

유창성장애 전문가들의 임상적 판단과 파라다이스-유창성 검사 결과 간의 비교 연구

이은주^a · 박정현^b · 신문자^c · 심현섭^{b,d,§}

^a단국대학교 특수교육과, ^b이화여자대학교 대학원 언어병리학협동과정, ^c조선대학교 언어치료학부,

^d이화여자대학교 특수교육과

§ 교신저자

심현섭

이화여자대학교 대학원

언어병리학협동과정 교수

서울특별시 서대문구 대현동 11-1

e-mail: simhs@ewha.ac.kr

tel.: 02-3277-2120

배경 및 목적: 본 연구에서는 유창성장애 전문가들의 주관적인 임상적 판단과 파라다이스-유창성 검사(Paradise-Fluency Assessment: P-FA, 이하 P-FA)를 이용한 객관적 결과를 비교함으로써, P-FA 개정 방향에 참고하고자 하였다. **방법:** 유창성장애 전문가 5명을 선정하여 취학전 아동, 초등학생, 성인 12명씩 총 36명의 P-FA 실시장면을 녹화한 자료를 시청한 후 1차와 2차에 걸쳐 말더듬 정도를 판단하도록 하였다. 그리고, 다른 연구자 2명은 P-FA 절차에 따라 자료를 분석하여 말더듬 정도를 산출하도록 하였다. **결과:** 연구자들 간의 판단 일치도는 상당히 높은 편이었으며, 1차와 2차 판단 간에도 큰 변화가 없었다. 연구자들의 판단은 P-FA 결과보다 낮은 경향을 보였다. 취학전 아동과 성인의 경우에는 전문가들의 임상적 판단과 P-FA 측정치들 사이에 유의한 상관관계가 없었으며, 초등학생의 경우에는 몇 가지 측정치에서 유의한 상관관계가 있었다. **논의 및 결론:** 본 연구의 결과는 P-FA의 개정방향 뿐만 아니라 말더듬의 평가에 대하여 의미있는 논의를 제공하고 있다. 『언어청각장애연구』, 2009;14:200-211.

핵심어: 파라다이스-유창성 검사, 임상적 판단, 전문가

I. 서론

파라다이스-유창성검사(Paradise-Fluency Assessment: P-FA, 이하 P-FA)(심현섭 · 신문자 · 이은주, 2004)는 2001년부터 2003년까지 표준화과정을 거쳐 개발된 유창성장애 검사도구로 2004년부터 연구와 임상현장에서 사용되고 있다. 현재 연구자들은 검사 과제와 해석방법을 수정보완하고 좀더 많은 말더듬 대상자들을 포함한 기준을 제시하기 위하여 P-FA의 개정을 준비하고 있다. P-FA 개정을 위한 기초 작업으로 P-FA를 직접 사용하고 있는 언어치료 전문가들과의 면담을 실시하였는데, 대부분의 참여자가 P-FA 결과와 주관적으로 판단한 인상(general impression)이 일치하지 않는 경우가 많아서 진단에 어려움을 느낀다고 보고하였다. 이러한 차이는 연령

이나 대상자의 증증도에 따라 다른 경향을 보이는데, 전체적으로 아동보다는 성인일 경우에 차이가 있고 자신이나 주위의 보고에 비해 P-FA 결과가 더 약한 편인 경우도 있지만, 주로 임상적 판단에 비해 P-FA 분석에 따른 증증도가 더 심하다고 느끼고 있는 것으로 나타났다. 몇몇 전문가들은 P-FA의 경우, 주저, 미완성 및 수정, 낱말이상의 반복과 같은 정상비유창성(normal disfluency: ND, 이하 ND)을 점수에 반영하기 때문에 이러한 차이가 발생하는 것으로 해석하였다(심현섭 · 신문자 · 이은주, 2009).

주관적 판단과 P-FA 결과 간의 불일치가 클수록 임상에서 치료사들, 특히 경험이 적은 치료사들은 진단을 하거나 중재를 계획하기 어려울 수 있다. 때문에 이러한 불일치가 실제로 나타나는지, 어느 정도로 불일치하는지를 파악하여야 할 것으로 보인다. 그러나

* 이 논문은 2008년 재단법인 파라다이스복지재단의 지원을 받아 연구되었음.

한편으로 이러한 어려움이 단순히 P-FA라는 검사도 구만의 문제라기보다 유창성장애의 본질적인 특성에서 기인할 수 있다는 점을 고려하여야 한다.

많은 연구자들이 지속적으로 관심을 갖고 있는 말더듬 관련 연구주제에는 말더듬 원인이나 중재 외에도 ‘말더듬 순간(stuttering event or moment)의 확인과 측정’이라는 주제가 포함되어 있다. 언어병리학의 연구와 치료는 인간 행동을 직접 관찰하는 것으로 이루어지기 때문에 신뢰도와 타당도를 확보하는 것이 매우 중요한데, 특히 유창성장애의 경우에는 말더듬 순간에 대한 타당한 정의와 신뢰할 만한 측정방법이 필요하기 때문이다(Cordes, 1994; Cordes & Ingham, 1994).

비유창성 유형의 분류에 대해서는 여러 가지 비유창성 유형을 이분법적으로 분류하거나, SSI-3 (Stuttering Severity Instrument for Children and adults-3, 이하 SSI-3)(Riley, 1994)와 같이 말더듬는 사람에게 더 많이 발생하는 범주만을 말더듬으로 간주하기도 한다. 하지만, SSI-3와 같은 방식은 정상인에게도 많이 발생할 수 있는 범주도 말더듬으로 판단되는 경우가 많기 때문에 동반되는 양상에 대한 고려 없이 유형만을 판단하여 일률적으로 제외시키는 것은 문제가 있다고 지적받고 있다(Cordes & Ingham, 1995; Yairi, 1996). 실제로 발화 순간에 예기치 않게 발생하는 비유창성은 단순히 유형에 따라 말더듬 여부를 판단하기 어렵다(Curlee, 1981). 말더듬의 판단에 영향을 미치는 여러 가지 변수들이 있기 때문이다. 이분법적 범주화에 따라 비유창성 유형을 말더듬과 정상비유창성으로 구분한 결과는 지각적인 말더듬 판단과 낮은 일치율을 보였다(Bothe, 2008; Onslow et al., 1992). Yairi (1996)는 Conture의 낱말내 비유창성과 낱말간 비유창성의 범주화를 비판하였는데, 그 이유는 1음절 낱말 혹은 다음절 낱말의 반복이 때에 따라 정상적일 수도 있고 말더듬일 수도 있기 때문이라고 하였다. 1음절 낱말 반복은 특히 취학전 말더듬 아동일 때, 다음절 낱말 반복과 같은 낱말간 비유창성은 말더듬 성인일 때 중요한 특징이라고 하였다. 이처럼, 낱말간 비유창성이 반복 횟수나 지각적 양상에 따라 말더듬으로 평가될 수 있기 때문에 연구자들은 비유창성 유형과 말더듬 여부를 판단은 별개로 생각해야 한다고 조언하기도 한다(Cordes & Ingham, 1994, 1996; Ryan, 1974).

