

# The Cleft Audit Protocol for Speech-Augmented-Korean Modification: Reliability Test

Seunghee Ha

*Division of Speech Pathology and Audiology, Audiology and Speech Pathology Research Institute, Hallym University, Chuncheon, Korea*

**Correspondence:** Seunghee Ha, PhD  
Division of Speech Pathology and Audiology,  
Audiology and Speech Pathology Research  
Institute, Hallym University, 1 Hallymdaehak-gil,  
Chuncheon 24252, Korea  
Tel: +82-33-248-2215  
Fax: +82-33-256-3420  
E-mail: shha@hallym.ac.kr

Received: July 6, 2018  
Revised: August 17, 2018  
Accepted: August 17, 2018

**Objectives:** There is no formal assessment protocol specialized for speech disorders related to cleft palate in Korea. This study aimed to develop an assessment tool for cleft palate speech of Korean-speaking children and to examine the inter-rater and intra-rater reliability. **Methods:** The preliminary version of the assessment tool was developed based on the Cleft Audit Protocol for Speech-Augmented. A modified version of the assessment tool was made through examination of validity by experts of cleft palate speech. Thirty-three undergraduate students majoring in speech-language pathology participated as raters in the reliability test on the use of the modified version. They assessed speech samples from 8 children with cleft palate on two separate occasions. Intraclass correlation coefficients were used for inter-rater and intra-rater reliability tests and calculated by 12 sections of the assessment tool. **Results:** Inter-rater reliability was rated as moderate for two sections, good for five sections, and excellent for three sections. Developmental errors and compensatory nasal articulation errors were rated as poor interrater reliability. Intra-rater reliability was rated as moderate for four sections. Compensatory co-articulation and audible nasal emission were rated as poor intra-rater reliability, whereas speech intelligibility and hypernasality were rated as good intra-rater reliability. **Conclusion:** The rationale for the sections which showed low inter-rater and intra-rater reliability were discussed. The results have implications in the direction and caution of listener training for the use of the modified version of the assessment tool. The study suggested that the modified version can be usefully and reliably used for assessing cleft palate speech following appropriate training.

**Keywords:** Cleft palate speech, Korean modification, Reliability

구개열 아동은 구개성형술 이후 말 발달 현황을 정기적으로 평가하면서 지연이나 문제가 확인될 경우 언어치료나 2차 구개성형술을 효과적으로 받으면 정상적인 말 산출 능력을 갖출 수 있다. 구개열 아동이 학령기 이전에 정상적인 말 산출 능력을 갖추기 위해서는 무엇보다 수술 전후의 구조적, 기능적 변화를 객관적으로 평가하고, 정상적인 말·언어발달을 위해 필요한 부가적인 의료 서비스를 결정하여 제공하는 것이 중요하다. 구개열 아동을 위한 임상적 의사결정은 연인두 기제를 직접적으로 관찰하여 남아있는 해부학적 결함을 확인하는 평가와 함께 아동의 말을 언어치료사의 숙련된 귀로 듣고 판단하는 청지각적 평가 결과를 토대로 이루어져야 한다(Kuehn & Moller, 2000). 구개열 아동이 빈번하게 어려움을 보이는 조음, 공명, 음성 영역에서의 청지각적 평가를 정확하고 일관

성 있게 하기 위해서는 구개열 아동의 말 특징에 대한 이해와 임상 경험이 필수적이다. 또한 구개열로 인한 말 문제를 체계적으로 평가하기 위해서는 타당도와 신뢰도를 갖춘 전문적인 검사도구가 필요하다. 하지만 현재 국내 임상 및 연구현장에서는 구개열로 인한 말 문제를 청지각적으로 평가할 때 사용할 수 있는 전문적인 검사도구가 없어, 일반적인 조음음운장에 검사도구를 이용하거나 비공식적인 검사를 주로 시행하고 있다.

영어권 국가에서 현재 가장 보편적으로 사용되고 있는 구개열 말 평가도구는 영국에서 2006년에 개발된 Cleft Audit Protocol for Speech-Augmented (CAPS-A)이다. CAPS-A는 영국에서 전국적으로 표준화된 기준으로 구개열 의료서비스 결과를 평가하기 위해 개발된 검사도구이다. 원래는 Sell, Harding과 Grunwell (1994, 1999)

이 개발하여 사용하던 The Great Ormond Street Speech Assessment을 수정하여 Cleft Audit Protocol for Speech로 먼저 제안하여 사용한 뒤에 타당도와 신뢰도 검증을 통해 다시 수정되었다(Britton et al., 2014; John, Sell, Sweeney, Harding-Bell, & Williams, 2006; Sell et al., 2009). 최근에는 미국 구개열 전문 언어병리학자들이 CAPS-A를 수정 도입하여 CAPS-A Americleft modification (Chapman et al., 2016)의 이름으로 사용하고 있다.

