

## 교육수화 통역의 평가자간 평가기준 조율 가능성에 대한 사례분석

민 은 주\*

나사렛대학교 수화통역학과

---

### 《 요 약 》

---

이 연구에서는 N대학교 수화통역과의 '교육통역' 수업을 수강한 학생들의 통역 능력에 대한 평가 결과를 비교하고 평가기준 조율이 가능한지를 알아보기 위해 수화통역 담당 교육자들에게 학생 평가를 의뢰하여 그 사례를 분석하였다.

2013년 1학기 교육통역 강의를 수강한 학생들 중 학기말 평가에서 우수한 점수를 획득한 5명의 통역 결과물에 대하여 4명의 평가자가 순위를 매겼고 평가의 근거가 되는 세부 평가 기준을 자유롭게 서술하도록 하였다. 그 결과 평가자들이 평가에 있어서 비슷한 기준을 가지고 있음이 드러났으나 그 적용에 있어서는 서로 다른 우선순위를 부여함으로써 평가 결과에 큰 차이를 보였다. 이러한 평가자 간의 차이를 극복하기 위한 조율이 가능한지를 알아보기 위하여 4명의 평가자 중 가장 큰 차이를 보였던 2명의 평가자에게 조율과정을 거치도록 하였다. 그 결과 평가기준의 조율이 가능하다는 것을 우선 발견했으며, 조율된 기준으로 수강 학생 전원의 기말시험을 평가토록 한 결과 66.7%의 일치율을 보였다. 마지막으로, 이전 평가에 관여하지 않았던 제3의 평가자를 투입하여 수강 학생 전원을 평가토록 한 결과 앞서 조율된 평가자들의 결과와 일치도가 확연히 낮음을 알 수 있었다. 이 결과는 평가기준의 조율이 평가의 객관성을 확보 할 수 있음을 입증하는 것으로 예비통역사 교육과정에서 평가기준 마련이 필요함을 시사하고 있다.

---

주제어 : 한국수화, 교육통역, 평가비교

---

\* 교신저자(unjumin@hanmail.net)

## 1. 서론

### 1. 연구의 필요성

2006년 제 1회 국가공인 수화통역사들을 배출한 이후 2011년 12월 말 기준으로 국내의 수화통역 자격증 소지자는 975명(황주희, 2013) 달한다. 수화통역의 수요가 늘어나고 소비자의 요구가 다양해지는 것과 비례해 수화통역사의 양성도 꾸준히 증가해왔다. 통역 소비자와 공급자의 양적 확장과 함께 수화통역사 양성과정에서의 교육의 질과 통역 평가기준의 질적 제고에 대한 요구도 높아지고 있다.

장애인관련 법이 제·개정 되는 등 어느 때보다 수화통역에 대한 시대적 요구가 높아짐에 따라 수화관련 연구도 다양한 주제로 진행되고 있으나 대부분이 수화통역사의 직무, 역할, 만족도나 인식조사, 통역 서비스 지원 실태 및 체계, 제도개선 등 주로 수화 환경과 관련된 연구이고 수화통역의 질적 개선에 영향을 미칠 수 있는 통역 기술이나 전략, 평가와 관련된 연구의 거의 없는 실정이다(민은주, 2013).

국가공인 수화통역사 자격시험에서는 1차 필기시험에 합격한 수험생을 대상으로 2차 실기시험을 치른다. 수화를 한국어로 번역하는 필기통역, 구어를 수화로 통역하는 수화통역, 수화를 구어로 통역하는 음성통역이 실기시험 과목이다. 그러나 오랫동안 수화통역사 자격시험을 진행해왔지만 평가기준을 설정하기 위한 심도 있는 연구는 이루어지지 못했다. 이에 대한 여러 가지 원인이 있겠으나 이 분야의 연구자의 부족, 국가차원의 연구 지원 미흡, 나아가 수화언어법 제정의 지연으로 인한 수화의 언어로서의 법적 지위 미확보 등이 직간접적으로 영향을 미쳤을 것이다.

수화통역 평가 기준에 대한 연구가 미흡한 가운데 일선의 통역사 양성과정에서도 교육과정 실행 단계의 학업수행을 평가하는 형성평가와 달성목표를 평가하는 총괄평가 과정에서도 체계적인 기준 적용이 어려운 실정이다. 통역분야를 비롯하여 모든 양성과정에는 필히 이러한 방식의 평가가 강조되기 마련인데 이는 평가가 교육활동의 핵심요소이기 때문이다(정철자, 2012).

평가에 있어서 물론 전문가의 판단은 존중되어야 하나 시험 설계와 평가를 위한 체계적인 접근방법이 없을 경우 평가의 타당도와 신뢰도가 위협받을 수 있을 것이다(Sawyer, 2001). 모든 평가 대상은 동일한 평가방식과 투명하고 객관적인 평가 기준, 공정한 절차 아래에서 평가되어야 하기에(이연향, 2009) 수화 통역에 있어서 평가 기준에 대한 논의는 속히 진행될 필요가 있을 것이다.

최근 들어 최상배(2010)와 민은주(2013) 등이 수화동시통역의 평가와 관련하여 통역의 정확성, 통역 기술 및 전략, 통역률, 시차(time-lag) 등 통역의 질에 영향을 미치는 다양한 요소들을 분석함으로써 이 분야의 통역평가 연구와 관련하여 기여하

고 있다. 그러나 평가의 기준으로 삼을 수 있는 평가체계 개발에는 아직 이르지 못하고 있다. 따라서 이 연구에서는 수화 통역 평가기준 마련을 위한 기초 연구의 형식으로, 통역 평가자 간의 평가기준이 개인별로 얼마나 차이가 나는지, 통역자 간 평가기준의 조율은 가능한 것인지, 조율된 평가기준과 그렇지 않은 평가 사이에 어떤 의미 있는 차이가 있는지 등을 알아보려고 한다. 이는 체계적인 평가기준 마련의 당위를 강조하기 위한 논리적 기초 실험 절차이며, 완성된 평가기준 제시를 위한 연구는 아니다. 이를 위해서는 관련 분야 전문가 집단의 공동연구 과정이 필요할 것이다.

