Validity and Responsiveness of VELO: A Velopharyngeal Insufficiency Quality of Life Measure

Jonathan R. Skirko, MD, MHPA, MPH, **Edward M. Weaver, MD, MPH**, **Jonathan A. Perkins, DO**, **Sara Kinter, MA, CCC-SLP**, **Linda Eblen, MA, CCC-SLP**, and **Kathleen C. Y. Sie, MD**
Department of Otolaryngology-Head & Neck Surgery, University of Washington (JS, EW, JP and KS), Division of Pediatric Otolaryngology-Head & Neck Surgery, Seattle Children's Hospital (JP and KS), Childhood Communication Center, Seattle Children's Hospital (JP and KS) and Speech and Language Services, Seattle Children's Hospital (SK and LE), Seattle, WA

Abstract

Objective—Test the Velopharyngeal Insufficiency (VPI) Effects on Life Outcomes (VELO) instrument for validity, reliability and responsiveness.

Study Design—Observational cohort

Setting—Academic tertiary medical center

Subjects—Children with VPI (n=59) and their parents (n=84) were prospectively enrolled from a pediatric VPI clinic.

Methods—Pediatric speech language pathologists diagnosed VPI using perceptual speech analysis and rated VPI severity and speech intelligibility deficit (each as minimal, mild, moderate or severe). All parents and youth 8+ years old (n=24) completed the VELO instrument and other quality-of-life questionnaires at baseline; the first 40 subjects completed the VELO instrument again two-weeks later. Treatments included Furlow palatoplasty (n=20), sphincter pharyngoplasty (n=14) or an obturator (n=2), and 29/36 (81%) subjects completed the questionnaires three months post-treatment. VELO was tested with correlations for criterion validity against VPI severity, construct validity against speech intelligibility and velopharyngeal gap size, and concurrent validity against other quality-of-life measures (r>0.40 demonstrating validity); for test-retest reliability using intraclass correlation (>0.6 demonstrating reliability); and for responsiveness with the 3-month post-treatment measure using the paired t-test.

Results—Parental responses are reported; youth responses showed similar results. The VELO instrument did not meet criterion validity (r=−0.18, p=0.10), or functional construct validity (r= −0.37, p=0.001), but did meet anatomic construct and concurrent validity (each r>0.50, p<0.01).

VELO scores demonstrated excellent test-retest reliability (r=0.85, p<0.001) and responsiveness (baseline 54+/−14 to post-treatment 70+/−18, p<0.001).

Conclusion—VELO provides a VPI specific quality-of-life instrument that demonstrates concurrent validity, test-retest reliability, and responsiveness to change in quality-of-life with treatment.

Keywords

Velopharyngeal insufficiency; quality-of-life; validation; reliability; responsiveness

Introduction

Quality-of-life (QOL) refers to judgment of value placed on patient's health-related experiences. Condition-specific QOL instruments are tailored to measure how the condition affects QOL and are better able to detect change than generic instruments.[1] Velopharyngeal insufficiency (VPI) affects speech, swallowing and many psychosocial aspects of life in a way that is different from other conditions. Accurately measuring QOL in children with VPI is an area in need of further research. One condition-specific measure for pediatric VPI is the VPI Effects on Life Outcomes (VELO) instrument. This instrument was developed to capture the effects of VPI on children's lives. While the initial instrument analyses have been encouraging[2,3] further analysis on validity, reliability, and responsiveness are needed to better understand this new instrument's psychometric properties.

Validity of an instrument is the accuracy of the instrument to measure what it purports to measure, in this case the effects of VPI on QOL. Validity can be tested in a number of ways. A previous study of VELO demonstrated discriminant validity (ability to detect differences between patients with VPI and controls) and concurrent validity (correlation to a related validated QOL instrument).[2] Criterion validity is demonstrated when there is adequate correlation with the "gold-standard," to the extent that one exists. While no such gold-standard criterion exists for VPI, perceptual speech analysis is the most widely used measure in diagnosis.[4] Construct validity is demonstrated when hypothesized associations between the instrument and related health states test positive.[5] Criterion and construct validity have not been reported for the VELO instrument, and they will help characterize the accuracy of this instrument to measure the effects of VPI.