CAPS-A에 대해 자세히 살펴보면 일반적인 조음음운능력 검사 도구처럼 표준화된 검사어에 의존하기 보다는 자발화를 비롯한 다양한 발화 샘플을 수집해서 사용하기를 권장하고 있다. 일반적으로 숫자세기와 같은 자동구어, 특정 구강자음이 집중적으로 들어간 낱말이나 문장(예: 양순파열음 문장), 자발화 등의 오디오/비디오 자료를 수집하여 CAPS-A에 포함된 평가 항목을 청지각적으로 판단하게 한다. CAPS-A의 평가 항목은 크게 (1) 말명료도, (2) 음성, (3) 공명, (4) 비누출, (5) 찡그림, (6) 자음 산출표, (7) 구개열 관련 조음 특성, (8) 언어치료 필요성 여부로 구성되어 있다. 이 중 공명 영역은 과다비성과 과소비성으로 구성되어 각각 5점, 3점의 등간척도를 이용하여 평가한다. 비누출은 들리는 비누출(audible nasal emission)과 비강 난기류(nasal turbulence)로 세부 항목을 두고 각각 3점 척도로 평가한다. 또한 평가 대상자가 말을 하는 동안 얼굴이나 코 찡그림을 보일 경우 따로 표기하도록 항목을 독립적으로 구분하였다. 조음 영역에서는 먼저 자음 산출표를 제시하여 목표 자음이 실현된 형태를 초성, 종성별로 기재하도록 하였다. 구개열 관련 조음 특성은 크게 발달적 오류, 보상적 오류, 필연적 오류로 유형을 분류한 뒤에, 각 오류 유형별로 세부 오류 형태를 제시하였다. 특히 보상적 오류의 경우에는 (1) 앞쪽 구강 기제를 사용하지만 조음위치나 방법면에서 오류를 보이는 패턴과, (2) 구강 뒤쪽으로 후방화된 패턴, (3) 성문음화나 인두음화와 같이 구강 기제를 사용하지 않는 비구강(non-oral) 보상조음 패턴으로 크게 분류하여 평가 대상자가 말 소리를 산출할 때 습관적으로 사용하는 조음기관의 위치를 체계적으로 살펴볼 수 있다. 조음영역은 세부 항목별로 3점 척도(0 = 없음; 1 = 2개 이하의 자음에서 관찰; 2 = 3개 이상의 자음에서 관찰)를 사용해서 특정 오조음 패턴의 유무뿐만 아니라 심각도를 대략 측정할 수 있게 하였다. 마지막으로 CAPS-A에서는 말명료도 및 조음, 공명, 음성 영역에서의 세부적인 평가뿐만 아니라 청지각적 말 평가결과를 종합하여 언어치료의 필요성을 판단하도록 하였다. CAPS-A는 여러 명의 구개열 전문 언어병리학자들의 지속적인 논의와 연구를 통해 수정 보완되어 최종적으로 7개의 평가 항목별로 70%에서 100% (평균 87%)의 준거 타당도가 보고되었다(John et al., 2006). 또한 검사자 간, 검사자 내 신뢰도는 Kappa값이 .41에서

1.00으로 양호한 신뢰도 결과가 나와, 적절한 훈련을 통해서 비교적 쉽게 사용할 수 있음이 입증되었다.

청지각적 평가에서 임상경험과 훈련은 신뢰도에 영향을 미치는 주요 변인 중 하나이다(Chan & Yiu, 2006; Eadie & Baylor, 2006; Ellis & Beltyukova, 2008; Lee, Whitehill, & Ciocca, 2009; Sell, 2005; Lewis, Watterson, & Houghton, 2003). 일반적으로 과다비성을 비롯해서 구개열과 관련된 말 문제의 청지각적 평가는 전문화된 훈련과 임상경험을 많이 가질수록 정확하고 일관성 있는 평가를 실시할 수 있다(Lee et al., 2009; Lewis et al., 2003). 현재 국내 임상현장의 현실을 고려해보면 언어병리학 전공 대학원은 구개열 대상자를 평가하고 치료하는 직무를 수행할 수 있는 역량을 갖추어야 한다. 따라서 구개열 전문 언어치료사뿐만 아니라 언어병리학 전공 대학생들도 구개열 말 특징과 관련된 교육과 훈련을 받은 후에 본 연구에서 개발된 검사도구를 이용하여 정확하고 일관성 있는 평가를 수행할 수 있는지를 살펴보는 것은 중요하다.

본 연구에서는 CAPS-A를 토대로 개발된 “한국판 구개열 말 평가(Cleft Audit Protocol for Speech -Augmented-Korean Modification)” 도구를 소개하고자 한다. 또한 구개열 언어재활 교과목을 수강하고, 구개열 말 특성에 대한 집중적인 듣기 훈련을 마친 언어병리학 전공 대학생을 대상으로 신뢰도를 측정하여 평가도구의 임상적 효율성과 추가적으로 개선되어야 할 사항에 대해서 살펴보고자 하였다.

## 연구방법

### 한국판 구개열 말 평가 개발

“한국판 구개열 말 평가”를 개발하기 위해 먼저 CAPS-A와 구개열 말 평가와 관련된 선행연구(Baylis, Chapman, & Whitehill, 2015; Chapman et al., 2016; Henningson et al., 2008; John et al., 2006; Peterson-Falzone, Trost-Cardamone, Karnell, & Hardin-Jones, 2010; Sell et al., 1994, 1999, 2009)를 문헌 조사하여 1차 평가 틀을 만들었다. 1차 평가 틀은 기본적으로 내용 및 기준 타당도가 입증된 CAPS-A를 토대로 하였다. 이후 1차적으로 개발된 평가 틀을 가지고 구개열 클리닉에서 구개열 아동을 전문으로 각각 10년과 6년 동안 평가 및 치료를 한 언어치료사 2인과 함께 내용 및 구성 타당도를 살펴보았다. 구체적으로 10개의 구개열 아동의 발화 샘플을 이용하여 독립적으로 평가한 뒤에, 평가결과를 함께 비교하면서 평가항목의 내용과 구성에 대해 논의하는 시간을 가졌다. 논의를 통해 수정이나 보완이 필요한 부분을 확인하고 도구의 임상적 타당성과 편리성을 점검하였다. 수정 작업에서는 타당도 검증에 참여

한 연구자와 언어치료사 2인의 의견에 따라 일부 항목의 순서와 내용, 평정법에서의 척도 개수를 수정하였다. 또한 선행연구의 문헌조사 결과를 참조하여 평가영역에 대한 보다 자세한 조작적 정의를 추가하였다. 이러한 타당도 검증과 수정 작업을 통해 본 연구에서 사용한 “한국판 구개열 말 평가”의 형식을 갖추었으며 이는 Appendix 1에 첨부하였다.

