alexandria Main University Hospital, during the year 2013.

All the children were physically fit. Children with acquired stuttering, brain damage, mental retardation, hearing impairment, highly unintelligible speech, psychiatric problems, a history of clinical and subclinical fits, or with any medical condition that seemed to interfere with proper analysis of the patient were excluded. Their age ranged between 4 years and 12 years 11 months. They were divided into two groups: group 1 included 30 children who stuttered; group 2 included 100 typically developing children as the control group, who were not suffering from any language or speech disorder. The pilot study included eight children. The proposed Arabic version of the TOCS was applied to the children of the pilot group to validate the materials used in the test (clear or not, suitable or not, valid to illustrate the item or not).

Methods

All studied individuals were subjected to the protocol of assessment of stuttering, which included assessment of stuttering severity using the Stuttering Severity Instrument-3 (SSI-3) [8].

The components of TOCS [7] include the following:

- (1) The standardized speech fluency measure.
- (2) Observational rating scales.
- (3) Supplemental clinical assessment activities.

The reliability of the test was assessed by test–retest after 8–12 days. The validity of the test was assessed on the basis of content description validity, internal consistency validity, convergent validity, and known group comparisons.

Informed written consent for participation in the study was obtained from the parents and/or the legally caring surrogates of the children; in addition, the child's assent was also obtained.

Data were analyzed using IBM SPSS software package version 20.0 (SPSS Inc., Chicago, Illinois, USA). Comparison between different groups with regard to categorical variables was performed using the χ^2 -test, Monte Carlo correction, or Fisher's exact test. For normally distributed data, two independent populations were compared using an independent t-test. For abnormally distributed data, the Mann–Whitney test was used to analyze stuttering and normal groups. Correlations between two quantitative and ordinal variables were assessed

using Spearman's coefficient. Sensitivity, specificity, positive predictive value, negative predictive value and accuracy were calculated. A receiver operating characteristic (ROC) curve was plotted to analyze the recommended cutoff; the area under the ROC curve (AUC) denotes the diagnostic performance of the test. An area greater than 50% represents acceptable performance and an area about 100% represents the best performance of the test. Reliability statistics were assessed using the test–retest method (Pearson's coefficient). Significant test results are quoted as two-tailed probabilities. Significance of the obtained results was judged at the 5% level.

Results

Pilot study

The pilot study was conducted on eight children who stuttered, after translation of the TOCS to the Arabic language and its modification to suit the Egyptian culture.

- (1) The modifications made to some of the pictures of the speech fluency task 1, Rapid Picture Naming, are as follows:
 - (a) In practice items, we substituted lawn mower with car.
 - (b) In the test items, we substituted the following:
 - (i) Crab with fish.
 - (ii) Rainbow with moon.
 - (iii) Stop sign with traffic lights.
 - (iv) Dinosaur with elephant.
 - (v) Lunch box with travel bag.
 - (vi) Snowflake with flower.
 - (vii) Ice cube with cube.
 - (viii) Kangaroo with lion.
 - (ix) Frog with cow.
- (2) The modifications made to some of the pictures of the speech fluency task 2, Modeled Sentences, are as follows:
 - (a) In practice items, we substituted 'The boy is riding the pig' with 'The boy is riding the donkey'.
 - (b) In the test items, we substituted 'The kids are taking a nap' with 'The kids are playing slide'.
- (3) The modification made to some of the pictures of the speech fluency task 3, Structured Conversation, is as follows: we substituted spaceship with airplane and modified the questions related to the picture according to the change.
- (4) And according to these changes, the pictures of speech fluency task 4, Narration, were changed.

Demographic data

Age

The controls and cases were classified into two age groups:

- (1) Age group A: from 4 years 0 months to 8 years 5 months.
- (2) Age group B: from 8 years 6 months to 12 years 11 months.

Statistically, there was an insignificant difference between the case and control groups (Table 1).

Sex

Nearly 80% of the studied population was male. No significant difference was found between cases and controls.

Test of Childhood Stuttering

- (1) The standardized speech fluency measure:

Table 2 shows the distribution of the studied groups with regard to the results of the standardized speech fluency measure of the TOCS [6,7].