Reliability of an instrument is the degree to which repeated iterations yield the same result.[1] More specifically, test-retest reliability measures score stability over a time period in which respondents are assumed not to change. Test-retest reliability is particularly important when the instrument will be used for serial measurements, such as before and after treatment. While previous analysis of the VELO instrument has shown adequate internal consistency[2] (another measure of reliability), test-retest reliability has not been reported for this instrument.

Responsiveness is an instrument's ability to detect meaningful within-person change in health status.[1] When a disorder has a treatment of known efficacy, responsiveness is demonstrated by significant change in the instrument score after the treatment.[6]

The primary goals of this study were to test the VELO instrument for criterion and construct validity, test-retest reliability, and responsiveness in a prospective cohort of VPI patients. Secondary goals were to further test concurrent validity (with different self-reported measures) and internal consistency in this new cohort.

Methods

Study Subjects—This prospective cohort study enrolled new and established subjects with VPI at Seattle Children's Hospital VPI Clinic. English speaking children (ages 3–22 years) with VPI were enrolled from January 2010 to February 2012. Exclusion criteria included severe intellectual disability (n=3) or VPI surgery within 6-months prior to enrollment (n=2). The study was approved by the Institutional Review Board at Seattle Children's Hospital prior to enrollment.

Subjects completed questionnaires during their VPI clinic visit or by mailed questionnaire (n=6) within 6 months of their VPI clinic assessment. The first 40 subjects completed the same questionnaire by mail 2-weeks after the first questionnaire to assess test-retest reliability. Medical records were monitored to identify subjects' treatment dates. Follow-up questionnaires were obtained by mail 3-months after treatment to assess responsiveness.

Patient-Reported Outcomes—The VELO instrument is a VPI-specific quality-of-life measure. Its precursor VPI Quality-of-Life instrument is a 48-item questionnaire originally developed from focus groups (which provides face validity). The VPI Quality-of-Life instrument was modified to reduce burden and refine questions, resulting in the VELO instrument, which was then tested for internal consistency, discriminant validity, and 4 concurrent validity.[2] VELO includes a 26-item parent version (VELO Parent) and a 23-item youth version (completed by children 8 years and older, VELO Youth). Both can be found in Appendix A online. Respondents are prompted: "In the past four weeks, how much of a problem has your child had with:[]" The response format is a 5-point Likert-type scale ranging from never (0) to almost always (4). The total score ranges 0 – 100 with 100 representing the highest QOL. Subscales are scored similarly.

The Pediatric Voice Outcomes Survey (PVOS) is a 4-item, voice-specific functional status measure validated in a general pediatric otolaryngology population[7] and found to be responsive to change with VPI surgery.[8] The Pediatric Voice Related Quality of Life (PVRQOL) survey is a 10-item voice-specific instrument studied in a pediatric otolaryngology population and found to be reliable and valid.[9] Both Instruments are scored from 0 to 100 with higher score representing better QOL. QOL was also measured with 100mm visual analog scales (VAS) on speech, swallowing and situation & social interactions. Subjects rated how much of a problem each was on a 100mm VAS with anchors of "None" (0mm) and "Severe" (100mm). VAS is a validated and widely utilized measurement modality[10,11] and is well suited to measuring uni-dimensional outcomes. VAS-combined was the mean of the three scales. The questionnaires also included patient reports of related medical conditions, including a congenital syndrome, cleft palate (with or without cleft lip), or hearing loss.

VPI Management—Pediatric speech and language pathologists conducted perceptual speech analysis with a standardized inventory.[12] VPI severity was rated as none, minimal, mild, moderate, and severe, as was speech intelligibility deficit. Nasal endoscopy was performed by a pediatric otolaryngologist and velopharyngeal gap was rated on a 4 point ordinal scale (none, mild, moderate, and large). Nasal endoscopy measures have good reproducibility (r=0.86).[13] The management algorithm has been previously discussed[14] and is not the focus of this study. Treatment options included Furlow palatoplasty, sphincter pharyngoplasty or obturator. Obturators were considered "treatment" when the speech and language pathologist felt the VPI was adequately treated.