평가자의 말더듬 판단 일치율에 관한 연구 결과들

을 살펴보면, 말더듬 빈도에 대한 일치율에 비해 발생 위치 및 말더듬 유형에 대한 일치율은 낮은 편이다(심현섭 · 이은주, 2002; McDonald & Martin, 1973; Young, 1969, 1975; Young & Downs, 1968). 따라서, 비유창성의 발생위치와 유형에 대한 판단 일치율을 높이기 위한 다양한 방법이 꾸준히 연구되고 있다(Cordes, 2000; Howell, Sackin & Glenn, 1997; Ingham, Cordes & Gow, 1993; Ingham et al., 1995; Yaruss, 1998).

일치율 연구들에 따르면, 유창성장애 평가에 숙련된 전문가들은 일반 언어치료사(general therapist), 언어병리학 전공 대학원생 등에 비해 일치율이 높은 것으로 나타났다(Brundage et al., 2006; Ingham & Cordes, 1997). 유창성장애의 판단 일치율에 관한 연구에서는 전문가를 피험자로 하거나 전문가들의 분석결과를 분석기준으로 삼는 연구들이 많다(Bothe, 2008; Brundage et al., 2006; Ingham & Cordes, 1997; Teesson, Packman & Onslow, 2003). 유창성장애는 다른 장애와 달리 평가자 개인에 따른 평가 결과의 편차가 크고, 말더듬에 대한 조작적 정의가 평가자나 검사도구에 따라 다를 수 있고, 말더듬으로 판단할 수 있는 행동의 범위가 넓어서 과대평가하거나 놓칠 위험이 크며, 무엇보다 말더듬는 사람에 따라 변이성이 커서 많은 경험과 훈련이 필요하기 때문이다.

발화샘플에서 말더듬을 분석하는 것은 진단과 예후 파악에서부터 치료효과의 검증에 이르는 중재과정에서 매우 중요하다. 그러나 말더듬의 분석은 시간이 많이 걸리고 에너지를 소모하게 되는 과정임에도 불구하고, 평가자에 따라 일치하기가 쉽지 않다(Howell, Sackin & Glenn, 1998; Onslow & O'Brian, 1998). 따라서 ‘청자의 지각에 근거한 평정척도’를 사용하여 말더듬 정도를 대략적으로 평가하거나, 말더듬 순간을 구체적으로 유형 분석하지 않고 빈도를 구하는 것으로 평가하기도 한다. 다양한 구간의 평정척도를 이용하는 평가방법은 언어치료사에 의해 또는 말더듬 대상자 본인에 의해 실제 연구와 치료에서 활발하게 사용되고 있다(Manning, 2001; O'Brian et al., 2004). 평정척도를 사용한 전반적인(overall) 말더듬 정도와 더듬은 음절 비율 간의 비교 연구에 따르면, 두 가지 방법 사이에는 유의한 상관관계가 있었다(O'Brian et al., 2004; Rousseau et al., 2008). 따라서, 유창성장애의 평가 및 치료 시에 즉석에서 편리하게 말더듬 정

도를 파악하는 것이 가능하며, 유창성장애 대상자가 치료실 밖에서 스스로 자신의 말을 관리하기 위해 사용할 수 있는 것으로 나타났다(O'Brian, Packman & Onslow, 2004).

이와 달리, 위에서 언급하였듯이 P-FA를 사용하고 있는 언어치료사들은 전반적인 임상적 인상과 P-FA 결과 간에 차이가 있다고 하였다. 따라서 본 연구에서는 전문가들을 대상으로 평정척도를 사용하여 말더듬을 평가하도록 하여 전문가들 간의 일치 정도를 살펴보고, 평정척도 결과와 P-FA 결과 사이에 어떠한 관계가 있는지 알아보고자 하였다.

II. 연구 방법

1. 연구 참여자

현재 학부 및 대학원에서 유창성장애 과목을 가르치고 있으며, 임상현장에서 유창성장애의 진단 및 치료 경험이 풍부하고, 유창성장애 아동 및 성인을 대상으로 한 논문으로 최종학위를 취득한 유창성장애 전문가들(이하 '전문가')이 연구에 참여하였다. 전문가들에 대한 정보는 <표-1>과 같다.

<표-1> 연구 참여자 특성

이름	최종학위논문영역	최종학위	언어치료사자격증	현재 소속
P1	유창성장애	박사	1급	00대학교 언어병리학협동과정
P2	유창성장애	박사	1급	00대학교 언어치료학부
P3	유창성장애	박사	1급	00대학교 언어치료학과
P4	유창성장애	석사	1급	000 언어치료실 (원장)
P5	유창성장애	박사	1급	00대학교 특수교육과

2. 연구절차

가. 피험자 선정

P-FA의 표준화 연구에 참여한 정상 및 말더듬 대상자들 중 취학전 아동 12명(남 11명, 여 1명), 초등학교

생 12명(남 12명), 성인 12명(남 10명, 여 2명)을 무작위로 선별하여 P-FA의 실시 과정을 디지털캠코더로 녹화한 자료를 사용하였다. 취학전 아동과 초등학교생의 자료는 '말하기 그림' 과제 실시 장면을, 성인 자료는 '대화' 과제 실시 장면을 사용하였다. 주관적 판단에 참여한 유창성장애 전문가 5인은 피험자 선정에 관여하지 않았으며, 본 연구의 공동연구자와 보조연구자가 피험자를 선정하였다. 연령별로 포함된 정상 피험자들은 P-FA 표준화 연구 참여시에 말더듬 문제가 없는 것으로 본인이나 부모가 답한 참여자들 중에 선정하였다.