본 연구에 사용한 “한국판 구개열 말 평가”를 자세히 소개하면 크게 (1) 조음, (2) 말명료도, (3) 공명, (4) 들리는 비누출, (5) 음성, (6) 임상적 권고사항 및 요약으로 구성되어 있다. 첫 번째 조음 영역에서는 CAPS-A와 동일하게 발달적 오류와 구개열 관련 보상적 오류, 필연적 오류로 분류한 뒤 보상적 오류와 필연적 오류는 하위 오류 유형별로 평가하도록 하였다. 보상적 오류에는 중설구개음화, 연구개음화, 인두음화, 성문음화, 후비강 마찰음화, 비강 마찰음화, 동시조음이 포함되어 있으며 필연적 오류에는 구강자음의 약화, 파열음의 비음화, 마찰음/파찰음의 활음화가 포함되어 있다. 각 하위 오류 유형별로 조작적 정의를 간략하게 제시하였으며 3점 척도를 이용하여 오류 유형별로 출현 여부와 영향 받은 자음의 개수에 대한 정보를 대략 반영할 수 있도록 하였다. 또한 목표음소에 대한 평가 대상자의 반응을 기록할 수 있게 CAPS-A와 같이 자음표를 함께 제시하였다. 두 번째 항목은 말명료도로 7점 척도를 이용하여 심각도를 평가할 수 있게 하였다. CAPS-A에서는 말명료도 항목을 5점 척도를 이용하여 각 척도별로 조작적 정의를 포함하였으나 기술상의 모호함이 있었다. 따라서 한국판에서는 아동의 말에서 이해 가능한 부분을 비율로 제시하여(예: 아동의 말 10%-30% 이해 가능) 좀 더 구체적이고 객관적으로 평가할 수 있도록 수정하였다. 세 번째 항목은 공명으로 과다비성과 과소비성으로 분류되며 각각 7점과 2점 척도로 평정하게 하였다. 아동의 과다비성 문제가 검사어의 음운환경과 길이 등에 따라 비일관적으로 지각될 경우가 있어서 CAPS-A와 다르게 과다비성 심각도 외에 일관성 여부와 관련 요인을 선택적으로 표기하게 하였다. 네 번째 항목은 들리는 비누출로 CAPS-A에서는 들리는 비누출 외에 비강 난기류를 구별하여 평정하게 하였다. 하지만 비강 난기류는 일부 전문가 사이에서는 들리는 비누출과 동일하거나 심각한 형태로서 기술되기도 하여(Kummer, 2008; Peterson-Falzone, Hardin-Jones, & Karnell, 2010) 두 유형을 구별하지 않고 들리는 비누출의 항목으로만 평정하도록 하였다. 또한 CAPS-A와 다르게 특정 자음에서만 지각되는지 표기할 수 있도록 선택 사항을 부가적으로 포함시켰다. 다음 항목으로는 음성으로 CAPS-A에서는 음성문제 유무만 표기하도록 되어 있으나 구개열 아동 중 성문파열음으로 인해 이차적으로 거친 음성과 같은 음질상의 문제 외에도 강도, 음도상의 문제를 보이는 아동도 임상에

서 종종 관찰되므로 강도, 음도, 음질 의 세부항목을 두어 음성의 각 영역별로 평가하도록 하였다. 들리는 비누출과 음성 항목은 2점 척도로 문제 여부만을 파악하도록 하였다. 마지막으로 임상적 권고사항 및 요약에서는 언어치료의 필요성만을 결정하게 하는 CAPS-A를 보충하여 평가대상자가 보인 주요한 말 문제 영역 및 평가결과를 요약하고 언어치료 또는 2차 구개성형술 필요여부를 판단할 수 있게 하였다.

## 신뢰도 검사

### 검사자

“한국판 구개열 말 평가”의 신뢰도 검증을 위해서 언어병리학 전공 대학생 집단이 참여하였다. 언어병리학 전공 학부생은 총 33명으로 연구 당시 구개열 언어재활 수업을 모두 수강하였으며, 듣기훈련에 모두 참여하였다. 연구 당시 수업 외에 구개열 관련 임상경험은 갖고 있지 않았다.

### 구개열 말 자료

“한국판 구개열 말 평가”의 신뢰도 검증에는 구개열 클리닉에서 외래 언어평가를 받는 상황에서 수집된 8명의 구개열 아동(남 5명, 여 3명)의 음성 자료를 사용하였다. 오디오 자료 수집 당시 아동의 생활연령은 2세 6개월에서 4세 6개월(평균 3세 4개월)이었으며 1차 구개성형술 이후 언어치료와 부가적인 수술의 필요성을 결정하기 위해 약 6개월마다 정기적으로 언어평가에 참여하고 있었다. 아동 발화는 주로 우리말조음음운검사의 낱말 검사, 1-10까지 숫자세기, Ha와 Shin (2017) 에서 사용된 구강자음으로만 구성된 고모음 문장, 저모음 문장, 혼합모음 문장의 따라말하기로 구성되었다. 평가자가 아동 및 아동의 개인 정보에 대해 알 수 없도록 자료 수집 당시 포함된 언어치료사 또는 보호자의 발화는 모두 삭제하고 아동의 발화만이 포함되게 편집하였으며 MP3 파일 형태로 저장하였다.

### 신뢰도 검사 절차

신뢰도 검사에 앞서 먼저 언어병리학 전공 대학생 집단은 한 학기 동안 3학점에 해당하는 구개열 언어재활 교과목을 수강하였다. 수업을 통해 구개열 아동의 말 특징과 “한국판 구개열 말 평가”에 포함된 모든 평가 항목의 조작적 정의 및 개념을 학습하였다. 또한 각 평가 항목에 해당하는 음성 샘플을 들으면서 구개열과 관련된 말 문제에 익숙해지는 시간을 가졌다. 특히 신뢰도 검사 전에 약 4주에 걸쳐 12시간 동안 음성샘플을 통해 보상적, 필연적 오류의 하위 유형에 익숙해지고, 다양한 보상적 오류가 들어간 아동 발화에 대한 음성 전사도 실시하면서 연습하는 시간을 가졌다. 듣기 연습

마지막에는 “한국판 구개열 말 평가”를 이용해서 8명의 구개열 아동 자료를 독립적으로 듣고 개별적으로 평정하는 시간을 가짐으로써 학생들이 평가도구 사용에 익숙해졌으며 신뢰도 검사에 참여할 수 있는지 최종 확인하였다. 모든 듣기 연습에는 신뢰도 검사에 포함되지 않은 구개열 아동 및 성인의 오디오/비디오 자료를 이용하였다.