Data Analysis

Validation—The primary criterion and construct validation analyses used correlations to test the associations between VELO scores (total score and subscale scores) and other measures. Spearman correlation was used when the other measure was on an ordinal scale, and Pearson correlation was used when the other measure was on a continuous (or near-continuous) scale. *A priori*, we considered a correlation r >0.40 as substantial enough to support validity because this correlation equates to a coefficient of determination of 0.16, which means >15% of the variance of VELO scores are related to or explained by the other measure.

Criterion validity was tested primarily with the correlation between VELO total score (0–100) and VPI severity measured on perceptual speech analysis (none, minimal, mild, moderate or severe). The VELO-speech subscale was also correlated with VPI severity to provide subscale criterion validity. While the VPI severity on perceptual speech analysis is not truly a gold standard, it is the most widely used measure.

Construct validity was tested primarily with two hypothesized associations, namely the correlation between VELO total score and 1) speech intelligibility deficit (function construct), and 2) velopharyngeal gap (anatomic construct). We also hypothesized secondarily that the VELO total score would be lower in patients with a syndrome, cleft palate, or hearing loss, with each tested with the Student's t-test. Related VELO subscales were also tested for association with these constructs.

Concurrent validity was also tested as a repeat analysis from the original VELO study,[2] but in this new cohort and with different self-reported health status and QOL measures. VELO total score and related subscales were tested for positive correlation with PVOS, PVRQOL and VAS-combined (QOL constructs).

Reliability—Test-retest reliability between baseline and 2-week scores was tested with the intraclass correlation (ICC) in the first 40 patients. A correlation greater than 0.6 (substantial agreement) was deemed adequate.[15] Chronbach's alpha was calculated using baseline data on all subjects to assess internal consistency of VELO total score as well as subscales as a repeat of the analysis performed previously in the original VELO study.[2] A Chronbach's alpha > 0.70 is considered acceptable.[16]

Responsiveness and Analysis of Effects—In subjects who received treatment, responsiveness of VELO and other QOL instruments were tested with the two-sided paired t-test for change from baseline to 3-month post-treatment score. Significance level (alpha) was set at 0.05 for all analyses. Effect size (Cohen's "d") was used to compare instrument performance and was calculated by dividing the change in total score after treatment by baseline standard deviation for each instrument and subscale.[17] Effect size of 0.2 was considered small but clinically important, 0.5 was considered medium and 0.8 was considered large.[18]

Results

We enrolled 84 subjects with mean age of 7.1 years, ranging from 3 to 20 years. The age distribution was skewed to young subjects with n=61 (73%) less than 8 years. The racial distribution is largely Caucasian and cleft palate with or without cleft lip comprises 60% (n=52) of the study subjects (Table 1).

Criterion Validity (Table 2)

VELO Parent total score trended to association with VPI severity (r=−0.18, p=0.10) among all age subjects, but the association was below the −0.40 validation threshold. Among subjects <8 years, VELO Parent total was associated with VPI severity (r=−0.26, p=0.05, Table 2), but also below −0.40. Among subjects 8+ years, VELO Parent total and VELO Youth total both were not associated with VPI severity (p=0.88 and p=0.89, respectively). There were limited subjects age 8+ with moderate (n=3) and severe (n=3) VPI. Box plot of VELO Parent total versus VPI severity shows trend to lower QOL and less variability in VELO Parent total among those with worse VPI (Figure 1A, p=0.09).

Construct Validity (Table 2)

VELO Parent total score was associated with speech intelligibility deficit (r=−0.37, p=0.001, Figure 1B) among all age subjects, but below the −0.40 validation threshold. Among subjects <8 years old, VELO Parent total was associated with speech intelligibility (r=−0.48, p=<0.001, Table 2). Among subjects 8+ years, VELO Parent total and VELO Youth total both were not associated with speech intelligibility (p=0.38 and p=0.12, respectively).