Table 1 Age distribution in the two groups

Age in years	Cases (n = 30) [n (%)]	Controls (n = 100) [n (%)]	Test of significance
Age group A 4–8.5	20 (66.7)	67 (67.0)	¹ P = 0.973
Age group B >8.5 to <13.0	10 (33.3)	33 (33.0)	
Mean ± SD	7.34 ± 2.36	7.56 ± 2.51	¹ P = 0.666
Median	6.31	7.40	

P, P-value for the comparison between case and control groups;
t, Student's t-test.

Table 2 Distribution of the two studied groups according to the standardized speech fluency measure

Standardized speech fluency measure	Cases (n = 30)			Controls (n = 100)			Test of significance between case and control groups (P ^{MW})
	Group A (4–8.5 years) (n = 20)	Group B (>8.5 to <13 years) (n = 10)	Total (n = 30)	Group A (4–8.5 years) (n = 67)	Group B (>8.5 to <13 years) (n = 33)	Total (n = 100)	
Rapid picture naming							
Mean ± SD	5.25 ± 6.77	6.0 ± 9.33	5.50 ± 7.56	0.06 ± 0.30	0.0 ± 0.0	0.04 ± 0.24	0.0001*
P ^{MW}	0.638			0.219			
Modeled sentences							
Mean ± SD	10.35 ± 7.52	11.10 ± 5.65	10.60 ± 6.86	0.06 ± 0.30	0.03 ± 0.17	0.05 ± 0.26	0.0001*
P ^{MW}	0.550			0.722			
Structured conversation							
Mean ± SD	15.15 ± 11.57	15.70 ± 10.47	15.33 ± 11.03	0.21 ± 0.64	0.03 ± 0.17	0.15 ± 0.54	0.0001*
P ^{MW}	0.708			0.185			
Narration							
Mean ± SD	4.0 ± 2.58	4.80 ± 3.22	4.27 ± 2.78	0.18 ± 0.52	0.03 ± 0.17	0.13 ± 0.44	0.0001*
P ^{MW}	0.625			0.138			
Total score							
Mean ± SD	34.70 ± 26.07	37.60 ± 27.03	35.67 ± 25.96	0.52 ± 1.31	0.09 ± 0.52	0.38 ± 1.13	0.0001*
P ^{MW}	0.741			0.053			
Severity [n (%)]							
Typical fluency	0	0	0	67 (100.0)	33 (100.0)	100 (100.0)	0.0001*
Mild	9 (45.0)	3 (30.0)	12 (40.0)	0	0	0	
Moderate	6 (30.0)	5 (50.0)	11 (36.7)	0	0	0	
Severe	5 (25.0)	2 (20.0)	7 (23.3)	0	0	0	
P ^{MC}	0.622			–			

MC, Monte Carlo test; MW, Mann–Whitney test; *Statistically highly significant at P ≤ 0.0001.

Statistically, there was a highly significant difference between the case and control groups, whereas there was an insignificant difference between groups A and B in the case and control groups.

- (2) Observational rating scales:

Table 3 shows the distribution of the studied cases with regard to the results of the observational rating scales of the TOCS [7].

Statistically, there was a highly significant difference between the case and control groups, according to the Mann–Whitney test (P < 0.001), whereas there was an insignificant difference between groups A and B in both case and control groups.

- (3) Supplemental clinical assessment activities:

Table 4 shows an insignificant difference between groups A and B in the case and control groups in speech naturalness, average number of units per repetition, speech rate, disfluency duration, stuttering frequency, and associated behaviors.

Standardization study

Reliability

Error due to time sampling refers to the extent to which a Pearson's test performance is constant over time. It was estimated by the test–retest method. The time laps between testing varied from 6 to 12 days. Test–retest reliability was measured in the pilot study group of eight children. Table 5 shows that all items of the TOCS are highly significant.