나. 파라다이스-유창성 검사 결과 분석

본 연구의 공동연구자 중 1명과 보조연구자 1명이 피험자 36명의 P-FA 녹화 자료를 분석하였다. 녹화된 동영상을 보면서 P-FA의 분석 절차에 따라 각 피험자의 발화를 모두 전사하고, 출현한 비유창성의 유형을 판단한 후에 과제별 점수, 말더듬 정도를 구하였다. 분석자들은 모두 언어병리학과에서 유창성장애 과목을 수강하였고, 임상현장에서 주로 유창성장애 아동 및 성인을 대상으로 평가와 치료를 하고 있으며, P-FA 개발 연구에 참여하였기 때문에 유창성장애 평가 경험이 매우 풍부하였다. 두 연구자는 주관적 판단에는 참여하지 않았다.

<표-2>는 피험자 36명의 P-FA 검사 결과를 말더듬 증증도에 따라 정리한 결과이다. 취학전 아동은 약함 3명, 중간 4명, 심함 5명이었으며, 초등학교생은 약함 3명, 심함 9명이었고, 성인은 약함 2명, 중간 7명, 심함 3명이었다.

<표-2> P-FA 검사 결과

	약함(1)	중간(2)	심함(3)	합
취학전	3 ^{a)}	4	5	12
초등학교생	3		9	12
성인	2	7	3	12
합	8	11	17	36

^{a)} 단위: 명

다. 1차 및 2차 판단과 결과

피험자들의 검사장면을 평가하는 순서는 취학전, 학령기, 성인 집단의 순이었으며, 집단 내에서는 무작위 순서로 하였다. 피험자에 대한 정보를 전혀 알려주

지 않고 검사장면을 시청하도록 한 후 전문가들 각자가 자신의 기록지에 피험자의 말더듬 정도를 1차 판단하여 기록하도록 하였다. 모든 참여자들이 1차 판단 기록을 마친 후에 연구진행자가 대상자의 연령, 말더듬 시작시기, 의사소통태도검사결과 등을 알려주고, 바로 판단을 수정할 수 있도록 2차 판단을 기록할 시간을 주었다. 말더듬 정도의 판단은 ‘정상, 약함, 중간, 심함’ 등으로 나누어 표로 제시하고 참여자들이 해당 칸에 표시하도록 하였다. 피험자 1명당 평균 시청 시간은 3~5분이었으며, 4명의 자료를 시청한 후 10분씩 휴식하도록 하였다. 전체 36명 피험자의 동영상 자료를 시청하고 판단하는데 총 4시간 가량의 시간이 소요되었다.

연구참여자들이 판단한 결과는 <부록-1>에 제시하였다. 각 피험자에 대한 참여자들의 말더듬정도 판단결과는 3명 이상이 일치하게 판단한 말더듬 정도로 정하였다. 3명 이상이 일치하게 판단하지 않았을 경우에는 ‘해당없음’으로 처리하였다.

라. 신뢰도 분석

말더듬 정도 및 비유창성 발생 횟수에 대한 신뢰도를 분석하기 위하여, 연구자 1명과 보조연구자 1명이 36명 중 6명의 자료를 각자 독립적으로 분석한 결과를 비교하였다. 비유창성의 발생 위치에 대한 일치율은 97.7%, 정상적 비유창성 유형에 대한 일치율은 94.8%, 비정상적 비유창성 유형에 대한 일치율은 99.2%였다.

3. 연구 분석방법

SPSS 14.0을 사용하여 통계처리하였으며, 피험자 수를 고려하여 비모수 검정인 Kruskal-Wallis 검정과 Spearman 상관분석을 실시하였다.

Ⅲ. 연구 결과

1. 전문가 판단

전문가들이 판단한 결과는 <부록-1>에 제시하였으며, 1차와 2차 판단 결과를 종합한 결과는 <표-3>과 같다. 전문가들 간에 일치율은 1차와 2차 판단에서 각각 76%로 나타났고, 거의 모든 경우에 5명

중 3명 이상이 일치한 판단을 한 것으로 볼 수 있다. 3명 이상의 일치가 이루어지지 않아 ‘해당없음(na)’으로 처리된 피험자는 1차 판단의 경우 취학전 아동 1명이었으며, 2차 판단의 경우 취학전 아동 2명이었다.

<표-3> 전문가들 판단의 결과 및 일치수

집단별 피험자 번호	취학전 아동		초등학생		성인		
	결과 ^{*)}	일치수	결과	일치수	결과	일치수	
판 단 1	1	1	4	2	4	1	4
	2	1	3	3	3	0	3
	3	3	3	3	3	0	5
	4	0	4	2	3	0	3
	5	1	3	2	4	1	5
	6	3	4	3	4	1	5
	7	na	na	1	3	1	5
	8	2	4	1	5	2	4
	9	1	5	1	5	3	3
	10	3	3	1	4	1	3
	11	1	5	0	5	3	3
	12	1	4	0	5	2	3
판 단 2	1	1	3	2	4	1	4
	2	1	4	3	3	1	3
	3	2	4	2	3	0	5
	4	0	4	2	3	1	3
	5	1	3	2	4	1	5
	6	3	5	3	4	1	5
	7	1	5	1	4	1	5
	8	na	na	1	4	2	3
	9	1	4	1	5	3	4
	10	3	3	1	5	1	3
	11	1	5	0	5	3	4
	12	na	na	0	5	2	3

*) ‘결과’는 3명 이상이 일치하게 판단한 말더듬 정도로 결정하였으며, ‘0’은 정상, ‘1’은 약함, ‘2’는 중간, ‘3’은 심함을 의미함.

전문가들의 판단결과에 따라 본 연구에 포함된 피험자들의 말더듬 정도를 요약하면 <표-4>와 같다. 첫 번째 판단에 따르면 전체 피험자 36명 중 정상 6명, 약함 15명, 중간 6명, 심함 8명이었으며, 두 번째 판단에 따르면 정상 5명, 약함 17명, 중간 7명, 심함 6명이었다.

<표 -4> 전문가 판단결과에 따른 피험자의 말더듬 정도

	정상		약함		중간		심함		na	
	1st	2nd	1st	2nd	1st	2nd	1st	2nd	1st	2nd
취학전	1	2	6	6	1	1	3	2	1	2
초등학생	2	2	4	4	3	4	3	2	0	0
성인	3	1	5	7	2	2	2	2	0	0
합계	6	5	15	17	6	7	8	6	1	2

<부록 - 1>에 따르면, 1차 판단 후 2차 판단을 하는 과정에서 발생한 변화는 <표 - 5>와 같다. P3는 피험자 36명 중 33명(92%)에 대해, P1, P2, P5는 30명(83%)에 대해, P4는 28명(78%)에 대해 일관된 판단을 하였다. 전문가 5명 모두 1차 판단 후 주어진 정보로부터 큰 영향을 받지 않는 경향을 보여주었다. 그리고, 판단이 변화된 경우에는 5명 모두 대부분 더 심한 쪽으로 바뀌었다. 전문가 5명의 1차와 2차 판단에 대한 상관분석을 실시한 결과, 상관계수는 .919로 매우 유의하였다($p < .01$).