## 2. 연구 문제

수화통역 교육상황에서 교수자 간 평가 편차를 극복하기 위해 평가기준 조율을 실시하고 극복방안을 모색한 후, 별도의 평가 작업을 실시해 조율된 기준의 효과를 모색하기 위하여 설정한 연구 문제는 다음과 같다.

첫째, 평가자별 평가의 편차를 분석하고 평가 근거를 비교한다.

둘째, 평가자간 평가기준 조율 가능성을 분석한다.

셋째, 조율된 평가기준 적용 결과와 제3의 평가자 투입 결과를 비교 분석한다.

## II. 연구 방법

### 1. 연구 대상

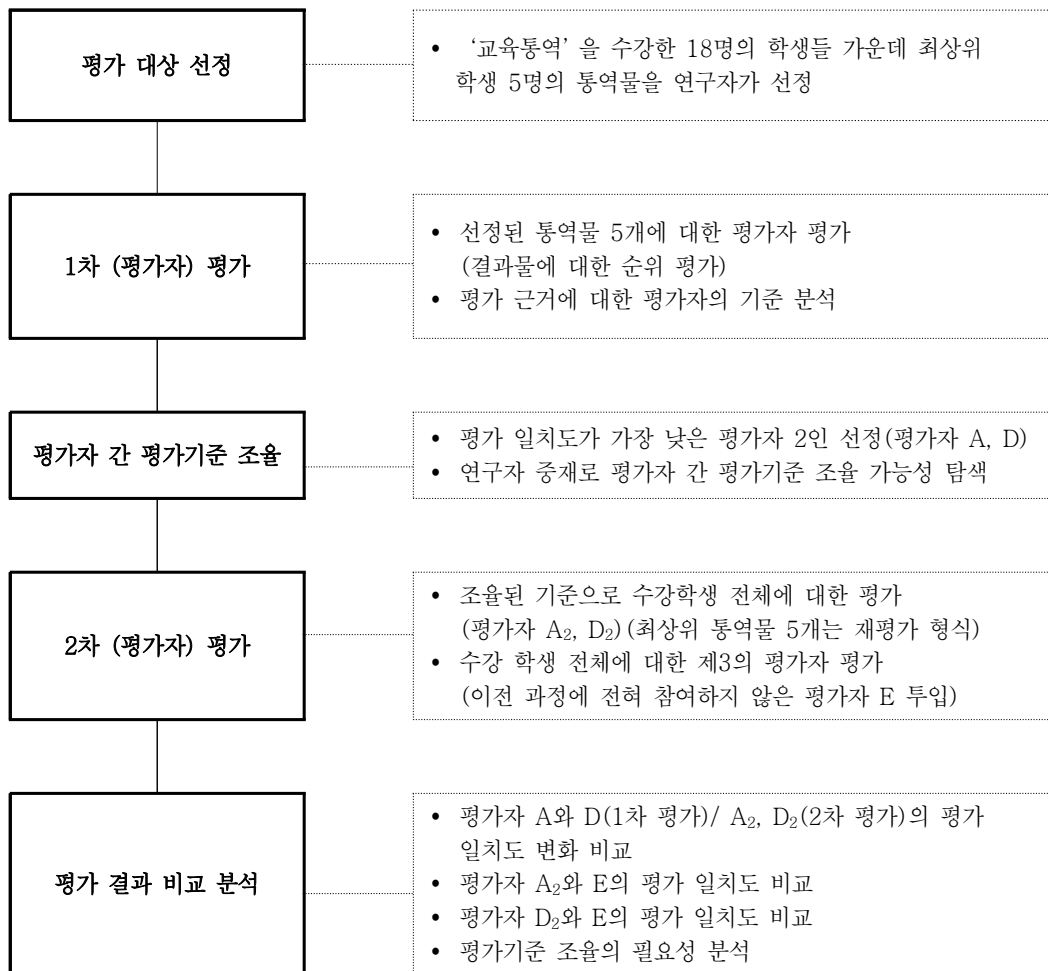
이 연구의 대상은 2013년 1학기 N대학의 ‘교육통역’ 수업을 수강한 18명의 학습자들이 기말시험으로 치른 통역물이며 이들 가운데 최상위 성적을 획득한 5명의 통역물이 1차 평가를 위한 연구 대상으로 선별되었다. ‘교육통역’은 수화통역 실기 수업중에서 3, 4학년 대상의 특화 심화과정이자 교육현장에서의 통역을 대비하기 위한 과정이다. 수업에서 다루는 내용은 대학 강의 가운데 농인 대학생들이 많이 수강하는 과목들과 농인 직업교육분야의 용어들을 학습하고 다양한 통역시연을 통해 졸업 후 현장에서 통역사로서 활동하기 위한 제반사항들을 학습한다. 상위 5명의 통역물을 1차 평가의 대상으로 삼은 이유는 통역 평가 시 개인 간 차가 심하지 않은 비교적 균질성이 높은 집단을 선별할 필요가 있었기 때문이다.

이 연구에 참여한 평가자 5명의 수화통역 평가자의 평가 결과가 이 연구의 또 다른 대상이다. 연구에 참여한 평가자들은 모두 국가공인 수화통역사 자격증을 소지

하였으며 평가자B를 제외한 모든 평가자들은 교육현장에서 3년 이상의 강의 경력 및 학생 평가 경험을 가진 자들이다. 1차 평가에 참여한 평가자는 평가자 A, B, C, D이며, 2차 평가에 참여한 평가자는 평가자 A, D, E이다. 2차 평가에도 참여한 평가자 A, D는 A<sub>2</sub>, D<sub>2</sub>라 칭한다. 평가자 E는 이전 평가에 관여하지 않고 2차 평가에만 참여한 평가자이다.

## 2. 연구 절차

이 연구의 진행 절차는 <그림 1>에 제시된 바와 같다.



<그림 1> 연구 진행 절차

이 연구의 전체 진행 절차는 평가 대상 선정, 1차 평가자 평가, 평가자 평가 결과 분석 및 평가기준 조율, 2차 평가자 평가, 평가 결과 비교 분석의 과정으로 이루어졌으며 이에 대하여 상술하면 다음과 같다.