Table 3 Distribution of the studied groups according to observational rating scales

Observational rating scales	Cases (n = 30)			Controls (n = 100)			Test of significance between case and control groups
	Group A (4–8.5 years) (n = 20)	Group B (>8.5 to <13 years) (n = 10)	Total (n = 30)	Group A (4–8.5 years) (n = 67)	Group B (>8.5 to <13 years) (n = 33)	Total (n = 100)	
The speech fluency rating scale							
I							
Mean ± SD	12.90 ± 5.58	15.70 ± 5.01	13.83 ± 5.48	0.39 ± 1.0	0.12 ± 0.70	0.30 ± 0.92	$P^{MW} = 0.0001^*$
P^{MW}		0.143			0.084		
II							
Mean ± SD	12.55 ± 5.40	12.50 ± 4.35	12.53 ± 5.0	0.25 ± 0.88	0.06 ± 0.35	0.19 ± 0.75	$P^{MW} = 0.0001^*$
P^{MW}		0.947			0.267		
III							
Mean ± SD	12.45 ± 5.60	14.30 ± 4.83	13.07 ± 5.34	0.25 ± 0.88	0.06 ± 0.35	0.14 ± 0.60	$P^{MW} = 0.001^*$
P^{MW}		0.364			0.359		
Severity [n (%)]							
Typical fluency	2 (10.0)	0	2 (6.7)	67 (100.0)	33 (100.0)	100 (100.0)	$P^{MW} = 0.0001^*$
Mild	12 (60.0)	5 (50.0)	17 (56.7)	0	0	0	
Moderate	6 (30.0)	5 (50.0)	11 (36.7)	0	0	0	
Severe	0	0	0	0	0	0	
P^{MC}		0.487			–		
The disfluency-related consequences rating scale							
I							
Mean ± SD	10.80 ± 6.97	13.90 ± 7.81	11.83 ± 7.28	0.3731 ± 1.027	0.0 ± 0.0	0.25 ± 0.86	$P^{MW} = 0.0001^*$
P^{MW}		0.332			0.280		
II							
Mean ± SD	11.0 ± 5.94	12.40 ± 5.93	11.47 ± 5.87	0.09 ± 0.42	0.0 ± 0.0	0.06 ± 0.34	$P^{MW} = 0.0001^*$
P^{MW}		0.596			0.219		
III							
Mean ± SD	10.50 ± 5.78	11.20 ± 5.75	10.73 ± 5.68	0.15 ± 0.50	0.06 ± 0.35	0.12 ± 0.46	$P^{MW} = 0.0001^*$
P^{MW}		0.706			0.285		
Consequences typical consequences [n (%)]							
Typical	4 (20.0)	2 (20.0)	6 (20.0)	67 (100.0)	33 (100.0)	100 (100.0)	$P^{\chi^2} = 0.0001^*$
Greater than typical	16 (80.0)	8 (80.0)	24 (80.0)	0	0	0	
P^{FE}		1.000			–		

FE, Fisher exact test; MC, Monte Carlo test; MW, Mann–Whitney test; *Statistically highly significant at $P \leq 0.0001$.

Table 4 Distribution of the two studied groups according to different studied parameters

Studied parameters	Cases (n = 30)			Controls (n = 100)			Test of significance between case and control groups
	Group A (4–8.5 years) (n = 20)	Group B (>8.5 to <13 years) (n = 10)	Total (n = 30)	Group A (4–8.5 years) (n = 67)	Group B (>8.5 to <13 years) (n = 33)	Total (n = 100)	
Speech naturalness							
Mean ± SD	4.0 ± 1.41	4.60 ± 1.58	4.20 ± 1.47	1.06 ± 0.24	1.03 ± 0.17	1.05 ± 0.22	0.0001*
P^{MW}		0.280			0.528		
Average number of units per repetition							
Mean ± SD	1.67 ± 0.43	1.71 ± 0.55	1.68 ± 0.46	0.18 ± 0.42	0.03 ± 0.19	0.13 ± 0.37	0.0001*
P^{MW}		0.596			0.056		
Speech rate (words per min or s)							
Mean ± SD	24.95 ± 9.95	24.20 ± 13.05	24.70 ± 10.85	28.19 ± 11.17	47.18 ± 15.65	34.46 ± 15.59	0.0001*
P^{MW}		0.613			0.0001*		
Disfluency duration (s)							
Mean ± SD	1.97 ± 0.83	2.31 ± 1.05	2.08 ± 0.91	0.23 ± 0.48	0.03 ± 0.19	0.16 ± 0.41	0.0001*
P^{MW}		0.402			0.027*		
Stuttering frequency (%)							
Mean ± SD	16.20 ± 12.36	17.70 ± 13.32	16.70 ± 12.48	0.25 ± 0.64	0.09 ± 0.52	0.20 ± 0.60	0.0001*
P^{MW}		0.757			0.063		
Associated behaviors [n (%)]							
No	6 (30.0)	2 (20.0)	8 (26.7)	67 (100.0)	33 (100.0)	100 (100.0)	0.0001*
Yes	14 (70.0)	8 (80.0)	22 (73.3)	0(0.0)	0(0.0)	0	
P^{FE}		0.682			–		