<표 -5> 2차 판단 과정에서 발생한 전문가들의 판단 변화

		P1	P2	P3	P4	P5
변화 없음		30명 (83%)	30명 (83%)	33명 (92%)	28명 (78%)	30명 (83%)
변화	보다 약하게	1명	2명	1명	2명	1명
	보다 심하게	5명 (17%)	4명 (17%)	2명 (8%)	6명 (22%)	5명 (17%)

<표 -6> 전문가 판단과 P-FA 결과의 비교

	취학전		학령기		성인		총	
	1차	2차	1차	2차	1차	2차	1차	2차
P-FA결과 > 전문가판단	7 58.3%	6 50%	8 66.7%	9 75%	8 66.7%	8 66.7%	23 63.9%	23 63.9%
P-FA결과 = 전문가판단	1 8.3%	1 8.3%	4 33.3%	3 25%	2 16.7%	2 16.7%	7 19.4%	6 50%
P-FA결과 < 전문가판단	3 25.0%	3 25%	0 0%	0 0%	2 16.7%	2 16.7%	5 13.9%	7 58.3%
해당없음	1 8.3%	2 16.7%	0 0%	0 0%	0 0%	0 0%	1 2.8%	2 16.7%
합계	12	12	12	12	12	12	36	36

2. 전문가 판단과 P-FA 검사 결과 간의 비교

P-FA 결과와 전문가들의 1차 판단과 2차 판단을 비교한 결과는 <표 -6>과 같다. 1차 판단과 2차 판단 모두 피험자 36명 중 23명(63.9%)에 대해 P-FA 결과에 따른 중증도가 전문가의 판단보다 심한 편으로 나타났다.

3. 집단별 전문가 판단과 P-FA 결과 간의 관계

전문가의 판단과 P-FA 수치들 간의 상관분석을 실시하기 전에, 집단 혹은 과제에 따라 어떤 차이가 있는지 파악하기 위하여 세 집단 간에 비모수검정인 Kruskal-Wallis 검정을 실시하였다. 전문가 판단, P-FA 결과, 부수행동, 의사소통태도검사결과, 각 비유창성유형 출현비율, 정상비유창성 비율, 비정상비유창성 비율 등을 분석 변수로 포함시켰다. 그 결과, 대부분의 변수에서 유의한 차이가 없었으며 간투사 비율($\chi^2 = 13.810, p = .001$), ND 비율($\chi^2 = 10.947, p = .004$), 낱말이하의 반복(R2, 이하 R2)($\chi^2 = 7.359, p = .025$)에서 집단 간에 유의한 차이가 있었다($p < .05$). 사후검정 결과, 간투사와 ND는 성인집단에서, R2는 초등학생 집단에서 유의하게 많이 발생한 것으로 나타났다.

전문가의 판단이 특정한 비유창성 유형이나 부수행동 등 어떤 변수와 관련되어 있는지 살펴보기 위하여, 집단별로 나누어 전문가의 판단과 P-FA 측정치들 사이에 비모수 상관분석을 실시한 결과는 <표 -7>과 같다.

<표 7> 전문가 판단과 P-FA 측정치 간의 상관분석 결과

	취학전		초등학생		성인	
	판단 1	판단 2	판단 1	판단 2	판단 1	판단 2
판단1		0.985**		0.972**		0.904**
판단2	0.985**		0.972**		0.904**	
P-F결과	-0.091	-0.073	0.692*	0.696*	0.269	0.278
부수행동	0.169	0.303	0.581*	0.603*	0.525	0.517
의사소통태도			0.648*	0.646*	0.114	-0.028
H ^{a)} 비율	0.307	0.148	0.414	0.345	-0.213	-0.120
I ^{b)} 비율	-0.010	0.157	0.035	0.105	-0.039	0.041
UR ^{c)} 비율	-0.154	-0.161	0.225	0.141	-0.011	-0.052
R1 ^{d)} 비율	-0.317	-0.314	0.123	0.199	0.248	0.088
ND ^{e)} 비율	-0.257	-0.193	0.232	0.257	-0.110	-0.070
Ha ^{f)} 비율	0.070	0.169	-0.452	-0.455	0.267	0.060
Ia ^{g)} 비율	0.326	0.295	-0.484	-0.487	0.211	0.271
URa ^{h)} 비율	0.046	-0.039	-0.203	-0.356	0.289	0.178
R1a ⁱ⁾ 비율	0.259	0.245	0.399	0.258	0.575	0.544
R2 ^{j)} 비율	-0.096	-0.035	0.640*	0.631*	0.565	0.523
DP ^{k)} 비율	0.015	-0.007	0.656*	0.703*	0.253	0.096
AD ^{l)} 비율	-0.015	-0.003	0.727**	0.721**	0.548	0.332
ND와 AD 비율의 합	-0.043	-0.083	0.662*	0.699*	0.270	0.183
AD 가중점수 ^{m)}	-0.015	-0.003	0.727**	0.721**	0.548	0.332
과제점수	-0.015	0.021	0.688*	0.728**	0.362	0.226

** $p < .01$, * $p < .05$

a) H: 주저; b) I: 간투사; c) UR: 미완성 및 수정; d) R1: 낱말이상의 반복; e) ND: 정상비유창성; f) Ha: 비정상적 주저; g) Ia: 비정상적 간투사
h) URa: 비정상적 미완성 및 수정; i) R1a: 비정상적 낱말이상의 반복; j) R2: 낱말이하의 반복; k) DP: 비운율적 발성; l) AD: 비정상비유창성
m) AD 가중점수: AD 비율×1.5

세 집단 모두 판단1과 판단2 사이에 유의하게 높은 상관관계($p < .01$)가 있었다. 취학전 아동과 성인 집단에서는 전문가 판단과 P-FA 측정치들 사이에 유의한 상관성이 없었다. 초등학생 집단에서는, 판단1과 판단2 모두 P-FA 결과($r = .692, p = .013$), 부수행동($r = .581, p = .047$), 의사소통태도결과($r = .648, p = .023$), R2 비율($r = .640, p = .025$), 비운율적 발성(DP, 이하 DP) 비율($r = .656, p = .020$), 비정상비유창성(AD, 이하 AD) 비율($r = .727, p = .007$), ND와 AD의 비율 합($r = .662, p = .019$) 등과 유의한 상관관계가 있었다.