VELO Parent total score was associated with velopharyngeal gap (r=−0.57, p<0.01, Figure 1C) among all age subjects. Among subjects 8+ years, VELO Parent total and VELO Youth total both were associated with velopharyngeal gap (r=−0.48, p=0.03 and r=−0.49, p=0.05, respectively). Secondary subscale results are reported in Table 2. Secondary tests of medical construct validity are summarized in Table 3.

Concurrent Validity (Table 2)

VELO Parent and VELO Youth total scores were associated with PVOS, PVRQOL and VAS-combined (Table 2.[2] Most of the VELO subscales were associated with the self-reported measures (Table 2).

Reliability (Table 4)

VELO Parent total showed excellent test-retest reliability (ICC=0.85, p<0.001). VELO Parent subscales and VELO Youth total and subscales had similar reproducibility (Table 4). PVOS had adequate test-retest reliability (ICC of 0.62, p<0.001) and the PVRQOL instrument had good test-retest reliability (ICC=0.72, p<0.001). Reliability was similar for subjects over and under 8 years old (data not shown). All instruments demonstrated acceptable internal consistency (Table 4.[2]

Responsiveness and Analysis of Effects (Table 5)

Follow up questionnaires were obtained in 29/36 (81%) subjects receiving VPI treatment. Treatment included Furlow palatoplasty (n=20), sphincter pharyngoplasty (n=14), or obturator (n=2). VELO Parent total mean (SD) score improved with treatment from 54(14) at baseline to 70(18) post-treatment (p<0.001). The effect size was 1.1 (large effect) for the VELO Parent total. Most VELO Parent subscales also showed improvement (Table 5). VELO Youth total mean (SD) score improved with treatment from 60(24) at baseline to 81(16) post-treatment (p=0.02). This resulted in an effect size of 0.9 (large effect) for the VELO Youth total. The study was underpowered to identify a change in VELO Youth subscales (Table 5). The other self-reported outcome measures showed significant improvements with medium effect sizes (Table 5).

Discussion

Most previous VPI studies have utilized postoperative perceptual speech analysis or velopharyngeal closure as their primary surgical outcomes. There is a paucity of studies utilizing patient-reported outcomes of validated condition-specific functional status or QOL. This study provides validation, reliability testing and responsiveness testing of the VELO, a condition specific quality-of-life instrument that can be used in future studies of patients with VPI.

While our criterion validity correlation did not reach the −0.40 validation threshold, there is an association among subjects under age 8 years old. Box plot of VELO Parent total score by VPI severity shows that subjects of any severity can have a low VELO (QOL), though subjects with more severe VPI are less likely to have a high VELO score. There was no association among subjects 8+ years old. This finding may be due to very small samples of subjects with moderate (n=3) and severe (n=3) VPI and high incidence of prior VPI surgery (46% among those 8+ years, 11% among those <8). Subjects who had prior treatment may have a different QOL for a given VPI severity than untreated subjects. While VPI severity is an important aspect of assessment, speech intelligibility was hypothesized to be a better overall assessment of speech dysfunction. The association between speech intelligibility and VELO Parent total score was present in younger subjects but not among older subjects. Given the overall functioning of the VELO measured here, the weak correlation with VPI severity highlights that VPI severity is not an adequate proxy for QOL, so QOL should be measured directly in these patients.

Anatomic construct validation tested the correlation between VELO score and velopharyngeal gap on endoscopy. VELO score was associated with nasendoscopy findings among older subjects, but not among younger subjects. Examination of the box plot show similar association between VELO score and velopharyngeal gap as discussed above with VPI severity. Subjects with any size of velopharyngeal gap, including those with a small gap, can have low QOL (VELO). While both nasal endoscopy and perceptual speech analysis are essential in assessing patients with VPI, perceptual speech analysis results have better correlation with QOL in young children, while nasendoscopy may be better in older children. This could be due to difficulty with the nasal endoscopy exam in young children or that nasal endoscopy is better at identifying residual VPI after treatment. While presence of a gap may be identified, differences in the young children's effort may also explain some of the difference in associations. Analysis in an independent sample and age specific reliability of these measurements would also help understand the association seen here.