첫째, 평가 대상 선정 과정에서는 2013년 1학기 ‘교육통역’ 강의를 수강한 학생 18명의 기말시험 결과 가운데 최상위 성적을 얻은 5명의 통역물을 선별하였다. 기말시험에서 학생들은 2개의 수화통역 과제를 받았다. 하나는 요시모토 요시오가 쓴 ‘스타벅스에서는 그란데를 사라’ 중 일부분을 발췌하여 1분 40초 길이의 음성 파일로 전환한 것이며, 다른 하나는 노무현 전대통령의 ‘독도’ 연설문에서 발췌한 2분 분량의 텍스트이다. 학생 18명이 기말시험으로 치른 통역물은 모두 녹화되었고 총 녹화시간은 66분이다( $\{스타벅스\ 1분\ 40초 \times 18\} + \{독도\ 2분 \times 18\} = 66분$ ). 예비 통역사들의 시험 결과에 대하여 연구자는 대학의 학점 평정 기준에 따라 평가 점수와 학점을 부여하였고 이들 중 A학점을 받아 균질성이 높은 5명의 통역물을 평가대상으로 삼았다.

둘째, 1차 평가자 평가 과정은 앞서 선별된 다섯 학생의 통역물에 대해 4명의 평가자가 실시한 평가이다. 4명의 평가자는 통역물에 대해 점수가 아닌 순위로 평가하였고 평가의 근거, 또는 평가기준을 서술하여 제시하였다. 평가를 점수가 아닌 순위로 평가한 이유는 구체적인 평가 점수 부여 기준이 평가자 간에 크게 차이가 날 수 있고 이 평가의 목적이 평가자 간 평가 결과의 비교 가능성을 확인하는 것이니 만큼 순위표시가 더 효율적일 것이라 판단하였기 때문이다(Wu, 2010). 모든 평가자는 이메일로 학생의 통역물 동영상 자료를 받아 평가를 실시하였고 그 결과에 대한 근거를 서술하여 다시 연구자에게 보내 주었다. 이 때 평가자 간에 소통은 없었고 학생들에 대한 정보도 제공하지 않았으며 모든 통역물은 이름 없이 번호로만 처리되었다.

셋째, 평가자 4명이 평가 근거로 제시한 서술식 평가기준을 분석하여 평가자들이 어떤 기준으로 통역을 평가하는지 파악하였다. 평가자들의 순위 평가 결과를 비교 분석하여 4명의 평가자 가운데 평가 편차가 가장 심한 2명의 평가자(A, D)를 선정하여 평가 조율 가능성을 탐색하였다. 편차가 가장 심한 2명의 평가자를 선정한 이유는 편차 극복 가능성을 탐색하기 위해서는 극복 과제가 가장 큰 사례를 선별하여 조율 가능성을 탐색하는 것이 연구의 목적에 부합할 것이라 판단했기 때문이다. 이들 두 평가자는 연구자의 중재로 직접 만나 자신들의 평가기준을 설명하고 연구자와 함께 충분한 논의 과정을 거쳐 연구자의 주도로 공통된 평가기준을 마련하였다. 그러나 이러한 과정이 연구자의 의도대로 항상 합의가 도출 될 수 있는 것은 아니라는 점이 향후 조절해야 할 과제로 남는다. 조율이라는 과정이 기계적 과정이 아니고 여러 사람의 주관이 충돌하여 최소한 조율자들 내에서는 하나의 객관을 만들어내는 과정이기 때문에 개인 간 또는 개인 내적 변인이 작용할 가능성이 매우 크다는 점이 고려되어야 한다.

넷째, 2차 평가자 평가에서는 위에서 조율된 새로운 평가기준으로 평가자 2인(이하 A<sub>2</sub>, D<sub>2</sub>라 칭한다.)에게 수강학생 전체의 통역물을 각각 평가토록 하였다. 이 가운데 최상위 5개의 평가물은 1차 평가(조율 전)와 2차 평가(조율 후)의 변화 추이가 분석되었다. 그리고 이전 평가 과정에 전혀 개입되지 않은 제3의 평가자를 투입하여 전체 통역물 및 최상위 통역물 5개를 평가토록 하였다. 이 과정에서 다시 한번 평가기준의 조율과정을 거친 평가와 그렇지 않은 평가 사이의 평가 편차를 확인하는 과정을 가졌다.

다섯째, 평가 결과에 대한 비교 분석에서는 평가자 A와 D의 1차 평가와 동일한 평가자인 2차 평가인 A<sub>2</sub>, D<sub>2</sub>를 비교함으로써 평가기준 조율이 평가에 미치는 영향을 조사하였다. 그리고 제3의 평가자 E의 평가 결과를 A<sub>2</sub>, D<sub>2</sub>와 비교함으로써 평가기준이 조율되지 않았을 때의 평가 편차를 다시 한 번 확인하였다. 마지막으로 전체 학생의 평가물에 대한 평가자 A<sub>2</sub>와 E, 평가자 D<sub>2</sub>와 E를 비교함으로써 평가기준 조율이 평가에 미치는 영향을 검증하였다.

### III. 연구 결과

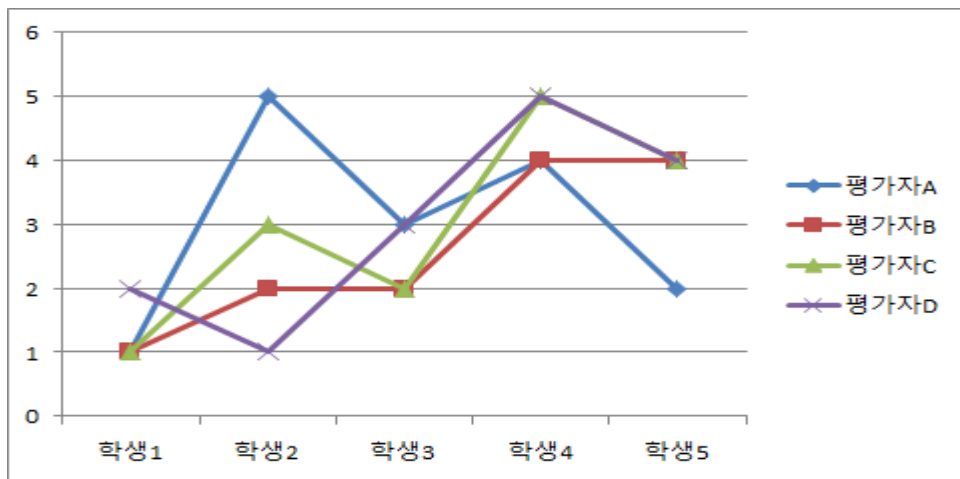
#### 1. 평가자별 평가기준 비교(1차 평가자 평가)