FE, Fisher Exact test; MW, Mann–Whitney test; P , P -value for comparing between groups A and B in each case and control group;

*Statistically highly significant at $P \leq 0.0001$.

Validity

(1) Content description validity:

Content description validity involves ‘systematic examination of the test content to determine

whether it covers a representative sample of the behavior domain to be measured’.

Selection of the test items was revised by two phoniatrists, who checked all the test items. All

relevant items were included, whereas all irrelevant ones were excluded.

(2) Internal consistency validity:

It is a measure of test homogeneity. It is measured by correlating each composite with the total test score using Pearson's correlation.

Table 6 shows correlation between each subscale and the total scale. The test proved to have good internal consistency, as there is significant positive correlation between each subtest score and the TOCS score.

(3) Convergent validity:

It examines the degree to which the test is similar to (converges on) other tests that it theoretically should be similar to. This was done by comparing TOCS with SSI-3:

Figure 1 shows that there was a significant positive correlation between the total TOCS score and SSI-3 in the stuttering group.

(4) Known group comparison:

Comparison of the mean and SD of the control and case groups showed that the score among cases was significantly higher than that among controls on all test items.

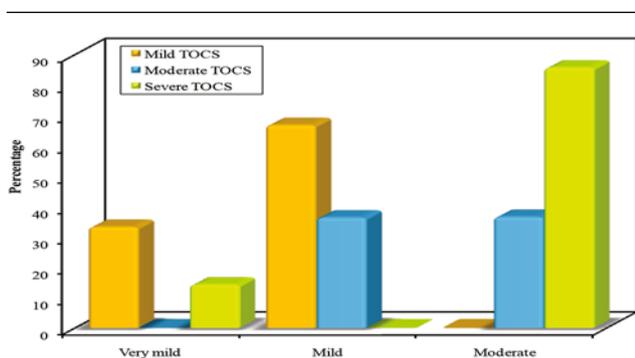
Table 7 shows the performance of the two groups on the standardized speech fluency measure of the TOCS in terms of mean and SD.

Table 8 shows the performance of the two groups on the observational rating scales of the TOCS in terms of mean and SD.

Sensitivity and specificity

The ROC curves with the AUCs and their statistical significance were used as an indicator for scale and subscale performances. Statistically significant AUCs denoted performances better than chance (AUC = 0.50), and the greater the AUC, the better

Figure 1



Correlation of the TOCS score with SSI-3 in the studied cases. SSI-3, Stuttering Severity Instrument-3; TOCS, Test of Childhood Stuttering.

was the performance, with a maximum AUC of 1.00 denoting a gold standard like performance. The cutoff values for diagnosis were identified just for statistically significant AUCs, where the value that maximized both sensitivity and specificity of the scale was chosen. Values above or the same as the identified cutoff values denote case values. The ROC curve was used to obtain cutoff values for the proposed test (Fig. 2).

Obtaining new cutoff values

Table 9 shows the cutoff scores of the total score of speech fluency measure of the TOCS. The test proved to have high sensitivity and specificity at the identified cutoff points ranging up to 7.