Perceptual speech analysis and nasendoscopy measure speech and anatomy, not quality-of-life. To understand these subjects' overall QOL and to test concurrent validity, QOL was measured in a variety of ways. Both VELO Parent and VELO Youth correlated well with these measures. VELO was previously shown to correlate with a generic pediatric QOL, the PedQL (r=0.73), providing evidence of its concurrent validity.[2] In the current study, the VELO also correlated well with two previously developed instruments (PVOS and PVRQOL) as well as the VAS-combined. The correlation of these QOL measures was similar across age groups, further highlighting the potential differences in speech and nasendoscopy assessment by age.

While the VELO total score attempts to measure overall QOL, monitoring subscale scores provides insight into the nuances of how VPI affects these children and how treatments are able to improve their QOL (the long-term goal of developing the VELO). Subscale concurrent validity also tested VELO subscales to a variety of specific measures. The correlation between VELO subscale and respective measures largely followed the hypotheses. While VASs may not be an ideal overall measure, they provide adequate measures for subscale validation purposes. The VELO situational difficulty and emotional impact did not reach the validation threshold correlation of 0.4 among younger children, but the subscales did show some level of association. The level of association could be due to limitations of the one-item VAS's ability to measure the complex construct of the psychosocial aspects of QOL. This could also be due to limitation of proxy assessment of psychosocial aspects of QOL in young children.

The study initially enrolled subjects with newly diagnosed VPI but age was skewed to young subjects. To ensure the VELO would be validated in older children, the protocol was changed to include subjects with VPI after prior treatment. The study ended with an adequate sample of children from age 8–20 for validation. Differences between the associations here highlight the need to include children in the assessment of QOL and not rely solely on parent proxy reporting.

The VELO total score, as well as each of the subscales were shown to have excellent test re-test reliability and internal consistency. This assessment is particularly important as the VELO is intended to measure change in QOL with treatment.

Treatment (surgical and obturator) improved VELO score, showing that the VELO Parent instrument is responsive to change in QOL. The subscales similarly showed responsiveness, except for VELO perception by others subscale. The effect size was large (>0.8) for the total VELO and several subscales. While the sample size was small (n=5), VELO Youth also improved with treatment. VELO outperformed the other QOL instruments on responsiveness and effect size, suggesting VELO may be better suited for detecting small changes in VPI-specific QOL. The smaller effect size on the perception by others and emotional impact subscales (both parent and youth report) may reflect slow improvement in psychosocial indicators. Additional studies following patients with VPI after treatment will help to determine if improvement continues to occur and when it stabilizes.

Minimal important change analysis seeks to determine the smallest change that is important to patients[16] and helps to provide context for measuring change. We utilized an anchor-based method to define a group of subjects with minimal change (n=2). The small sample size limited the interpretation of this analysis (data not shown). Future studies measuring VELO after a treatment with smaller effect may help provide a larger sample of subjects with minimal important change.

Accurately measuring condition-specific QOL is important for understanding current treatment and management of VPI. The VELO instrument provides a rigorously tested instrument to measure patient-centered outcomes in children with VPI. This work provides a foundation for future investigations of VPI treatment with a focus on a patient-centered measure.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Patrick DL, Deyo RA. Generic and disease-specific measures in assessing health status and quality of life. Med Care. Mar; 1989 27(3 Suppl):S217–232. [PubMed: 2646490]

2. Skirko JR, Weaver EM, Perkins J, et al. Modification and Evaluation of a Velopharyngeal Insufficiency Quality-of-Life Instrument. Arch Otolaryngol Head Neck Surg. Oct 1; 2012 138(10): 929–935. [PubMed: 23069823]

3. Barr L, Thibeault SL, Muntz H, de Serres L. Quality of life in children with velopharyngeal insufficiency. Arch Otolaryngol Head Neck Surg. Mar; 2007 133(3):224–229. [PubMed: 17372078]