IV. 논의 및 결론

본 연구의 결과를 요약하면 다음과 같다. 첫째, 유창성장애 전문가들 사이에는 말더듬 정도의 주관적

판단의 일치율이 높은 편이었으며, 대상자에 대한 정보가 추가되더라도 판단이 변화하는 경우가 드물었다($r = .919$). 이는 숙련된 전문가들의 경우 많은 유창성장애 아동 및 성인을 평가하고 치료해왔기 때문에 발화상황에서 나타나는 구어와 비구어적인 정보를 종합적으로 직관적으로 파악하여 개인의 말더듬 정도를 판단하는 능력을 갖추고 있는 것으로 해석할 수 있다.

둘째, 전문가 판단과 P-FA 결과를 비교한 결과, 대체적으로 P-FA 결과에 따른 중증도가 전문가 판단보다 심한 편으로 나타났다(63.9%). 이러한 결과는 P-FA 개정을 위한 질적 연구(심현섭 · 신문자 · 이은주, 2009)에서 면담 참여자들이 염려하는 바와 일치하지만, 몇몇 선행연구와 불일치하는 결과인데 이는 본 연구에서 선택한 비유창성 분석방법과 선행연구에서 사용한 분석방법의 차이에서 비롯되었거나, 피험자의 특성으로 볼 수도 있다(O'Brian et al., 2004; Rousseau

et al., 2008).

마지막으로 전문가 판단과 P-FA 측정치들 간에는 취학전 아동과 성인의 경우에는 유의한 상관관계가 없었으며, 초등학교 집단인 경우에만 P-FA 결과, 의사소통태도검사결과, 부수행동점수, R2와 DP, AD 비율, ND와 AD의 비율 합, 과제 점수가 전문가 판단과 유의한 상관관계가 있는 것으로 나타났다. 본 연구에서는 간투사를 포함한 ND가 성인에서 더 많이 나타났고, R2는 초등학교에서 더 많이 나타나는 등, 비유창성 유형의 발생양상이 연령에 따라 차이가 있었다. 이러한 외현적 특성이 주관적 판단에 영향을 미칠 수 있다. 또한, 본 연구에 포함된 초등학교 집단의 P-FA 결과를 살펴보면 약함(3명)과 심함(9명)을 포함하고 있고 중간에 해당하는 피험자가 없는 등 다른 집단에 비해 분포가 고르지 않았다. 선행연구에 따르면 말더듬 중증도가 심해질수록 전문가들 간의 판단 일치율이 높아지는 경향을 보였는데(권보영, 2008), 본 연구에서도 초등학교 집단의 '심함'에 치우친 중증도 분포의 특성이 상관분석 결과에 영향을 미칠 수 있다.

선행연구에 따르면 전문가들의 비유창성 판단이나 말더듬정도 판단은 다른 집단에 비해 매우 일치율이 높은 것으로 나타났는데 본 연구의 결과는 이와 일치하는 것이다(Brundage et al., 2006; Ingham & Cordes, 1997; O'Brian et al., 2004; O'Brian, Packman & Onslow, 2004; Rousseau et al., 2008). 한편, 선행 연구에서는 평정척도와 객관적인 평가는 서로 대체할 수 있을 정도로 일치율이 높고 유의한 상관관계가 있었으나, 본 연구에서는 부분적인 일치만을 보여주었다(O'Brian et al., 2004; Rousseau et al., 2008). 이러한 결과는 선행연구에서는 말을 더듬은 음절의 비율을 사용하였으나 본 연구에서는 비유창성 유형과 빈도를 모두 고려하는 방법을 사용하여 분석방법에 차이가 있기 때문일 수 있다. 또한 선행연구에서는 다양한 연령집단을 대상으로 하지 않았으나, 본 연구에서는 취학전 아동, 초등학교, 성인과 같은 다양한 연령 집단을 포함시켰고, 집단별 피험자 숫자가 제한적이어서 예외적인 측정치의 영향을 크게 받을 수 있고, 말더듬 정도가 고루 포함되지 못하였기 때문에 이러한 결과를 일반화하여 해석하기는 어려울 것이다.

서론에서 언급하였듯이 말더듬 여부를 판단하는 것과 비유창성 유형을 판단하는 것은 별개의 문제이다(Cordes & Ingham, 1995, 1996; Ryan, 1974). 본

연구에 참여한 전문가들은 실험 후에 실시한 토의에서 전반적인 말더듬 정도를 평가할 때, 전체적인 비유창성 빈도 만큼이나 눈에 띄는 비유창성 유형, 그리고 순간적으로 동반되는 긴장과 같은 비유창성의 질적 측면 등을 함께 고려하게 된다고 하였다. P-FA와 같은 공식검사에서 실시하는 비유창성 유형의 판단이 말더듬의 이해와 치료에 실제적인 도움이 되도록 하기 위해서는, 비유창성 유형을 확인하는 작업을 통해 유창성장애 대상자의 '문제'를 희석시키는 결과를 낳지 않도록 말더듬 여부를 판단하는 데 영향을 미치는 변수를 보다 분명하게 하는 것이 필요할 것으로 보인다(Einarsdottir & Ingham, 2005).

본 연구의 제한점 및 임상적 의의는 다음과 같다. 본 연구에서는 주관적 판단과 P-FA 결과 간의 불일치가 무엇 때문에 발생하는지를 밝히지 못하였는데, 추후에 유창성장애 대상자들의 다양한 특성을 통제한 연구를 통해 전반적 판단과 객관적인 평가결과 간에 나타나는 차이가 무엇 때문인지를 명확하게 밝혀야 할 것이다.

또한, 본 연구에서는 P-FA에 포함된 과제 중 하나의 과제에 대해서만 전문가들이 판단하도록 하였는데, 추후 연구에서는 말더듬의 상황에 따른 가변성을 고려할 때 주관적 판단과 객관적 평가를 모두 같은 과제에 대하여 실시하고 비교할 필요가 있다. 연령에 따라 전반적 판단과 P-FA 결과 간의 상관관계가 다르게 나타난 점은 P-FA의 개정과정에서 충분히 고려하여 유창성장애의 연령과 과제에 따른 가변성을 반영할 수 있도록 보완하여야 할 것이다.