5명의 최상위 학생의 통역물을 4명의 평가자가 순위별로 평가한 결과는 <표 1>과 같다. 학생들은 수화동시통역 능력 면에서 개인적인 차이가 있을 수 있겠으나 동시통역 실기 평가 시 차등화가 매우 어려운, 균질성이 매우 높은 집단이었다. 따라서 평가자들이 이들 학생들의 순위를 결정하기 위해서는 섬세한 평가기준을 가지고 있어야 했다. 1차 평가자 평가의 주된 목적은 평가자의 평가기준을 알아보기 위함이며 동시에 상대적 순위 부여를 통해 평가의 편차를 가장 심하게 보이는 대표적인 평가자 2명을 선정하기 위함이므로 구체적인 점수 부여가 필요치 않았다.

<표 1> 통역사례 평가 순위 표

|     | 평가자A | 평가자B | 평가자C | 평가자D |
|-----|------|------|------|------|
| 학생1 | 1    | 1    | 1    | 2    |
| 학생2 | 5    | 2    | 3    | 1    |
| 학생3 | 3    | 2    | 2    | 3    |
| 학생4 | 4    | 4    | 5    | 5    |
| 학생5 | 2    | 4    | 4    | 4    |

평가자B와 평가자 C의 평가결과가 가장 높은 비교가능성을 보였다. 두 평가자 모두 학생1, 학생3 그리고 학생5에 대해 동일한 평가결과를 보여 60%의 비교가능성을 나타내었고, 평가자 A와 평가자 B는 학생 1과 학생 4의 순위가 같아 비교가능성 40%를 보였다. 학생 1에 대한 평가는 평가자 D가 2를 부여한 것을 제외하고 모든 평가자가 1위를 부여해 거의 동일한 평가결과를 보였고, 학생 3과 학생 4의 경우에도 각각 2-3위, 4-5의 순위를 부여함으로써 비슷한 평가 결과를 보였다. 평가자간 순위 결과가 2등급이상 차이가 나는 경우는 학생 2와 학생 5인데, 평가자 A는 학생 2에게 5위를, 평가자 D는 1위를 부여함으로써 가장 큰 평가 간 차이를 보였으며, 학생 5에 대해서도 평가자 A가 2위를, 평가자B, C, D는 4위를 부여함으로써 2등급 이상의 차이를 보였다. 평가 간의 가장 큰 차이를 보이는 두 평가는 A와 C, 또는 A와 D로 이들은 학생 1의 평가를 제외하면 비교가능성은 0% 수준이다. <표 2>을 보다 쉽게 비교할 수 있도록 한 <그림 2>의 그래프를 보면 학생 2를 기준으로 평가자 A와 D사이의 차이(4등급)를 확연히 볼 수 있다. 그러나 나머지 평가자들은 거의 평가등급이 일치하거나 한 등급 간의 차이만을 보인다.



<그림 2> 통역사례 평가 순위표

<표 2>에 나타난 평가자들의 평가기준을 살펴보면, 모든 평가자들이 공통적으로 통역의 정확성(accuracy)과 원문에 대한 충실성(fidelity)을 중요한 기준으로 삼고 있는 것을 알 수 있다. 이는 Kurz(2001)가 지적한 전문가 집단의 통역 평가에 대한 기준과도 일치하는 것이다.

<표 2>

평가자별 평가기준 요약

| 평가자  | 평가기준                                    |                              |                |
|------|---|------------------------------|----------------|
| 평가자A | 1. 원문의 내용 이해도<br>4. 수화 문장 구성력           | 2. 통역의 정확성<br>5. 전체적인 통역 분위기 | 3. 수화 어휘 표현력   |
| 평가자B | 1. 통역의 오역<br>4. 수화 구사력(강약, 비수지기호, 공간활용) | 2. 내용의 생략                    | 3. 통역시 표정과 분위기 |
| 평가자C | 1. 원문의 내용 전달력<br>4. 수화 표현력(비수지기호, 역할전환) | 2. 정확성과 유창성                  | 3. 통역태도        |
| 평가자D | 1. 수화어휘력과 표현력                           | 2. 원문의 이해도                   | 3. 통역분위기       |

평가자 A는 평가자 중 가장 구체적인 기준을 제시하였고 학생별로 자신의 평가 기준 1-5에 따라 획득 점수를 매겨 순위를 정했다. 그는 다른 평가자들처럼 정확성도 중요한 평가기준으로 삼고 있지만 논리적 응집성과 언어구사력을 중요한 평가기준으로 보았다. 통역의 정확성에 대한 평가자 A의 기준은 예를 들어 주어부와 술어부의 묶음과 연결이 자연스러운지, 문장과 문장 간의 구분이 정확한지 등에 대한 평가기준을 가지고 있었으며, 통역의 적절성이라고 할 수 있는 수화어휘 표현력과 문장 구성력 부분에서도 비수지, 수형, 수동, 수위의 완전한 실현 여부와 수화문장 구성 시 한국어 어순에 따른 수화문장을 구사하고 있는지(한국어 영향을 받았다는 것을 암시함)를 세밀하게 평가하였다. 뿐만 아니라 전체적인 통역흐름에서 문맥의 끊김 현상과 통역태도를 평가기준으로 삼고 있었다.

평가자 B 또한 매우 다양한 평가항목들을 제시하였으며 그 내용을 보면 원문의 내용을 제대로 이해하고 통역을 하는지를 평가하는 통역의 충실성에 중점을 두었다. 내용의 충실한 전달보다는 단어 통역에만 집중하여 문장의 주어와 목적어 등이 불확실하게 나타나는 실수 여부를 주 평가기준으로 삼았다. 그는 또한 틀린 통역과 부정확한 수화표현 등을 거론함으로써 통역의 정확성에 비중을 두었고 통역의 적절성에 해당하는 통역사의 불필요한 몸짓과 버릇도 평가의 기준으로 삼았다. 이러한 통역사의 통역태도는 몰입에 방해가 됨을 지적하였다.