4. Lam DJ, Starr JR, Perkins JA, et al. A comparison of nasendoscopy and multiview videofluoroscopy in assessing velopharyngeal insufficiency. Otolaryngol Head Neck Surg. Mar; 2006 134(3):394–402. [PubMed: 16500434]

5. Health outcomes methodology. Med Care. Sep; 2000 38(9 Suppl):II7–13. [PubMed: 10982086]

6. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Med Care. Sep; 2000 38(9 Suppl):II84–90. [PubMed: 10982093]

7. Hartnick CJ. Validation of a pediatric voice quality-of-life instrument: the pediatric voice outcome survey. Arch Otolaryngol Head Neck Surg. Aug; 2002 128(8):919–922. [PubMed: 12162771]

8. Boseley ME, Hartnick CJ. Assessing the outcome of surgery to correct velopharyngeal insufficiency with the pediatric voice outcomes survey. Int J Pediatr Otorhinolaryngol. Nov; 2004 68(11):1429–1433. [PubMed: 15488976]

9. Boseley ME, Cunningham MJ, Volk MS, Hartnick CJ. Validation of the Pediatric Voice-Related Quality-of-Life survey. Arch Otolaryngol Head Neck Surg. Jul; 2006 132(7):717–720. [PubMed: 16847178]

10. Huskisson EC. Measurement of pain. Lancet. Nov 9; 1974 2(7889):1127–1131. [PubMed: 4139420]

11. Scott J, Huskisson EC. Graphic representation of pain. Pain. Jun; 1976 2(2):175–184. [PubMed: 1026900]

12. Sie KC. Cleft palate speech and velopharyngeal insufficiency: surgical approach. B-ENT. 2006; 2(Suppl 4):85–94. [PubMed: 17366852]

13. Yoon PJ, Starr JR, Perkins JA, Bloom D, Sie KC. Interrater and intrarater reliability in the evaluation of velopharyngeal insufficiency within a single institution. Arch Otolaryngol Head Neck Surg. Sep; 2006 132(9):947–951. [PubMed: 16982971]

14. Sie KC, Chen EY. Management of velopharyngeal insufficiency: development of a protocol and modifications of sphincter pharyngoplasty. Facial Plast Surg. May; 2007 23(2):128–139. [PubMed: 17516340]

15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. Mar; 1977 33(1):159–174. [PubMed: 843571]

16. Terwee CB, Bot SD, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. Jan; 2007 60(1):34–42. [PubMed: 17161752]

17. Cohen J. A power primer. Psychol Bull. Jul; 1992 112(1):155–159. [PubMed: 19565683]

18. Kazis LE, Anderson JJ, Meenan RF. Effect sizes for interpreting changes in health status. Med Care. Mar; 1989 27(3 Suppl):S178–189. [PubMed: 2646488]
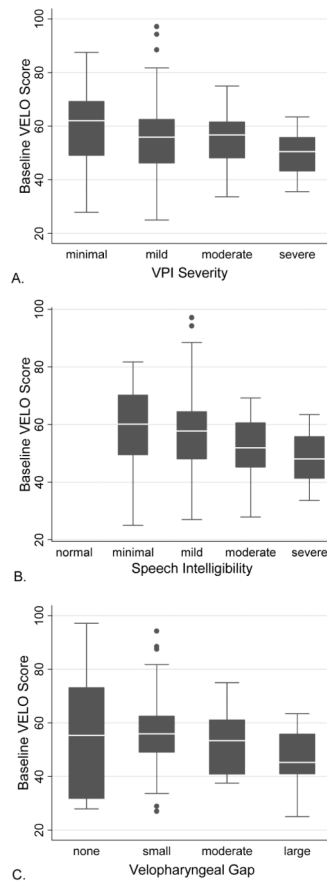
Figure 1.
Box plot graphs of baseline VELO parent score by A) VPI severity, B) speech intelligibility and C) velopharyngeal gap for subjects of all ages. Gray box shows 25th to 75th percentile with white line marking median. Black lines (whisker) shows 95% confidence interval. Asterisks mark data points outside the 95% CI.