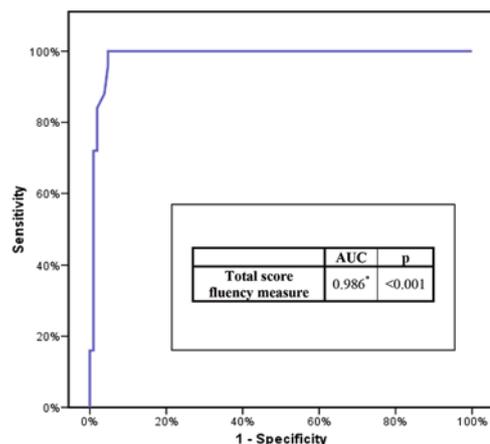
Table 5 The children values of the items that patients scored differently on retest

Different test items	<i>r</i>	<i>P</i>
The standardized speech fluency measure		
Rapid picture naming	0.984	0.0001***
Modeled sentences	0.965	0.0001***
Structured conversation	0.986	0.0001***
Narration	0.772	0.025*
Total scores	0.979	0.0001***
Severity	0.999	0.0001***
Observational rating scales		
The speech fluency rating scale		
I	0.765	0.027*
II	0.938	0.001**
III	0.751	0.032*
Severity	0.745	0.034*
The disfluency-related consequences rating scale		
I	0.883	0.004**
II	0.890	0.003**
III	0.898	0.002**
Typical consequences	0.999	0.001**

r, Pearson's coefficient; *Statistically significant at $P \leq 0.05$;

Statistically highly significant at $P \leq 0.001$; *Statistically very highly significant at $P \leq 0.0001$.

Figure 2



ROC curve for the TOCS total score vs. SSI-3. ROC, receiver operating characteristic; SSI-3, Stuttering Severity Instrument-3; TOCS, Test of Childhood Stuttering.

Discussion

About two-third of the patients in this sample were between 4 and 8 years, as the primary symptoms of developmental stuttering are first observed between

the ages of 2 and 4 years, following a period of relatively fluent speech production [6,9]. Data obtained from the University of Illinois Stuttering Research Program revealed that in 65% of the participating children, stuttering onset occurred before the age of 3 years; the figure rose to 85% by 3.5 years of age [9].

Table 6 Correlation between each subtotal with total scale

Subtotal and total scale	<i>r</i>	<i>P</i>
The standardized speech fluency measure		
Rapid picture naming	0.898	0.00001*
Modeled sentences	0.905	0.00001*
Structured conversation	0.956	0.00001*
Narration	0.856	0.00001*
Observational rating scales		
The speech fluency rating scale		
I	0.930	0.00001*
II	0.905	0.00001*
III	0.966	0.00001*
The disfluency-related consequences rating scale		
I	0.973	0.00001*
II	0.949	0.00001*
III	0.979	0.00001*

r, Pearson's coefficient; *Statistically highly significant at $P \leq 0.00001$.

In the present study, boys constituted 76.7% of the cases and 80% of the control group. There was an insignificant difference between the case and control groups and between the two age groups in both case and control groups. This corresponds to several studies reporting male predominance. A strong association is seen between sex and stuttering, especially in terms of a higher incidence and prevalence among male than among female individuals on the one hand [10] and a higher percentage of natural recovery on the other hand [6]. Previous research on various stuttering parameters, for example, disfluency characteristics,

Table 7 Comparison between the two studied groups according to the standardized speech fluency measure

Standardized speech fluency measure	Cases (<i>n</i> = 30)	Controls (<i>n</i> = 100)	<i>Z</i>	<i>P</i>
Rapid picture naming				
Mean ± SD	5.50 ± 7.56	0.04 ± 0.24	8.670	0.0001*
Modeled Sentences				
Mean ± SD	10.60 ± 6.86	0.05 ± 0.26	10.463	0.0001*
Structured Conversation				
Mean ± SD	15.33 ± 11.03	0.15 ± 0.54	10.313	0.0001*
Narration				
Mean ± SD	4.27 ± 2.78	0.13 ± 0.44	9.744	0.0001*
Total score				
Mean ± SD	35.67 ± 25.96	0.38 ± 1.13	9.980	0.0001*
Severity [<i>n</i> (%)]				
Typical fluency	0	100 (100.0)	11.245	0.0001*
Mild	12 (40.0)	0		
Moderate	11 (36.7)	0		
Severe	7 (23.3)	0		

Z, *Z* for Mann–Whitney test; *Statistically highly significant at $P \leq 0.0001$.