평가자 C는 원문의 이해 정도와 통역의 정확성 그리고 수화의 유창성 등을 평가기준으로 제시하였다. 평가자 C의 평가기준은 평가자B의 평가기준과 가장 유사한데 이는 1차 평가의 결과에도 반영되었다. 즉, 1차 평가에서 학생1, 3, 5의 평가 결과가 두 평가자에게서 일치하는 것은 평가기준이 유사하기 때문이다. 다른 평가자와는 달리 평가자 C는 통역의 정확성에서의 통역 오류를 가장 중요한 평가기준으로 설정하고 통역의 오류 발생 회수를 체크하여 평가에 반영하였다. 그러나 오류의 유형에 따라 평가자는 판단하기 어려운, 통역자 자신만이 알 수 있는 오류유형이 많기에 오류



발생 횟수에 전적으로 의지하지 않고 전체적인 통역 완성도를 고려하여 평가하였다.

통역사 D는 수화어휘력과 표현력에 가장 비중을 두어 평가하였다고 자신의 평가 기준을 서술하였다. 따라서 통역 어휘수가 가장 빈약했던 학생4에게 5순위를 부여하였다. 그러나 통역 어휘수가 가장 풍부했던 학생1에게는 1순위가 아닌 2순위를 부여하였다. 이는 비록 학생1이 사용 어휘수가 풍부하기는 했지만 공간을 활용해서 표현해야 하는 수화의 특성을 충분히 살리지 못했기 때문이다. 따라서 통역사D의 경우 비록 수화어휘 사용 능력에 가장 비중을 두었음에도 불구하고 부정확하거나 부적절한 수화표현에 대해서는 분명히 평가에 반영하는 태도를 보였다.

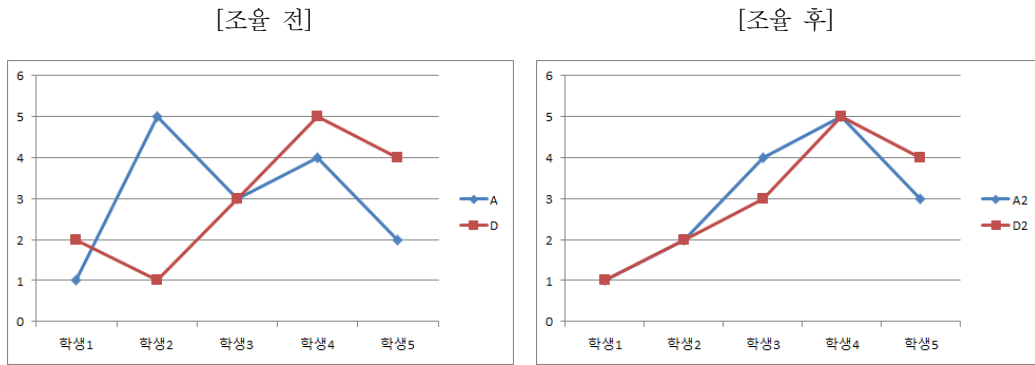
통역의 평가기준과 관련하여 학자들에 따라 접근과 분석방법이 다양하고 평가기준에 대한 개념이 서로 다르게 해석되어 사용되고 있지만(이혜승, 2008) 공통적으로 나타나는 통역 평가기준을 보면 등가와 정확성을 포함하는 충실성, 문체적, 형식적 규범의 준수여부를 평가하는 적절성과 유용성(Riccardi, 2002), 그리고 문장의 완결성, 속도, 유창성, 태도 등을 포함하는 부가적 차원(Shlesinger, 1997) 등을 들 수 있다. 일반 음성통역에서 통용되는 이러한 평가기준들은 수화통역의 평가기준과도 크게 다르지 않음을 앞서 살펴보았던 수화통역 평가자들의 평가기준 분석을 통하여 발견할 수 있었다.

## 2. 평가자 간 평가기준 조율 가능성 분석

학생들의 통역물에 대하여 4명의 평가자들은 전반적으로 유사한 평가기준들을 적용하는 것으로 나타났으나 평가 결과가 상이하게 나타난 것은 평가기준이 되는 요고 평가하였기 때문인 것으로 판단된다. 특히 평가자 간 가장 큰 차이를 보였던 평가자 A와 D의 경우에도 평가기준에 있어서는 많은 일치점을 보인다. 이에 연구자는 평가자 A와 D를 중재하여 1차 통역 평가 결과를 공유하고 상호 평가기준에 대한 논의 과정을 거쳐 평가항목을 크게 통역의 '정확성 및 충실성-오역과 생략'에 70%, '적절성-수화표현력(비수지, 강약, 공간활용)과 태도'에 30%의 배점을 부여하는 것으로 합의점에 도달하였다. 그리고 새롭게 조율한 평가기준으로 최상위 학생 5명의 통역물을 다시 평가토록 한 결과 <그림 3>와 같은 결과를 얻었다. 여기서 1차 평가 결과는 평가자 A, D로 동일 평가자 이면서 2차 평가 결과는 평가자 A<sub>2</sub>, D<sub>2</sub>로 표기 하였다.

<그림 3>를 보면 평가기준 조율 전 두 평가자 간 비교가능성은 학생3에서만 일치하였고 학생2에서 4등급, 학생5에서 2등급, 나머지 학생에서 1등급씩 차이를 보였으나 평가기준 조율 후에는 학생1, 2, 5에서 일치를 보였으며 나머지 학생의 경우에도 경향에서 일치를 보이며 1등급의 차이만을 보였다. 이는 평가의 기준에 대한

조율과정이 평가 집단 내에서 평가의 객관성을 확보하는 것으로 볼 수 있다.



<그림 3> 평가기준 조율 전과 후 비교

### 3. 조율된 평가기준에 대한 제3의 평가 검증

이에 대한 검증을 위하여 이번에는 평가에 전혀 개입하지 않은 제3의 평가자 E를 투입하여 동일한 평가물에 대하여 순위 평가를 하도록 하였다. 즉 평가기준의 조율 과정을 거친 평가자와 그렇지 않은 평가자 간의 차이를 조사하기 위함이었다. 그 결과 <표 3>과 같은 결과를 얻었다.