Table 1

Characteristics of VPI subjects

Parameter	VPI Subjects n=84
Child's Age; years	7.1 +/− 4.3
Child's Sex; n (%) Female	45 (52%)
Hispanic	9 (11%)
Race	
Caucasian	60 (71%)
Asian	15 (18%)
American Indian	3 (4%)
African American	2 (2%)
Other	4 (5%)
Medical Comorbidities	
Cleft Lip & Palate	30 (35%)
Cleft Palate Alone	22 (26%)
No Cleft	33 (39%)
Child with Syndrome	30 (36%)
Hearing Loss	21 (25%)
VPI Severity	
None	0 (0%) [*]
Minimal	11 (13%)
Mild	31 (37%)
Moderate	27 (33%)
Severe	14 (17%)
Speech Intelligibility Deficit	
None	1 (1%)
Minimal	16 (18%)
Mild	21 (24%)
Moderate	26 (30%)
Severe	18 (20%)
Velopharyngeal Gap	
None	9 (12%)
Mild	40 (54%)
Moderate	16 (22%)
Large	9 (12%)
Quality of Life	
VELO Parent Total, n=83	56 +/− 16
VELO Youth Total, n=24	65 +/− 120
PVOS n=85	64 +/− 21
PVRQOL n=84	73 +/− 17
VAS Speech	36 +/− 30
VAS Swallow	14 +/− 22

Parameter	VPI Subjects n=84
VAS Situational	26 +/− 29
Bothered most by:	
Speech problem	52 (66%)
Swallowing problem	3 (4%)
Social Interactions	20 (25%)
Other	4 (5%)

Data presented as mean +/− SD and n (%)

*
Inclusion in the cohort required presence of at least minimal VPI.

Table 2

VELO Validation – Criterion, Construct and Concurrent Validity

VELO Measure:	Correlated with:	Parent age 3–7 n=61		Parent age 8+ n=23		Youth age 8+ n=23	
		r	p-value	r	p-value	r	p-value
Criterion Validity:							
VELO Total	VPI Severity	−0.26[a]	0.05	−0.09[a]	0.71	−0.16[a]	0.49
VELO-Speech	VPI Severity	−0.16[a]	0.21	−0.10[a]	0.66	−0.06[a]	0.79
Construct Validity:							
VELO Total	Intelligibility	**−0.48[a]**	**<0.001**	−0.21[a]	0.38	−0.34[a]	0.12
VELO-Speech	Intelligibility	−0.23[a]	0.07	−0.04[a]	0.87	−0.21[a]	0.35
VELO Total	VP Gap	−0.13[a]	0.32	**−0.49[a]**	**0.03**	**−0.48[a]**	**0.05**
Concurrent Validity:							
VELO Total	PVOS	**0.63**	**<0.001**	**0.60**	**<0.01**	**0.47**	**<0.001**
VELO Total	PVRQOL	**0.54**	**<0.001**	**0.69**	**<0.001**	**0.55**	**0.004**
VELO Total	VAS-combined	**−0.60**	**<0.001**	**−0.71**	**<0.001**	**−0.83**	**<0.001**
VELO-Speech	PVOS	**0.53**	**<0.001**	**0.60**	**<0.01**	0.39	0.07
VELO-Speech	PVRQOL	**0.51**	**<0.001**	**0.51**	**0.01**	**0.42**	**0.05**
VELO-Speech	Speech VAS	−0.35	0.007	**−0.56**	**0.01**	**−0.67**	**<0.001**
VELO-Swallow	Swallow VAS	−0.28	0.04	**−0.91**	**<0.001**	**−0.82**	**<0.001**
VELO-Situational	Situational VAS	−0.25	0.06	−0.42	0.06	**−0.59**	**0.003**
VELO-Emotional	Situational VAS	−0.29	0.03	**−0.53**	**0.01**	**−0.65**	**<0.001**
VELO-Perception	Situational VAS	**−0.57**	**<0.001**	**−0.54**	**0.01**	**−0.73**	**<0.001**

[a] Spearman correlation results shown (otherwise Pearson's correlation result shown), correlations greater than 0.40 shown in bold. VP, Velopharyngeal; VAS, Visual analog scale.