Table 8 Comparison between the two studied groups according to observational rating scales

Observational rating scales	Cases (<i>n</i> = 30)	Controls (<i>n</i> = 100)	Test of significance	<i>P</i>
The speech fluency rating scale				
I				
Mean ± SD	13.83 ± 5.48	0.30 ± 0.92	<i>Z</i> = 10.047	0.0001*
II				
Mean ± SD	12.53 ± 5.0	0.19 ± 0.75	<i>Z</i> = 10.395	0.0001*
III				
Mean ± SD	13.07 ± 5.34	0.14 ± 0.60	<i>Z</i> = 10.508	0.0001*
Severity [<i>n</i> (%)]				
Typical fluency	2 (6.7)	100 (100.0)	<i>Z</i> = 10.789	0.0001*
Mild	17 (56.7)	0		
Moderate	11 (36.7)	0		
Severe	0	0		
The disfluency-related consequences rating scale				
I				
Mean ± SD	11.83 ± 7.28	0.25 ± 0.86	<i>Z</i> = 9.905	0.0001*
II				
Mean ± SD	11.47 ± 5.87	0.06 ± 0.34	<i>Z</i> = 10.588	0.0001*
III				
Mean ± SD	10.73 ± 5.68	0.12 ± 0.46	<i>Z</i> = 10.395	0.0001*
Consequences [<i>n</i> (%)]				
Typical consequences	6 (20.0)	100 (100.0)	P^{c^2} = 98.113	0.0001*
Greater than typical consequences	24 (80.0)	0		

P, *P*-value for comparing between the two studied groups; *Z*, *Z* for; Mann–Whitney test; *Statistically highly significant at $P \leq 0.0001$.

Table 9 Agreement (sensitivity, specificity, and accuracy) for the TOCS total score

TOCS total score	SSI-3		Sensitivity	Specificity	Positive predictive value	Negative predictive value	Accuracy
	Typical fluency	Other					
TOCS total score							
≤7	100	0	100.0	95.24	83.33	100.0	96.15
>7	5	25					
≤9	100	1	96.0	95.24	82.76	99.01	95.38
>9	5	24					

TOCS, Test of Childhood Stuttering.

has reported few or no statistically significant sex differences [6].

The SSI-3 is the most commonly used measure of stuttering severity. The SSI-3 measures the frequency of symptoms, the duration of the three longest symptoms, and physical concomitant behavior and combines them into an overall score [8]. In the current study, there was a highly significant difference between the case and control groups, whereas there was an insignificant difference between groups A and B of both case and control groups. Riley [8] indicated that SSI-3 can be used as part of diagnostic evaluations, can assist in tracking changes in severity during and after treatments, and can be used to validate other assessment instruments. Lewis [11] did not support Riley's reliability procedures and found that two of the three parameters of SSI-3 failed to satisfy the fundamental requirement for measures of judge agreement. He concluded that SSI-3 scores do not accurately reflect observations of stuttering behaviors. The SSI-3 has the advantage of being the only available standardized measure of stuttering severity.

Reliability of the TOCS was proven to be high by the test-retest method. We investigated test-retest reliability using a group of eight children. These children had all been previously diagnosed with stutters. The time lapse between testing varied from 6 to 12 days. After the testing was completed, the raw scores for each test were correlated. The results indicated highly significant correlations between initial and retest scores, proven statistically by the correlation coefficient (r), which ranged from 0.745 to 0.999, hence proving the reliability of the test. This is in concordance with the results of the original test, in which r ranged from 0.86 to 0.91.

Validity of the TOCS was proven on the basis of content description validity, internal consistency validity, convergent validity, and known group comparisons.

Content description validity is the adequacy with which the test items adequately and representatively sample the content area to be measured. It was

examined by three phoniatricians to prove that the test measured what it was designed to measure. There was agreement on most test items and disagreement on some items with regard to the formulation of some of the questions; these were changed to be more clear and suitable.

Internal consistency validity, which is a measure of the homogeneity of the test itself, is measured by correcting each composite with the total test scores using Pearson correlation [12]. Significant positive correlation was found between each subtest score and the TOCS total score. The correlation coefficient (r) ranged from 0.856 to 0.979.