검사자 내 신뢰도를 위해서 모든 평가자는 1차 평가 실시 후 동일한 음성자료를 2주일 뒤에 재평가하였다. 검사자 간 신뢰도는 각 평가자가 두 번째로 평가했던 자료결과를 토대로 구하였다. 모든 평가자는 조용한 음성학 실습실 또는 강의실에서 노트북에 저장된 평가자료의 MP3 파일을 개별적으로 이어폰을 통해 들으면서 평가하였다. 음성의 강도는 평가자가 듣기 편안한 수준으로 조절할 수 있도록 하였다. 평가항목이 많으므로 아동의 파일을 평가자가 자율적으로 반복적으로 들을 수 있도록 허용하였다. 평가자가 순수하게 음성 자극만을 토대로 청지각적인 판단을 할 수 있도록 평가자에게는 2-4세 구개열 아동의 음성 자료라는 정보 외에는 구개열 아동의 개별적인 정보는 제공하지 않았다. 8개의 음성 자료를 듣고 “한국판 구개열 말 평가”의 모든 항목을 평가하는 데 평가자당 평균 30분 정도 소요되었다.

**자료 및 통계 분석**

검사자 간, 검사자 내 신뢰도는 “한국판 구개열 말 평가”에 구성된 하위 영역별로 구하였는데 조음과 공명 영역은 세부적으로 나누어서 구하였다. 특히 조음 영역에서 보상적 조음오류는 구강기제를 사용한 오류와 성문음화와 인두음화처럼 구강에서 조음되지 않은 비구강 조음오류, 비강 마찰음, 후비강 마찰음과 같이 기류 조

절의 실패로 나타난 조음오류, 구강과 비구강 기제를 동시에 사용하여 조음하는 동시조음으로 좀 더 세부적으로 나누어서 신뢰도를 살펴보았다. 따라서 신뢰도 검사결과는 (1) 발달적 조음오류, (2) 보상적 구강조음오류, (3) 보상적 비구강 조음오류, (4) 보상적 비강 오류, (5) 보상적 동시조음, (6) 필연적 조음오류, (7) 말명료도, (8) 과다비성, (9) 과소비성, (10) 들리는 비누출, (11) 음성, (12) 임상적 권고사항 및 요약으로 항목별로 분류해서 살펴보았다. 임상적 권고사항 및 요약은 언어치료와 2차 구개성형술의 필요성 여부와 관련된 판단 부분의 신뢰도를 살펴보았다. 검사자 간, 검사자 내 신뢰도는 급내 상관계수(intraclass correlation coefficient, ICC)를 이용하여 구하였다. 급내 상관계수는 SPSS 22.0 프로그램을 사용하여 95%의 신뢰구간을 설정하고 절대 동의와 이원혼합모형을 기본으로 구하였다.

**연구결과**

**검사자 간 신뢰도**

검사자 간 신뢰도 분석결과는 Table 1에 정리하였다. 급내 상관계수와 관련된 선행연구를 참조하여 급내 상관계수가 .50 미만인 경우는 낮은(poor) 상관관계로, .50과 .75 사이는 중간(moderate) 상관관계로, .75와 .90 사이는 양호한(good) 상관관계로 .90 이상은 우수한(excellent) 상관관계로 해석하였다(Fleis, 1986; Koo & Li, 2016; Shrout & Fleiss, 1979). 세부 항목별로 살펴보기 전에 모든 항목을 종합한 전체 검사자 간 신뢰도도 살펴보았는데 급내 상관계수가 .97로 우수한 상관관계를 보였다. 세부 항목별로 살펴보면 발달

**Table 1.** Inter-rater and intra-rater reliability results

Section	Inter-rater reliability		Intra-rater reliability	
	ICC	Interpretation	ICC	Interpretation
Entire sections	.97	Excellent	-	-
Developmental articulation errors	.08	Poor	.60	Moderate
Compensatory oral articulation errors	.81	Good	.68	Moderate
Compensatory non-oral articulation errors	.80	Good	.58	Moderate
Compensatory nasal articulation errors	.28	Poor	.52	Moderate
Compensatory co-articulation	.64	Moderate	.39	Poor
Obligatory articulation errors	.86	Good	.59	Moderate
Speech intelligibility	.98	Excellent	.87	Good
Hypernasality	.94	Excellent	.79	Good
Hyponasality	.83	Good	.58	Moderate
Audible nasal emission	.69	Excellent	.48	Poor
Voice	.83	Good	.59	Moderate
Clinical recommendations	.94	Excellent	.70	Moderate

ICC = intraclass correlation coefficient.

적 오류와 비강으로 기류를 보내면서 보상적으로 오조음하는 후비강 마찰음과 비강 마찰음으로 구성된 보상적 비강오류 영역에서 급내 상관계수가 .30 미만으로 낮은 상관관계를 보였다. 다음으로 보상적 동시조음과 들리는 비누출의 급내 상관계수가 .60대로 중간 정도의 검사자 간 신뢰도를 보였다. 보상적 구강조음오류, 보상적 비구강조음오류, 필연적 조음오류, 과소비성, 음성 항목에서는 급내 상관계수가 .80대로 양호한 검사자 간 신뢰도를 보였다. 말명료도, 과다비성, 임상적 권고 사항 항목에서는 급내 상관계수가 .90 이상으로 우수한 검사자 간 신뢰도를 보였다.