유창성장애 전문가들의 주관적 판단 일치율이 높기 때문에 공식검사를 실시하기 어렵거나 공식검사결과를 확인할 수 없을 때 제한적이지만 효과적인 정보 전달과 비교 평가가 가능할 것이다. 추후 유창성장애 청소년이나 성인을 대상으로 한 연구를 통해 이러한 주관적 평정척도의 신뢰도를 검증할 수 있다면, 전문가들 뿐만 아니라 유창성장애 대상자가 직접 자신의 말더듬 정도를 평정할 수 있도록 훈련하여 임상에서 효율적으로 사용할 수 있을 것이다.

본 연구를 통해 유창성장애 전문가들이 한 자리에 모여 유창성장애의 평가와 치료가 나아갈 방향에 대하여 함께 논의하고 다른 전문가들의 의견을 직접적으로 청취할 수 있었던 것도 본 연구의 큰 의의였다고 볼 수 있다.

참고 문헌

- 권보영(2008). 말더듬 중증도에 따른 평가자집단 간 말더듬 평가 비교. 이화여자대학교 대학원 석사학위논문.
- 심현섭 · 이은주(2002). SSI-3 하위영역들에 관한 평가자간 일치율 연구. 『언어청각장애연구』, 7(1), 130-154.
- 심현섭 · 신문자 · 이은주(2004). 『파라다이스-유창성검사』. 서울: 파라다이스복지재단.
- 심현섭 · 신문자 · 이은주(2009). 『파라다이스-유창성 검사도구의 개정판 연구원료보고서』. 서울: 파라다이스복지재단.
- Bothe, A. K. (2008). Identification of children's stuttered and nonstuttered speech by highly experienced judges: Binary judgments and comparisons with disfluency-types definitions. *Journal of Speech, Language, and Hearing Research*, 51, 867-878.
- Brundage, S. B., Bothe, A. K., Lengeling, A. N., & Evans, J. J. (2006). Comparing judgements of stuttering made by students, clinicians, and highly experienced judges. *Journal of Fluency Disorders*, 31, 271-283.
- Cordes, A. K. (1994). The reliability of observational data: I. Theories and methods for speech-language pathology. *Journal of Speech and Hearing Research*, 37, 264-278.
- Cordes, A. K. (2000). Individual and consensus judgments of disfluency types in the speech of persons who stutter. *Journal of Speech, Language, and Hearing Research*, 43(4), 941-964.
- Cordes, A. K., & Ingham, R. J. (1994). The reliability of observational data: II. Issues in the identification and measurement of stuttering events. *Journal of Speech and Hearing Research*, 37, 279-294.
- Cordes, A. K., & Ingham, R. J. (1995). Stuttering includes both within-word and between-word disfluencies. *Journal of Speech and Hearing Research*, 38, 382-386.
- Cordes, A. K., & Ingham, R. J. (1996). Disfluency types and stuttering measurement: A necessary connection? *Journal of Speech and Hearing Research*, 39, 404-405.
- Curlee, R. F. (1981). Observer agreement on disfluency and stuttering. *Journal of Speech and Hearing Research*, 24, 595-600.
- Einarsdottir, J., & Ingham, R. J. (2005). Have disfluency-type measures contributed to the understanding and treatment of developmental stuttering? *American Journal of Speech-Language Pathology*, 14, 260-273.
- Howell, P., Sackin, S., & Glenn, K. (1997). Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: II. ANN recognition of repetitions and prolongations with supplied word segment markers. *Journal of Speech, Language, and Hearing Research*, 40(5), 1085-1096.
- Howell, P., Sackin, S., & Glenn, K. (1998). Assessment procedures for locating stuttered events. In E. C. Healey & H. F. Peters (Eds.), *Proceedings of the second world congress on Fluency Disorders* (212-218). Nijmegen, The Netherlands: University Press Nijmegen.
- Ingham, R. J., & Cordes, A. K. (1997). Identifying the authoritative judgements of stuttering: Comparisons of self-judgments and observer judgments. *Journal of Speech, Language, and Hearing Research*, 40, 581-594.
- Ingham, R. J., Cordes, A. K., & Gow, M. L. (1993). Time-interval measurement of stuttering: Modifying interjudge agreement. *Journal of Speech, Language, and Hearing Research*, 36, 503-515.
- Ingham, R. J., Cordes, A. K., Ingham, J. C., & Gow, M. L. (1995). Identifying the onset and offset of stuttering events. *Journal of Speech and Hearing Research*, 38, 315-326.
- Manning, W. H. (2001). *Clinical decision-making in fluency disorders* (2nd ed.). Vancouver: Singular & Thomson Learning.
- McDonald, J. D., & Martin, R. R. (1973). Stuttering and disfluency as two reliable and unambiguous response classes. *Journal of Speech and Hearing Research*, 16, 691-699.
- O'Brian, S., Packman, A., & Onslow, M. (2004). Self rating of stuttering severity as a clinical tool. *American Journal of Speech-Language Pathology*, 13, 219-226.
- O'Brian, S., Packman, A., Onslow, M., & O'Brian, N. (2004). Measurement of stuttering in adults: Comparison of stuttering-rate and severity-scaling methods. *Journal of Speech, Language, and Hearing Research*, 47, 1081-1087.
- Onslow, M., & O'Brian, S. (1998). Reliability of clinician's judgments about prolonged-speech targets. *Journal of Speech, Language, and Hearing Research*, 41, 969-975.
- Onslow, M., Gardner, K., Bryant, K. M., Stuckings, C. L., & Knight, T. (1992). Stuttered and normal speech events in early childhood: The validity of a behavioral data language. *Journal of Speech and Hearing Research*, 35, 79-87.
- Riley, G. (1994). *Stuttering severity instrument for children and adults* (3rd ed.). Austin, TX: Pro-ed.
- Rousseau, I., Onslow, M., Packman, A., & Jones, M. (2008). Comparisons of audio and audiovisual measure of stuttering frequency and severity in preschool-age children. *American Journal of Speech-Language Pathology*, 17, 173-178.
- Ryan, B. P. (1974). *Programmed therapy for stuttering in children and adults*. Springfield, IL: CC Thomas.
- Teesson, K., Packman, A., & Onslow, M. (2003). The lidcombe behavioral data language of stuttering. *Journal of Speech and Hearing Research*, 46, 1009-1015.
- Yairi, E. (1996). Applications of disfluencies in measurements of stuttering. *Journal of Speech, Language, and Hearing Research*, 39, 402-404.
- Yaruss, J. S. (1998). Real-Time analysis of speech fluency:

- Procedures and reliability training. *American Journal of Speech-Language Pathology*, 7, 25-37.
- Young, M. A. (1969). Observer agreement: Cumulative effects of repeated ratings of the same samples and of knowledge of group results. *Journal of Speech and Hearing Research*, 12, 144-155.
- Young, M. A. (1975). Observer agreement for marking moments of stuttering. *Journal of Speech and Hearing Research*, 18, 530-540.
- Young, M. A., & Downs, T. D. (1968). Testing the significance of the agreement among observers. *Journal of Speech and Hearing Research*, 11, 5-17.