Table 3

VELO Validation – Medical Construct Validity

VELO Measure	Associated with	Parent Age 3–7			Parent Age 8+			Youth Age 8+		
		+	–	p-value	+	–	p-value	+	–	p-value
		Mean (SD)	Mean (SD)		Mean (SD)	Mean (SD)		Mean (SD)	Mean (SD)	
VELO Total	Syndrome	56(13)	55(14)	0.87	48(23)	63(23)	0.14	54(21)	72(16)	0.04
	N	19	41		8	11		10	11	
VELO Total	Cleft Palate	56(13)	55(14)	0.76	62(21)	46(15)	0.10	68(19)	57(21)	0.27
	N	26	34		14	7		16	7	
VELO Total	Hearing Loss	55(14)	55(14)	0.94	57(18)	56(22)	0.89	59(12)	66(23)	0.41
	N	13	46		6	14		8	14	

+ = with medical condition, – = without medical condition.

Table 4

Test-Retest Reliability and Internal Consistency of VELO, PVOS and PVRQOL instruments

	ICC	95% CI	Cronbach's α
VELO Parent Total	0.85	(0.77 – 0.93)	0.92
Speech	0.80	(0.70 – 0.91)	0.80
Swallow	0.87	(0.80 – 0.94)	0.85
Situational Difficulty	0.80	(0.68 – 0.90)	0.90
Perception	0.73	(0.59 – 0.87)	0.75
Emotional	0.76	(0.63 – 0.89)	0.78
Care Giver Impact	0.81	(0.70 – 0.91)	0.68
VELO Youth Total	0.83	(0.68 – 0.99)	0.93
Speech	0.63	(0.32 – 0.95)	0.80
Swallow	0.90	(0.81 – 0.99)	0.89
Situational Difficulty	0.79	(0.60 – 0.98)	0.92
Perception	0.71	(0.46 – 0.97)	0.83
Emotional	0.72	(0.47 – 0.97)	0.78
PVOS	0.62	(0.43 – 0.80)	0.76
PVRQOL	0.72	(0.58–0.86)	0.73

ICC, Intraclass Correlation, used to test test-retest reliability. ICC > 0.6 is considered substantial agreement. Cronbach's α used to test internal consistency.

Table 5

Change in QOL 3 months after VPI treatment

	Baseline mean (95% CI)	Three Month mean(95% CI)	p-value[a]	Effect Size[b]
VELO Parent Total (n=29)	54 (48 – 60)	70.2 (63 – 76)	<0.001	1.1
speech limitation	41 (34 – 49)	65 (56 – 73)	<0.001	1.3
swallowing	87 (79 – 93)	95 (90 – 99)	<0.005	0.4
situational difficulty	32 (30 –39)	53 (43 – 62)	<0.001	1.1
emotional impact	64 (55 – 72)	75 (67 – 83)	<0.005	0.5
perception by others	72 (64 – 80)	78 (69 – 86)	0.10	0.3
caregiver impact	52 (45 – 59)	70 (63 – 77)	<0.001	1.0
VELO Youth Total (n=5)	60 (31 – 89)	81 (61 – 101)	0.02	0.9
speech limitation	41 (12 – 70)	74 (56 – 92)	0.02	1.4
swallowing	95 (86 – 104)	100 (100 – 100)	0.21	0.7
situational difficulty	51 (19 – 83)	81 (55 – 108)	0.05	1.2
emotional impact	65 (13 – 118)	79 (32 – 126)	0.12	0.3
perception by others	75 (30 – 120)	81 (64 – 99)	0.62	0.2
PVOS (n=29)	63 (57 – 72)	75 (70 – 81)	<0.001	0.6
PVRQOL (n=29)	75 (70 – 82)	83 (78 – 89)	0.002	0.5

[a]paired t-test.

[b]Effect size (Cohen's "d") = change / baseline SD