### 검사자 내 신뢰도

“한국판 구개열 말 평가”의 항목별로 검사자 내 신뢰도 분석 결과는 Table 1에 정리하였다. 보상적 동시조음과 들리는 비누출은 급내 상관계수가 .50 미만으로 낮은 검사자 내 신뢰도를 보였으며, 말명료도와 과다비성을 제외한 나머지 평가항목은 .50과 .75 사이로 중간 정도의 검사자 내 신뢰도를 보였다. 말명료도와 과다비성은 급내 상관계수가 각각 .87과 .79로 양호한 검사자 내 신뢰도를 보였다.

### 논의 및 결론

본 연구에서 소개한 “한국판 구개열 말 평가”는 CAPS-A를 우리말로 번역하고 구개열 말 특징 및 평가와 관련된 선행연구의 문헌 조사를 바탕으로 일부 항목을 수정하였다. 그리고 구개열 말 언어장애 전문 언어치료사 2인과 함께 실제 구개열 아동의 음성자료를 평가하면서 1차적으로 개발된 평가도구의 내용 및 구성 타당도를 살펴보았다. 타당도 검증 결과를 토대로 수정 작업을 실시하여 최종적으로 본 연구에서 사용된 평가도구로 개발하였다. 본 연구에서는 이러한 과정을 통해 개발된 “한국판 구개열 말 평가”를 소개하고, 구개열 언어재활 교과목 수강과 듣기 훈련을 마친 언어병리학 전공 대학생을 대상으로 검사도구의 신뢰도를 측정하여 평가도구의 임상적 효율성과 개선 사항들을 점검하고자 하였다.

구개열 언어재활 교과목을 통해 구개열 말 특성에 관해 학습하고 듣기 훈련을 마친 대학생이 구개열 아동의 음성자료를 평가했을 때 세부적으로 항목마다 신뢰도 수준은 다양하게 나타났으나 검사자간 신뢰도는 비교적 높게 나타났다. 특히 모든 항목을 종합해서 급내 상관계수를 구한 결과는 .97로 아주 우수한 검사자 간 신뢰도를 보였다. 세부 항목별 급내 상관계수를 토대로 검사자 간 신뢰도 결과를 살펴보면, 말명료도와 과다비성, 임상적 권고사항에서는 .90 이상으로 우수한 신뢰도를 보였다. 말명료도의 경우에는 아

동의 말을 평가자가 이해한 정도로 다른 항목에 비해 평가자의 주관적 판단이 들어갈 수 있어 평가자 간 다양한 평정을 내릴 수 있는 항목이다. 하지만 개발된 검사도구에서는 등간척도의 점수별로 정량화된 조작적 정의(예: 아동의 말 70% 이상 이해가능)를 함께 제시했기 때문에 평가자 간 일치도가 높게 나타난 것으로 판단된다. 과다비성과 임상적 권고사항도 우수한 신뢰도를 보였는데 다양한 사례의 오디오/비디오 자료를 이용하여 집중적인 듣기 연습을 했던 것이 효과가 있었던 것으로 보인다. 과다비성 심각도의 각 등급별로 3개 이상의 대표적 음성을 들려주었으며, 임상적 권고사항 항목도 사례자료를 제시한 후에 실제 임상에서 적용된 결과를 알려주는 등의 듣기 훈련이 유용했던 것으로 생각해볼 수 있다. 반면에 발달적 오류와 보상적 비강오류 영역에서 급내 상관계수가 .30 미만으로 낮은 검사자 간 신뢰도를 보였다. 발달적 오류 항목에서 낮은 신뢰도는 구개열 말 평가도구와 관련된 선행연구에서도 보고되고 있다(Chapman et al., 2016; Sell et al., 2009). 본 연구에서 발달적 오류 항목에서 검사자 간 신뢰도가 낮았던 이유로는 발달적 오류의 판단 기준 및 조작적 정의가 부족했던 점과 듣기 훈련이 구개열 말 특성에 해당하는 보상적 및 필연적 조음 오류에 초점을 맞춰 이루어진 점을 생각해볼 수 있다. 또한 음성 자료가 수집된 당시의 구개열 아동의 정확한 생활연령을 평가자에게 제공하지 않아, 발달적 오류 여부에 대해 평가자마다 다양한 판단을 하게 된 것 같다. 이와 비슷하게 Sell 등(2009)도 발달적 오류와 관련된 낮은 신뢰도의 이유로 훈련 부족, 주관적 판단의 여지가 많은 발달적 오류의 특성, 서로 중첩되는 부분이 있는 구개열로 인한 말 문제와 일반적인 말 발달 지연상의 문제 등을 제시하였다. 발달적 오류와 함께 낮은 검사자 간 신뢰도를 보인 보상적 비강 오류 영역에는 비강 마찰음과 후비강 마찰음이 포함되어 있다. 두 조음오류는 모두 아동이 구강 자음을 산출할 때 구개 천공(fistula)이나 불완전한 연인두 폐쇄로 인해 기류가 구강이 아닌 비강으로 들어가면서 소리가 만들어지는 유형이다(Peterson-Falzone et al., 2010). 비강 마찰음은 입술, 치조 또는 연구개 부위에서 조음 움직임은 있지만 기류는 구강이 아닌 비강으로 통과하고 소음을 동반하지 않은 채 오조음되는 패턴이다. 비강 마찰음은 종종 코 찡그림을 동반하거나 비강 마찰음을 산출할 때 대상자의 코를 막으면 말소리가 들리지 않는 것으로 확인할 수 있다. 따라서 오디오 자료만으로는 산출 여부를 정확하게 평가하는 것이 다른 오류 유형에 비해 상대적으로 어려울 수 있다. 마찬가지로 후비강 마찰음은 구개 천공이나 좁은 연인두 틈으로 기류가 통과하면서 만들어지는 소음을 단서로 평가하기 때문에 소음을 지각하지 못할 경우 정확하게 판단하기 어려울 수 있다. 이러한 평가의 어려움으로 인해 보상적 비강 오류영역에서 검사자간

에 통일성과 일관성이 떨어지는 검사결과가 나온 것으로 보인다.