Convergent validity examines the degree to which a test is similar to tests that it should theoretically be similar to [12]. In this study, some of the test items were compared with SSI-3 [8] to look for correlations; these items have been selected because of the resemblance in the aim of testing between them and similar items in the TOCS. There is a highly significant positive correlation, proven statistically by the correlation coefficient (r) of 0.904, between both tests in the total severity scores.

Known group comparison is one way of establishing a test's validity by studying the performance of different groups of people on the test. Each group's results should make sense in relation to what is known about the relationship of the test's content with the group. In the case of TOCS, a measure of speech fluency, a rating scale for speech fluency, and a rating scale for disfluency and its related consequences, one would expect that individuals with identified stuttering would do worse than children with typical fluency. The means for the control group were much lower than those for cases, and all attained a highly significant probability.

In the current study, the sensitivity index reflects the ability of a test to correctly identify individuals who were previously identified as stutters. The specificity index reflects the ability of a test to correctly identify individuals who were not identified as stutters. The positive predictive value reflects the proportion of

individuals who were identified as stutters among all those identified by the measure as having stuttering-related problems.

The results of the TOCS were highly sensitive and specific. The percentage of sensitivity and specificity ranged from 95 to 100% for cutoff scores (≤ 7) of the total score of speech fluency measure of the TOCS, whereas in the original test of TOCS, the percentage of sensitivity and specificity ranged from 79 to 90% for cutoff scores (≤ 9) of the total score of speech fluency measure.

The ROC curve was plotted and the AUC was calculated for the TOCS total score, with an SSI-3 value of 0.986, indicating good discriminatory potential between typical fluency children and children with a stutter. The cutoff score of the total score of speech fluency measure of the TOCS was established at 7 or lower, showing a sensitivity of 100% and a specificity of 95.24%. The cutoff scores at which the sensitivity and specificity were maximum were 7 or lower. A raw score of 7 or higher by any rater is considered to be a high rating for a child to be categorized as a stutterer, and ratings of 6 or lower are considered to represent children with typical fluency.

Conclusion

- (1) TOCS offer clinicians straightforward and efficient fluency assessment techniques for children between 4 and 12 years of age.
- (2) The Arabic version of the TOCS is a valid and reliable test for evaluation of Arabic speaking

children with stutters and to determine the severity of a child's stuttering.

- (3) The results were highly significant and were capable of discriminating between children with normal disfluency and those who stutter.

Acknowledgements

Conflicts of interest

None declared.

References

- 1 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, text revision (DSM-IV-TR), 4th ed. Washington: American Psychiatric Association; 2000.
- 2 Guitar BE. Stuttering and stammering. *Pediatr Rev* 1985; 7:163–168.
- 3 Maguire G. *Without hesitation: speaking to the silence and the science of stuttering*. New York: National Stuttering Association; 2010.
- 4 Mansson H.. Childhood stuttering: incidence and development. *J Fluency Disord* 2000; 25:47–57.
- 5 Bloodstein O. *A handbook on stuttering*. 5th ed. San Diego: Singular Publishing Group; 1995.
- 6 Yairi E, Ambrose NG. Early childhood stuttering I: persistency and recovery rates. *J Speech Lang Hear Res* 1999; 42:1097–1112.
- 7 Ronald BG, Kenneth JL, Pearson NA. *Test of childhood stuttering (TOCS)*. Austin, Texas: Pro-Ed; 2009.
- 8 Riley G. *Stuttering severity instrument for children and adults*. 3rd ed. Austin, Texas: Pro-Ed; 1994.
- 9 Yairi E, Ambrose NG. *Early childhood stuttering*. Austin: Pro-Ed; 2005
- 10 Van Borsel J, Moeyaert E, Rosseel M, Van Loo E, Van Renterghem L. Prevalence of stuttering in regular and special school population in Belgium based on teacher perception. *Folia Phoniatr Logop* 2006; 58:289–302.
- 11 Lewis KE. Do SSI-3 scores adequately reflect observations of stuttering behaviors?. *Am J Speech Lang Pathol* 1995; 4:46–59.
- 12 Dunn W. Reliability and validity. In: Miller LJ, editor. *Developing norm-referenced standardized tests*. New York: Haworth Press; 1989.