<부록 - 1> 전문가 5명의 1차 및 2차 평가 결과

		P1			P2			P3			P4			P5		
		1st	2nd	변화 ^{a)}	1st	2nd	변화	1st	2nd	변화	1st	2nd	변화	1st	2nd	변화
취학 전 아동	1	1	1	=	1	1	=	1	1	=	1	2	↑	2	2	=
	2	1	1	=	1	1	=	0	0	=	0	1	↑	1	1	=
	3	2	3	↑	2	3	=	3	3	=	3	3	=	3	2	↓
	4	0	0	=	0	0	=	1	1	=	0	0	=	0	0	=
	5	1	1	=	2	2	=	1	1	=	0	1	↑	1	2	↑
	6	2	3	↑	3	3	=	3	3	=	3	3	=	3	3	=
	7	1	1	=	0	1	↑	2	1	↓	0	1	↑	1	1	=
	8	2	3	↑	2	2	=	1	1	=	2	2	=	2	3	↑
	9	1	0	↓	1	1	=	1	1	=	1	1	=	1	1	=
	10	3	3	=	3	3	=	1	1	=	2	2	=	3	3	=
	11	1	1	=	1	1	=	1	1	=	1	1	=	1	1	=
	12	1	1	=	1	1	=	1	2	↑	1	2	↑	2	3	↑
초등 학생	13	2	2	=	2	2	=	2	2	=	3	3	=	2	2	=
	14	3	3	=	3	3	=	2	2	=	2	2	=	3	3	=
	15	3	3	=	3	2	↓	2	2	=	2	2	=	3	3	=
	16	2	2	=	2	2	=	1	1	=	3	3	=	2	2	=
	17	2	2	=	3	3	=	2	2	=	2	2	=	2	2	=
	18	2	2	=	3	3	=	3	3	=	3	3	=	3	3	=
	19	1	1	=	1	1	=	0	0	=	2	1	↓	1	1	=
	20	1	1	=	1	1	=	1	1	=	1	1	=	1	2	↑
	21	1	1	=	1	1	=	1	1	=	1	1	=	1	1	=
	22	1	1	=	1	1	=	1	1	=	2	1	↓	1	1	=
23	0	0	=	0	0	=	0	0	=	0	0	=	0	0	=	
24	0	0	=	0	0	=	0	0	=	0	0	=	0	0	=	
성인	25	0	1	↑	1	0	↓	1	1	=	1	1	=	1	1	=
	26	1	1	=	0	1	↑	0	0	=	1	1	=	0	0	=
	27	0	0	=	0	0	=	0	0	=	0	0	=	0	0	=
	28	0	1	↑	1	1	=	0	0	=	1	1	=	0	0	=
	29	1	1	=	1	1	=	1	1	=	1	1	=	1	1	=
	30	1	1	=	1	1	=	1	1	=	1	1	=	1	1	=
	31	1	1	=	1	1	=	1	1	=	1	1	=	1	1	=
	32	2	2	=	2	2	=	2	2	=	3	3	=	2	3	↑
	33	3	3	=	2	3	↑	1	2	↑	3	3	=	3	3	=
	34	3	3	=	2	2	=	1	1	=	1	1	=	1	1	=
35	3	3	=	3	3	=	2	2	=	2	3	↑	3	3	=	
36	3	3	=	3	3	=	2	2	=	2	2	=	2	2	=	

a) '='은 변화없음을, '↓'은 더 약한 쪽으로의 변화를, '↑'은 더 심한 쪽으로의 변화를 의미함.

ABSTRACT

Comparing the Clinical Judgments of Stuttering Specialists to Stuttering Severity Rated by the Paradise-Fluency Assessment (P-FA)

Eun-Ju Lee^a · Jung-Hyun Park^b · Moon-Ja Shin^c · Hyun-Sub Sim^{b,d,s}

^a Department of Special Education, DanKook University, Seoul, Korea

^b Interdisciplinary Program of Communication Disorders, Ewha Womans University, Seoul, Korea

^c Department of Speech Therapy, Chosun University, Gwangju, Korea

^d Department of Special Education, Ewha Womans University, Seoul, Korea

^s Correspondence to

Prof. Hyun-Sub Sim, PhD,
Interdisciplinary Program of
Communication Disorders and
Department of Special
Education, Ewha Womans
University, 11-1
Daehyun-dong,
Seodaemun-gu, Seoul, Korea
e-mail: simhs@ewha.ac.kr
tel.: +82 2 3277 2120

Background & Objectives: The purpose of this study was to compare fluency specialists' subjective clinical judgments of stuttering with objective severity, as rated by the Paradise-Fluency Assessment (P-FA). We aimed to provide a reference that would be of use in revising the P-FA. **Methods:** Five fluency specialists were asked to judge the severity of stuttering of 36 people who stutter (i.e., 12 preschoolers, 12 school-age children, and 12 adults) on a four-point scale after watching video clips recording the participants' utterances elicited by the P-FA items. Additionally, two different researchers rated the same 36 participants' speech samples based on the P-FA. **Results:** Findings showed high inter-rater agreement in clinical judgments of stuttering severity. Similarly, a high degree of intra-rater agreement was observed in clinical judgments across the two evaluations. However, the judges tended to report lower stuttering severity on the four-point scale compared to severity ratings calculated using the P-FA. No statistically significant relationship was found between the clinical judgments and the P-FA scores in the case of preschool children and adults. However, there was a statistically significant relationship for school-age children. **Discussion & Conclusion:** These findings suggest guidelines for the revision of the P-FA. The data also provide a springboard for further discussion of stuttering assessment. (*Korean Journal of Communication Disorders* 2009;14:200-211)

Key Words: Paradise-Fluency Assessment (P-FA), clinical judgments, fluency specialist