1차적으로 평가를 실시한 후에 약 2주 후에 다시 실시한 평가를 토대로 측정된 검사자 내 신뢰도는 대부분의 항목에서 중간 또는 양호한 검사자 내 신뢰도를 보이기는 했으나 전반적으로 검사자 간 신뢰도보다 낮은 상관계수를 보였다. 특히 보상적 동시조음과 들리는 비누출은 급내 상관계수가 .50 미만으로 낮은 검사자 내 신뢰도를 보여 구개열 말 평가항목 중 평가자가 일관성 있게 판단하는 것이 어려운 항목임을 나타내고 있다. 반면에 말명료도와 과다비성은 검사자 간 신뢰도 결과와 마찬가지로 세부 항목 중에서 가장 높은 신뢰도를 보였다.

“한국판 구개열 말 평가”의 신뢰도 검사결과는 영국판, 미국판 구개열 말 평가를 이용하여 각각 신뢰도를 살펴본 John 등(2006)과 Chapman 등(2016)의 연구결과와 비교해서 검사자 간 신뢰도 부분에서는 일치된 결과를 보여주고 있다. 특히 구개열 말과 관련된 훈련 후에 검사자 간 신뢰도를 살펴본 Chapman 등(2016)은 모든 항목에서 일정하게 나타나는 않았지만 체계적인 훈련 후에 검사자 간 신뢰도가 향상되었음을 보여주었다. 또한 John 등(2006)의 연구에서도 검사자간, 검사자 내 신뢰도는 .41에서 1.00으로 양호한 신뢰도 결과가 나왔으며, 적절한 훈련을 통해서 CAPS-A를 비교적 쉽게 사용할 수 있다고 제안하였다. Chapman 등(2016)은 본 연구결과와는 다르게 검사자 내 신뢰도가 검사자간 신뢰도 보다 전반적으로 높게 나타났다. 이러한 상이한 결과는 검사자의 임상경험의 차이로 인한 것일 수 있다. Chapman 등(2016)은 6명의 구개열 전문 언어치료를 대상으로 개발된 평가도구를 이용하여 집중적인 훈련을 실시한 후에 신뢰도를 측정하였다. 따라서 본 연구에서 평가자로 참여한 구개열 관련 임상경험이 없는 대학생 집단보다는 구개열 말 문제와 관련하여 안정적인 내적 기준을 가지고 있어 상대적으로 높은 검사-재검사 간 일치도를 보였을 것이다. 이러한 결과는 집중적인 듣기 훈련을 통해 비교적 짧은 시간 내에 구개열 말 특성과 관련된 비슷한 내적 기준을 평가자 간에 갖출 수는 있으나 평가자가 시간이 지나도 변하지 않는 안정적인 내적 기준을 갖기 위해서는 임상경험이나 다른 요소가 추가적으로 필요함을 제시하고 있다. 여러 선행연구가 장애 음성에 대한 가변적인 내적 기준을 보완할 수 있는 방법으로 외적 기준 사용이 효과적이라고 제안하였다 (Chan & Yiu, 2006; Gerratt, Kreiman, Antonanzas-Barroso, & Berke, 1993; Ha, 2010; Kreiman et al., 1993). 추후에는 구개열 말 특성 각각을 대표하는 한국 아동 또는 성인의 말샘플 모음집을 체계적으로 만들어 언어치료전공 학생과 언어치료를 위한 전문적인 훈련에 활용하여 정확하고 일관성 있는 구개열 말 평가가 이루어질 수 있는 기반을 마련하는 것이 필요하다.

본 연구에서 개발된 “한국판 구개열 말 평가”는 영국에서 타당도와 신뢰도 검증을 통해 구개열 말 문제를 전문적으로 평가할 수 있는 도구로서의 유용성이 입증된 CAPS-A를 기본으로 하고 있다. 두 검사도구는 공통적으로 음성전사방법에 의존하지 않고 오류 패턴의 유무를 확인하고 영향 받는 자음의 개수를 토대로 대략적인 심각도 정보를 기록할 수 있다. 말소리를 자세하게 음성전사하는 것이 말 문제를 분석하는 것에는 이상적이지만 구개열 말과 같은 특수장애영역의 음성전사는 높은 전사자 간/전사자 내 신뢰도를 이루기가 어렵다(Gooch, Hardin-Jones, Chapman, Trost-Cardamone, & Sussman, 2001; Kent 1996). 따라서 음성전사의 대안으로 오류 유형의 유무와 심각도를 판단하는 CAPS-A는 실제 임상현장에서는 보다 편리하게 사용될 수 있다. 본 연구는 구개열과 관련된 임상경험이 없는 대학생들도 구개열 말 특성에 대한 학습과 듣기 훈련을 받은 후에는 개발된 평가도구를 이용하여 구개열 말의 일부 특성을 청각적으로 일관성 있게 평가할 수 있음을 보여주었다.

본 연구에서는 표준화된 조음음운장애 검사도구의 낱말 리스트, 자동구어, 고모음과 저모음으로 구성된 문장, 자발화 등의 현재 임상에서 사용되고 있는 말자료를 토대로 평가하였다. 구개열 말 문제를 좀 더 정확하고 타당하게 평가하기 위해서는 표준화된 검사어와 문장을 사용하는 것이 효과적이다. 추후에는 구개열 말 특징이 보다 뚜렷하게 나타나는 음운 환경을 고려해서 검사어와 단어를 개발하고 이를 검증하는 과정이 필요하다. 또한 검사자 간, 검사자 내 낮은 신뢰도를 보였던 항목에 대해서는 추후 듣기 훈련에서 좀 더 다양한 음성 샘플을 이용하여 강조하거나 검사도구 내에 조작적 정의 및 설명을 좀 더 보완해야 한다. 또한 평가항목 및 분류 체계, 평가척도 유형, 평가등급의 개수 등에 대한 지속적인 검증작업을 실시하여, 임상현장에서 구개열 말 특성을 보다 정확하고 일관성 있게 평가할 수 있도록 개선시켜야 한다.