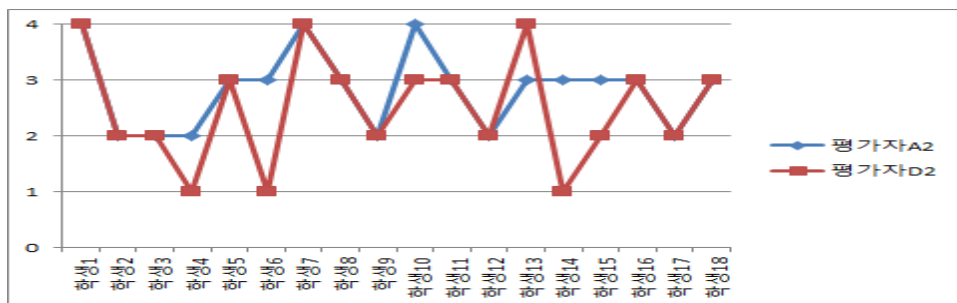
<표 3> 평가기준 조율 후의 평가 결과와 제3의 평가 결과 비교

| 학생  | 평가자     |                       | 평가자     |                       | 평가자 E |
|-----|---------|-----------------------|---------|-----------------------|-------|
|     | A(조율 전) | A <sub>2</sub> (조율 후) | D(조율 전) | D <sub>2</sub> (조율 후) |       |
| 학생1 | 1       | 1                     | 2       | 1                     | 1     |
| 학생2 | 5       | 2                     | 1       | 2                     | 3     |
| 학생3 | 3       | 4                     | 3       | 3                     | 2     |
| 학생4 | 4       | 5                     | 5       | 5                     | 4     |
| 학생5 | 2       | 3                     | 4       | 4                     | 5     |

먼저 평가자 A, D에 대한 평가기준 조율 전과 후의 비교 결과를 보면 평가자 A에게서 많은 변화가 있음을 알 수 있다. 이는 평가자 A가 평가기준 조율과정에서 평가자 D에 비해 더 많은 영향을 받았음을 의미한다. 두 평가자는 평가기준 조율 후

60%의 일치율을 보임으로써 조율 전의 20%보다 비교가능성이 높아졌다. 반면 새롭게 평가에 투입된 평가자 E와의 일치정도를 보면 평가자 A<sub>2</sub>와 20%, 평가자 D<sub>2</sub>와도 20%의 일치율을 보인다. 이는 1차 평가에서 조율 전 평가자들이 보였던 일치정도와 비슷한 수준이다.

최상위 학생 5명의 통역물이 균질성이 높고 차등화가 어려워 가치 있는 평가 대상이라고는 하지만 대상이 5명이라는 점은 결과의 신뢰도를 확보하기에는 어려운 숫자이다. 그래서 이번에는 18명의 수강학생 전체의 통역물을 대상으로 평가기준의 조율이 미치는 영향을 검증하기 위하여 전체평가를 실시하였다. 평가자들은 학생 18명이 제출한 기말시험 통역물에 대하여 가장 높은 수준(1수준)에서부터 가장 낮은 수준(4수준)까지 4등급으로 평가하도록 하였다. 먼저 평가자 A<sub>2</sub>와 D<sub>2</sub>의 평가 결과를 비교한 결과는 <그림 4>과 같다. 평가자 D<sub>2</sub>가 A<sub>2</sub>보다 전반적으로 높은 점수를 부여한 차이점은 있지만 두 평가자가 평가의 일치율을 보인 학생 수가 12명으로 일치율이 66.7%를 보였다. 그리고 평가 결과 2수준까지 차이가 나는 2명의 학생을 제외하고는 나머지 4명 모두 1수준의 평가 차이를 보여 상당히 일치하는 평가 경향을 보였다. 이것은 두 평가자가 유사한 평가 요소들을 상당히 근접한 비중으로 적용한 것으로 해석 할 수 있다. 이는 1차평가에서 학생3의 평가를 제외하면 두 평가자 사이의 일치도가 0%인 점을 고려할 때 두 평가자간 조율가능성은 크게 개선된 것으로 평가된다.

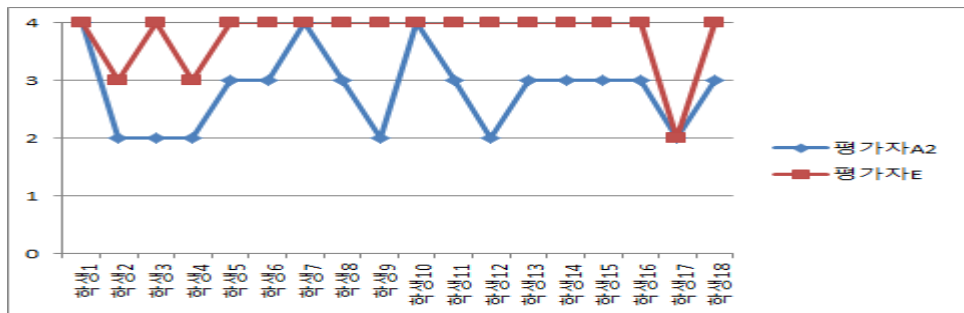


<그림 4> 평가자 A<sub>2</sub>와 평가자 D<sub>2</sub>의 평가 결과 비교(2차 평가)

평가자 A<sub>2</sub>와 D<sub>2</sub>가 보인 평가 일치 정도에 대한 의미를 좀 더 명확하게 알기 위하여 이번에는 이전 조율 과정에 전혀 개입되지 않은 제3의 평가자 E를 투입하여 학생 18명의 통역물을 동일한 방법으로 평가하도록 하였다. 평가자 E의 평가 결과는 조율내용을 공유한 평가자 A<sub>2</sub> 및 D<sub>2</sub>와 각각 비교되었으며 그 결과는 <그림 4>, <그림 5>과 같다. 즉, 이 평가에서 평가자 E는 조율내용을 공유하지 않았으므로 개인의