REFERENCES

- Bothe, A. K. (2008). Identification of children's stuttered and nonstuttered speech by highly experienced judges: Binary judgments and comparisons with disfluency-types definitions. *Journal of Speech, Language, and Hearing Research, 51*, 867-878.
- Brundage, S. B., Bothe, A. K., Lengeling, A. N., & Evans, J. J. (2006). Comparing judgements of stuttering made by students, clinicians, and highly experienced judges. *Journal of Fluency Disorders, 31*, 271-283.
- Cordes, A. K. (1994). The reliability of observational data: I. Theories and methods for speech-language pathology. *Journal of Speech and Hearing Research, 37*, 264-278.
- Cordes, A. K. (2000). Individual and consensus judgments of disfluency types in the speech of persons who stutter. *Journal of Speech, Language, and Hearing Research, 43*(4), 941-964.
- Cordes, A. K., & Ingham, R. J. (1994). The reliability of observational data: II. Issues in the identification and measurement of stuttering events. *Journal of Speech and Hearing Research, 37*, 279-294.
- Cordes, A. K., & Ingham, R. J. (1995). Stuttering includes both within-word and between-word disfluencies. *Journal of Speech and Hearing Research, 38*, 382-386.
- Cordes, A. K., & Ingham, R. J. (1996). Disfluency types and stuttering measurement: A necessary connection? *Journal of Speech and Hearing Research, 39*, 404-405.

* This work was supported by the Paradise Welfare Foundation (2008).

■ Received April 17, 2009 ■ Final revision received June 22, 2009 ■ Accepted June 23, 2009.

© 2009 The Korean Academy of Speech-Language Pathology and Audiology <http://www.kasa1986.or.kr>

- Curlee, R. F. (1981). Observer agreement on disfluency and stuttering. *Journal of Speech and Hearing Research, 24*, 595-600.
- Einarsdottir, J., & Ingham, R. J. (2005). Have disfluency-type measures contributed to the understanding and treatment of developmental stuttering? *American Journal of Speech-Language Pathology, 14*, 260-273.
- Howell, P., Sackin, S., & Glenn, K. (1997). Development of a two-stage procedure for the automatic recognition of dysfluencies in the speech of children who stutter: II. ANN recognition of repetitions and prolongations with supplied word segment markers. *Journal of Speech, Language, and Hearing Research, 40*(5), 1085-1096.
- Howell, P., Sackin, S., & Glenn, K. (1998). Assessment procedures for locating stuttered events. In E. C. Healey & H. F. Peters (Eds.), *Proceedings of the second world congress on Fluency Disorders* (212-218). Nijmegen, The Netherlands: University Press Nijmegen.
- Ingham, R. J., & Cordes, A. K. (1997). Identifying the authoritative judgements of stuttering: Comparisons of self-judgments and observer judgments. *Journal of Speech, Language, and Hearing Research, 40*, 581-594.
- Ingham, R. J., Cordes, A. K., & Gow, M. L. (1993). Time-interval measurement of stuttering: Modifying interjudge agreement. *Journal of Speech, Language, and Hearing Research, 36*, 503-515.
- Ingham, R. J., Cordes, A. K., Ingham, J. C., & Gow, M. L. (1995). Identifying the onset and offset of stuttering events. *Journal of Speech and Hearing Research, 38*, 315-326.
- Kwon, B-Y. (2008). *Evaluations of stuttered speech samples with different stuttering severity: Speech language pathologist group comparisons*. Unpublished master's thesis. Ewha Womans University, Seoul.
- Manning, W. H. (2001). *Clinical decision-making in fluency disorders* (2nd ed.). Vancouver: Singular & Thomson Learning.
- McDonald, J. D., & Martin, R. R. (1973). Stuttering and disfluency as two reliable and unambiguous response classes. *Journal of Speech and Hearing Research, 16*, 691-699.
- O'Brian, S., Packman, A., & Onslow, M. (2004). Self rating of stuttering severity as a clinical tool. *American Journal of Speech-Language Pathology, 13*, 219-226.
- O'Brian, S., Packman, A., Onslow, M., & O'Brian, N. (2004). Measurement of stuttering in adults: Comparison of stuttering-rate and severity-scaling methods. *Journal of Speech, Language, and Hearing Research, 47*, 1081-1087.
- Onslow, M., & O'Brian, S. (1998). Reliability of clinician's judgments about prolonged-speech targets. *Journal of Speech, Language, and Hearing Research, 41*, 969-975.
- Onslow, M., Gardner, K., Bryant, K. M., Stuckings, C. L., & Knight, T. (1992). Stuttered and normal speech events in early childhood: The validity of a behavioral data language. *Journal of Speech and Hearing Research, 35*, 79-87.
- Riley, G. (1994). *Stuttering severity instrument for children and adults* (3rd ed.). Austin, TX: Pro-ed.
- Rousseau, I., Onslow, M., Packman, A., & Jones, M. (2008). Comparisons of audio and audiovisual measure of stuttering frequency and severity in preschool-age children. *American Journal of Speech-Language Pathology, 17*, 173-178.
- Ryan, B. P. (1974). *Programmed therapy for stuttering in children and adults*. Springfield, IL: CC Thomas.
- Sim, H. S., & Lee, E. J. (2002). An interjudge reliability study on the subtypes of SSI-3. *Korean Journal of Communication Disorders, 7*(1), 130-154.
- Sim, H. S., Shin, M. J., & Lee, E. J. (2004). *Paradise-Fluency Assessment*. Seoul: Paradise Welfare Foundation.
- Sim, H. S., Shin, M. J., & Lee, E. J. (2009). *Revised version of Paradise-Fluency Assessment* (Final Report). Seoul: Paradise Welfare Foundation.
- Teeson, K., Packman, A., & Onslow, M. (2003). The lidcombe behavioral data language of stuttering. *Journal of Speech and Hearing Research, 46*, 1009-1015.
- Yairi, E. (1996). Applications of disfluencies in measurements of stuttering. *Journal of Speech, Language, and Hearing Research, 39*, 402-404.
- Yaruss, J. S. (1998). Real-Time analysis of speech fluency: Procedures and reliability training. *American Journal of Speech-Language Pathology, 7*, 25-37.
- Young, M. A. (1969). Observer agreement: Cumulative effects of repeated ratings of the same samples and of knowledge of group results. *Journal of Speech and Hearing Research, 12*, 144-155.
- Young, M. A. (1975). Observer agreement for marking moments of stuttering. *Journal of Speech and Hearing Research, 18*, 530-540.
- Young, M. A., & Downs, T. D. (1968). Testing the significance of the agreement among observers. *Journal of Speech and Hearing Research, 11*, 5-17.