## ACKNOWLEDGEMENT

한국판 구개열 말평가 도구의 타당성을 검증해주신 정승은, 문희원 선생님께 감사드립니다.

## REFERENCES

- Baylis, A., Chapman, K., & Whitehill, T. L. (2015). Validity and reliability of visual analog scaling for assessment of hypernasality and audible nasal emission in children with repaired cleft palate. *The Cleft Palate-Craniofacial Journal*, 52, 660-670.

- Britton, L., Albery, L., Bowden, M., Harding-Bell, A., Phippen, G., & Sell, D. (2014). A cross-sectional cohort study of speech in five-year-olds with cleft palate±lip to support development of national audit standards: benchmarking speech standards in the United Kingdom. *The Cleft Palate-Craniofacial Journal*, *51*, 431-451.
- Chan, K. M., & Yiu, E. M. L. (2006). A comparison of two perceptual voice evaluation training programs for naive listeners. *Journal of Voice*, *20*, 229-241.
- Chapman, K. L., Baylis, A., Trost-Cardamone, J., Cordero, K. N., Dixon, A., Dobbelsteyn, C., ... & Stoddard, G. (2016). The Americleft Speech Project: a training and reliability study. *The Cleft Palate-Craniofacial Journal*, *53*, 93-108.
- Eadie, T. L., & Baylor, C. R. (2006). The effect of perceptual training on inexperienced listeners' judgments of dysphonic voice. *Journal of Voice*, *20*, 527-544.
- Ellis, L. W., & Belyukova, S. A. (2008). Effects of training on naïve listeners' judgments of the speech intelligibility of children with severe-to-profound hearing loss. *Journal of Speech, Language, and Hearing Research*, *51*, 1114-1123.
- Fleis, J. (1986). *Design and analysis of clinical experiments*. New York, NY: John Wiley & Sons.
- Gerratt, B. R., Kreiman, J., Antonanzas-Barroso, N., & Berke, G. S. (1993). Comparing internal and external standards in voice quality judgments. *Journal of Speech, Language, and Hearing Research*, *36*, 14-20.
- Gooch, J. L., Hardin-Jones, M., Chapman, K. L., Trost-Cardamone, J. E., & Sussman, J. (2001). Reliability of listener transcriptions of compensatory articulations. *The Cleft Palate-Craniofacial Journal*, *38*, 59-67.
- Ha, S. (2010). Effects of listener training and external standard on the reliability of perceptual judgment of hypernasality. *Korean Journal of Communication & Disorders*, *15*, 411-421.
- Ha, S., & Shin, I. (2017). Effects of vowel context and stimulus length on test-retest nasalance variability between children and adults with normal speech. *Communication Sciences & Disorders*, *22*, 578-586.
- Henningsson, G., Kuehn, D. P., Sell, D., Sweeney, T., Trost-Cardamone, J. E., & Whitehill, T. L. (2008). Universal parameters for reporting speech outcomes in individuals with cleft palate. *The Cleft Palate-Craniofacial Journal*, *45*, 1-17.
- John, A., Sell, D., Sweeney, T., Harding-Bell, A., & Williams, A. (2006). The cleft audit protocol for speech—augmented: a validated and reliable measure for auditing cleft speech. *The Cleft Palate-Craniofacial Journal*, *43*, 272-288.
- Kent, R. D. (1996). Hearing and believing: some limits to the auditory-perceptual assessment of speech and voice disorders. *American Journal of Speech-Language Pathology*, *5*, 7-23.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*, 155-163.
- Kreiman, J., Gerratt, B., Kempster, G., Erman, A., Berke, G. (1993). Perceptual evaluation of voice quality: review, tutorial, and a framework for future research. *J Speech Hear Res*, *36*, 21-40.
- Kuehn, D. P., & Moller, K. T. (2000). Speech and language issues in the cleft palate population: the state of the art. *The Cleft Palate-Craniofacial Journal*, *37*, 1-35.
- Kummer, A. W. (2008). *Cleft palate and craniofacial anomalies: effects on speech and resonance*. Clifton Park, NY: Delmar/Cengage Learning.
- Lee, A., Whitehill, T. L., & Ciocca, V. (2009). Effect of listener training on perceptual judgement of hypernasality. *Clinical Linguistics & Phonetics*, *23*, 319-334.
- Lewis, K. E., Watterson, T. L., & Houghton, S. M. (2003). The influence of listener experience and academic training on ratings of nasality. *Journal of Communication Disorders*, *36*, 49-58.
- Peterson-Falzone, S. J., Hardin-Jones, M. A., & Karnell, M. P. (2010). *Cleft palate speech* (4th ed.). St. Louis, MO: Mosby.
- Peterson-Falzone, S. J., Trost-Cardamone, J. E., Karnell, M. P. & Hardin-Jones, M. A. (2010). *The clinician's guide to cleft palate speech*. St. Louis, MO: Mosby.
- Sell, D. (2005). Issues in perceptual speech analysis in cleft palate and related disorders: a review. *International Journal of Language & Communication Disorders*, *40*, 103-121.
- Sell, D., Harding, A. & Grunwell, P. (1994). GOS.SP.ASS. A screening assessment of cleft palate speech. *European Journal of Disorders of Communication*, *29*, 1-15.
- Sell, D., Harding, A., & Grunwell, P. (1999). Revised GOS. SP. ASS (98): Speech assessment for children with cleft palate and/or velopharyngeal dysfunction. *International Journal of Disorders of Communication*, *34*, 7-33.
- Sell, D., John, A., Harding-Bell, A., Sweeney, T., Hegarty, F., & Freeman, J. (2009). Cleft Audit Protocol for Speech (CAPS-A): a comprehensive training package for speech analysis. *International Journal of Language & Communication Disorders*, *44*, 529-548.
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin*, *86*, 420-428.