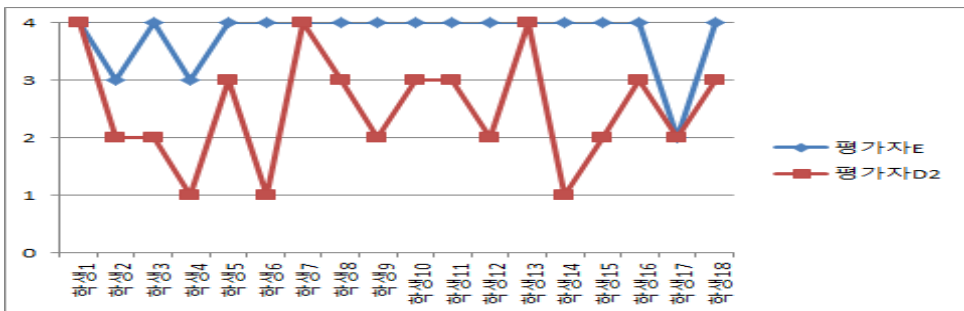
평가기준을 적용하도록 하여 조율효과를 진단해보고자 하였다.

평가자 A<sub>2</sub>와 평가자 E의 평가결과는 <그림 5>에서와 같이 평가 일치수가 4명으로 22.2%의 일치율을 보였으며 평가자 A<sub>2</sub>는 평가자 E에 비해 최저점을 주는 경우가 적었으나 두 평가자 모두 최고점은 부여하지 않았다. 평가가 2 수준 이상 차이 나는 학생도 3명으로 학생들의 실력차가 그리 크지 않아 평가의 어려움이 반영된 것으로 해석할 수 있을 것이다.



<그림 5> 평가자 A<sub>2</sub>와 평가자 E의 평가 결과 비교

평가자 D<sub>2</sub>와 평가자 E의 결과를 비교해보면 두 평가자간 결과가 일치하는 경우는 4건으로 16.7%에 해당한다. 반면 2수준의 평가 차이가 5건, 심지어 3수준의 차이도 2건이나 있었다. 이는 동일한 통역물에 대하여 평가자의 기준에 따라 매우 큰 차이가 날 수 있음을 명확하게 보여주는 사례라고 할 수 있다.



<그림 6> 평가자 D<sub>2</sub>와 평가자 E의 평가 결과 비교

앞서 언급한대로 모든 평가자들의 평가기준 자체는 크게 차이가 나지 않는다.

대부분의 평가자들은 통역에 있어서 정확성과 충실성, 그리고 적합성을 매우 중요한 평가기준으로 판단하고 있다. 그러나 이러한 기준을 적용함에 있어서 적용의 비중을 어떤 요소에 더 둘 것인지, 그리고 어떤 요소를 우선적으로 평가의 요소로 삼을 것인지 등에 대한 조율이 없을 경우 평가자 개인의 주관적 판단과 가치에 따라 평가의 결과는 매우 큰 차이가 발생할 수 있다. 이상의 결과를 통해 평가기준의 조율 과정이 평가에 좀 더 객관성을 담보할 수 있다.

#### IV. 결론 및 제언

모든 평가가 그렇듯 통역 평가에서도 평가자 간 비교가 가능할 수 있는 평가기준의 타당성과 신뢰성, 공평성이 중요한 문제로 인식되고 있다(Clifford, 2001). 이 연구의 출발점 역시 바로 수화통역 수업에서의 평가의 타당성과 신뢰성, 공평성 확보에 대한 고민에서 시작되었다. 연구자는 수화통역사 양성과정의 ‘교육통역’ 수업을 진행하면서 연구자 자신의 총괄평가는 과연 평가기준의 타당성과 신뢰성, 공평성을 확보하고 있는가에 대한 의심과 함께 평가자들은 과연 학생들의 통역물을 평가함에 있어서 어떤 기준을 가지고 있는지 그 기준의 적용은 평가에서 어떤 현상으로 나타나는지를 알아보고 싶었다. 그리고 평가에 있어서 기준이나 그 기준의 적용이 평가자 간 큰 편차를 보일 경우 평가기준의 조율은 가능한지, 그리고 평가기준을 조율할 경우 실제 평가의 객관성에 긍정적인 영향을 미치는지 알아보고자 하였다.

연구는 다면적으로 진행되었다. 먼저 ‘교육통역’ 수업 참여자 가운데 수화능력이 우수하고 균질성이 높은 학생의 시험 결과물을 평가대상으로 선정하였고, 이 통역 결과물에 대하여 대학에서 강의와 평가 경력이 있는 평가자들에게 순위평가를 의뢰하였다. 그 과정에서 평가자 간 가장 큰 차이를 보인 2명의 평가자를 선별하였고, 이들과 조율내용을 논의한 후 이 조율된 평가기준으로 재평가토록 하였다. 조율된 평가기준의 타당성과 신뢰성을 알아보기 위해 조율내용에 노출되지 않은 제3의 평가자를 투입하여 평가기준의 조율 가능성과 조율의 영향에 대하여 진단하였다. 그 결과 다음과 같은 결론에 도달하였다.

첫째, 평가자들 간의 평가기준을 분석한 결과 평가자들이 서술한 평가의 기준에 유사점이 발견되었다. 이들 평가자들은 표현의 차이는 있었지만 통역의 정확성과 충실성, 그리고 적절성과 관련된 평가 요소들을 평가의 기준으로 삼고 있었다. 이는 일반 음성통역에서 통용되는 충실성, 적절성, 유용성, 완결성과 같은 평가기준들과 크게 다르지 않은 것으로써 통역의 범주에서 평가기준은 음성언어나 수화에서 동일하게

적용될 수 있음을 시사한다.

둘째, 평가자 간 평가기준의 조율 가능성을 알아보기 위하여 참여자들의 논의 하에 조율된 평가 기준으로 정확성 및 충실성(오역과 생략), 그리고 적절성(수화표현력 및 태도)을 7:3의 비율로 정하였다. 조율된 평가기준을 가장 큰 평가 편차를 보였던 2명의 평가자에게 적용하여 재평가 하도록 한 결과 1차 평가보다 평가 일치율이 증가함으로 보였다. 따라서 평가자 간 평가기준 조율의 가능성이 높은 것으로 분석되었다. 이는 평가의 타당성과 신뢰도, 공정성을 높이기 위해 평가기준의 합리적 조율이 필요함을 시사한다.

셋째, 두 번째의 분석을 검증하기 위하여 조율내용에 노출되지 않은 제3의 평가자를 투입하여 평가토록하고 그 결과를 조율된 평가기준을 적용한 평가와 비교한 결과 조율을 거친 평가의 경우가 그렇지 않은 경우보다 현저히 높은 일치율을 보였다. 이는 통역사 평가에 있어 평가기준을 수립해야 할 필요성을 강하게 시사하는 결과로 볼 수 있다.

이러한 결과에 따라 향후 논의해야 할 점을 찾는다면 다음과 같다.

첫째, 음성통역 분야에서 통역의 평가기준에 대한 논의가 활발히 이루어지고 있는 것에 반해 수화통역 분야에서 아직 평가기준에 대한 논의가 거의 이루어지고 있지 않다. 이는 현재 양성과정에서 다양한 형식으로 진행되고 있는 통역 평가가 타당성과 신뢰성, 그리고 공정성을 확보하지 못하였을 수 있음을 시사한다. 따라서 수화통역 평가자들의 평가기준에 대한 논의와 후속 연구를 장려할 필요가 있을 것이다. 단, 수화는 음성언어와 다른 언어적 고유성을 가지고 있으므로 통역에 있어서도 수화통역 평가의 고유성이 고려되어야 할 것이다.

둘째, 평가기준의 설계 못지않게 평가기준의 적용방법에 대한 논의가 이루어져야 할 것이다. 통역평가는 단일기준에 의해 평가될 수 없는 만큼 복수의 평가기준이 어떤 비중으로 적용되어야 할지, 또 통역의 영역에 따라 기준의 적용 정도가 어떻게 달라질 수 있는지도 논의될 필요가 있을 것이다. 예를 들어, 법률통역이나 의료통역의 경우 다른 통역에 비해 낱어 전달보다는 원문의 단어까지 섬세하게 전달하는 것에 더 비중을 둘 수 있을 것이며, 반대로 종교통역이나 교양강좌통역의 경우에는 일대일 대응 통역이 꼭 필요하지 않을 수 있다. 이러한 통역 분야 간의 차이를 고려한 평가기준의 개발이 논의 되어야 할 것이다.

## 참고문헌

- 민은주 (2013). 한국어-한국수화 동시통역 사례분석 및 평가에 관한 연구. 박사학위 논문, 나사렛대학교 대학원.
- 이연향, 한미선, 오미형 역 (2009). **통역학 입문**. Pöchhacker, F.의 *Introducing Interpreting Studies*. 서울: 이화여자대학교 출판부.
- 이혜승 (2008). 동시통역 품질 평가에 대한 경험적 연구 고찰. **통번역학연구**, 12(1), 91-107.
- 정철자 (2012). 통역 평가 비교가능성: 교육상황에서의 사례연구. **통번역학연구**, 16(2), 219-239.
- 최상배 (2010). 장애학생지원센터 수화통역사의 수화언어 동시통역 과정 분석. **특수아동교육연구**, 12(1), 23-45.
- 황주희 (2013) 청각장애인을 위한 수화통역서비스 의무제공 현황 미 활성화방안 모색-사범, 의료영역을 중심으로. **한국장애인복지학**, 20, 63-87.
- Clifford, A. (2001). Discourse theory and performance-based assessment: Two tools for professional interpreting. *Meta: Translators' Journal*, 46(2), 365-378.
- Kurz, I. (2001). Conference Interpreting: Quality in the Ears of the User. *Meta*, 45(2), 394-409.
- Riccardi, A. (2002). Evaluation in Interpreting, *Teaching Translation and Interpreting 4*, Building bridge. Amsterdam and Philadelphia: John Benjamins, 115-126.
- Sawyer, D. B. (2001). The Intergration of Curriculum and Assessment in Interpreter Education: A Case Study doctoral dissertation, University of Mainz(<http://archimed.uni-mainz.de/pub/2001/0097/diss.pdf>).
- Shlesinger, M. (1997). Quality in simultaneous interpreting. In Y. Gambier, D. Gile & C. Taylor(eds.), *Conference interpreting: current trends in research*. Amsterdam: Benjamins, 123-131.
- Wu, S. (2010). Some reliability issues of simultaneous interpreting assessment within the education context, in Pellat, V., Griffiths, K., & Wu, S. (eds.). *Teaching and testing interpreting and translating*, 301-325.

## An Case Analysis on the Possibility of Coordinating Evaluation Standards among Evaluators in Relation to Educational Sign Language Interpretation

Min, Un Ju

Korea Nazarene University

<Abstract>

This study compared the results of evaluating the interpretation competencies of the students taking 'Educational Interpretation' in Department of Sign Language Interpreting Education in N University. Student evaluations were requested from educators in charge of sign language interpretation to investigate whether or not evaluation standards could be coordinated and the cases were analyzed.

Four evaluators graded the interpreting results of five students who obtained excellent scored in the final evaluation among students who took the course, 'Educational Interpretation' in the first semester of 2013 and freely describe detailed evaluation standards. As a result, even though they had similar evaluation standards, evaluators gave different priorities in application to show big differences in evaluation results. Two evaluators who had showed the greatest difference among four evaluators passed through the coordination process to investigate where or not the coordination could be made to overcome the difference among evaluators. As a result, it was found that evaluation standards were able to be coordinated. There was the accordance rate of 66.7% as a result of evaluating the final exams of all students who had taken the course based on the coordinated standards. Lastly, as a result of putting the third evaluator who had not been involved in the previous evaluation to evaluate all students, the accordance rate with the results of the previously coordinated evaluators was significantly low. This result proves that the objectivity of evaluation can be secured through the coordination of evaluation standards. Therefore, it is implied that it is necessary to prepare evaluation standards in the curriculum for prospective interpreters.

**Key Words :** Korean sign language, educational interpretation and comparison of evaluations

---

논문 접수: 2013. 11. 05 심사 시작: 2013. 11. 10 게재 확정: 2013. 11. 26