Appendix 1. 한국판 구개열 말 평가(The Cleft Audit Protocol for Speech-Augmented-Korean modification)

아동명(파일명)		성별	
생년월일		검사일	
생활연령		구개열 유형	

1. 조음

(1) 자음표

자음	m	p	p*	p <sup>h</sup>	n	l	t	t*	t <sup>h</sup>	s	s*	ʈ	ʈ*	ʈ <sup>h</sup>	k	k*	k <sup>h</sup>	ŋ	h
초성																			
어중중성																			
어말중성																			

(2) 오류형태

	없음	2개이하 자음에서 나타남	3개이상 자음에서 나타남	오류패턴(예)
(2a)발달적 오류				

(2b)구개열 보상조음	없음	2개이하 자음	3개이상 자음	설명
중설 구개음화 (mid dorsum palatalization)				가운데 혀 등을 경구개에 대고 조음
연구개음화				조음점이 연구개/목젖으로 후방화
인두음화				인두파열음/인두마찰음으로 대치
성문음화				성문파열음/성문마찰음으로 대치
후비강 마찰음화				연인두 부위의 마찰성 소음 동반
비강 마찰음화				구강이 닫힌 채 기류가 비강으로 통과
동시조음				구강 기제와 인두/성문 함께 사용

(2c)필연적 오류	없음	2개이하 자음	3개이상 자음	설명
구강자음의 약화				구강음의 에너지 약화, 비강공명지각
파열음의 비음화				구강음이 조음점 같은 비음으로 산출
마찰음/파찰음의 활음화				마찰음/파찰음이 [w, j]로 대체됨

2. 말명료도

등급	설명	평가 결과
0	아동의 말 모두 이해가능하고 정상임	
1	아동의 말 모두 이해가능하나 또래에 비해 떨어짐	
2	아동의 말 70% 이상 이해가능	
3	아동의 말 70-50% 이해가능	
4	아동의 말 30-50% 이해가능	
5	아동의 말 10-30% 이해가능	
6	아동의 말 모두 이해 불가능	

(Continued to the next page)



Appendix 1. Continued

3. 공명

(1) 과다비성 (hypernasality)

심각도 (severity)		
등급	설명	평가 결과
0	정상 범위	
1	약간의 비성이 지각됨(minimal)	
2	경도 수준의 비성 지각됨 (mild)	
3	경중도 (Mild to moderate)	
4	중도 (Moderate)	
5	중심도 (Moderate to severe)	
6	심도(Severe)	
(선택 사항) 일관성 (consistency) - 녹음된 발화 샘플 내에서 변화하는 정도		
등급	설명	평가 결과
0	비성 정도가 일관적임	
1	비성 정도가 비일관적임	
N/A	평가가 불가능하거나 제한적임	
일관성 관련 요인		

(2) 과소비성 (hyponasality)

등급	설명	평가 결과
0	없음	
1	있음	

4. 들리는 비누출 (Audible nasal emission)

등급	설명	평가 결과		
0	없음			
1	있음			
(선택 사항) 관련 요인		예	아니오	모름
특정 자음에서만 지각된다				
특정 자음에서만 지각될 경우 해당 자음은?				

5. 음성

등급	설명	강도 (loudness)	음도 (pitch)	음질 (quality)
0	정상범위			
1	정상범위에서 뚜렷하게 벗어남			
비고)				

6. 임상적 권고사항 및 요약

대상자가 보인 주요한 말 문제 영역 및 결과 요약			
	예	아니오	좀 더 지켜봐야 함
언어치료 필요 여부			
2차 구개성형술 필요 여부			

## 국문초록

### 한국판 구개열 말 평가 개발 및 신뢰도 검사

하승희

한림대학교 언어청각학부, 청각언어연구소

**배경 및 목적:** 현재 국내 임상현장에는 구개열 화자의 말 문제를 청지각적으로 평가하는 데 사용할 수 있는 전문적인 평가도구가 없다. 따라서 본 연구는 한국판 구개열 말 평가도구를 개발하고, 평가도구의 항목별 검사자 간, 검사자 내 신뢰도를 살펴보았다. **방법:** Cleft Audit Protocol for Speech-Augmented를 한국어로 번역하고 일부 항목의 형식과 정의를 수정하였다. 구개열 전문 언어치료사 2인과 함께 타당성을 검증하고 “한국판 구개열 말 평가”를 개발하였다. 구개열 언어재활 교과목을 수강하고 듣기 훈련을 마친 언어병리학 전공 대학생이 개발된 평가도구를 이용하여 구개열 아동 음성자료를 두 번씩 평가하였다. 급내 상관계수를 이용하여 검사자 간, 검사자 내 신뢰도를 구하였다. **결과:** 평가항목 중 말명료도, 과다비성, 임상적 권고사항에서는 검사자 간 신뢰도가 높았다. 낮은 검사자 간 신뢰도를 보인 발달적 오류와 보상적 비강 오류를 제외하고는 모든 항목에서 중간 또는 양호한 검사자 간 신뢰도를 보였다. 검사자 간 신뢰도보다 검사자 내 신뢰도가 전반적으로 낮았는데 대부분 중간 정도의 신뢰도를 보였다. 보상적 동시조음과 들리는 비누출은 낮은 신뢰도를 보였으나, 말명료도와 과다비성은 양호한 신뢰도를 보였다. **논의 및 결론:** 일정 수준의 훈련을 받은 후에 개발된 평가도구를 이용해서 구개열 대상자의 말 특성을 일관성 있게 평가할 수 있었다. 신뢰도 검사결과와 평가도구에서 개선해야 할 사항과 평가자를 위한 훈련의 방향과 강조점에 대해 시사하고 있다.

**핵심어:** 구개열, 말 평가, 검사자 간 신뢰도, 검사자 내 신뢰도

## 참고문헌

하승희, 신일산(2017). 검사어의 모음 환경과 길이에 따른 성인과 아동의 비음치 변이성. *언어청각장애연구*, 22, 578-586.

하승희(2010). 듣기 훈련과 외적 음성 기준 사용이 과다비성의 청지각적 평가 신뢰도에 미치는 효과. *언어청각장애연구*, 15, 411-421.