

음악능력의 정의와 평가방법에 관한 고찰*

(음악교육평가의 음악심리학적 접근방법)

조 호 임**

차 례

- I. 들어가는 말
- II. 테스트를 통한 음악능력의 차별화
 - 1. 테스트란 무엇인가?
 - 2. 음악능력의 구조적 관련자질
 - 3. 구조적 관련자질에 관한 논쟁
- III. 음악테스트 과제의 구조
 - 1. 음악테스트 과제의 선택과 개발
 - 2. 지각과 개념과제
 - 3. 연주과제
- IV. 맺음말

I. 들어가는 말

일반적으로 '음악능력'을 말할 때, 여러 가지 다른 음가, 음정, 강약을 식별하는 능력이나 선율을 조화시키는 능력, 시창력, 악기를 연주하는 능력과 같은 숙련성을 의미한다. 이러한 능력들은 다양한 숙련의 정도에 따라 학습되는 것이므로 음악행동과 그것이 행해지는 일체의 조건을 전제로 하는 음악경험과 훈련, 선천적 자질 따위의 복합적인 요소를 내포하고 있다. 따라서 이러한 능력들은 어떤 성분으로 구성되었으며 그 특성은 무엇인지? 우리는 이를 어떻게 측정하고 평가할 수 있는지? 또한 음악능력 테스트의 평가 데이터를 음악교육에 효율적으로 적용

* 이 논문은 1994년도 교육부 학술진흥 재단의 공보과제 연구비에 의하여 연구 되었음.

** 서울 교육대학교 음악교육과, 부교수(Ph. D.), 현 하바드음대 교환교수.

하는 방안은 무엇인지가 이 글의 논점이다.

주지하는 바와 같이 개인의 음악능력과 행동을 평가하려는 노력은 서양음악과 심리학의 전통에 그 뿌리를 두고 있다. 미국을 중심으로한 초기의 시도는 전적으로 일반적인 심리학적 측정 방법론에 의존하였으며, 특히 청각적인 감각을 분별하는데 이를 주로 사용하였다. 그후 음악적 개성의 차이를 설명하려는 연구는 음악적성 심리학이나 음악개인차 심리학에서 전통적으로 몇 가지의 구조적 자질의 분류로 발전하였다. 감상테스트는 음악에 대한 심미적 반응들의 표준화를 추구하였으며, 음악지식 테스트와 성취테스트는 주로 작곡가, 악보, 악기, 음계 등의 이름에 대한 습득된 반응을 취급한데 반하여 음악능력 또는 음악재능 테스트에서는 음악적 성공을 예보하는 것을 시도하는데 전력을 기울여 왔다. 이는 한 개인이 이 테스트에서 좋은 성적을 나타낼 경우에 음악성취를 위한 훈련에서 이점을 유리하게 특히 연주분야에서 유용하다는 것을 인정하는데 근거를 둔다. 물론 음악행동의 복잡성으로 인하여 음악연주의 예견을 그 목적으로 하는 기존의 테스트들의 효율성은 이 분야의 여러 가지 학설을 살펴봄으로써 최소한 동의할 것을 구하여 보고자 한다. 따라서 본 논문에서는 그 역사적 규명 과정을 주요문헌 및 논문을 통하여 개관하며, 다양한 음악능력 테스트 방법에 대한 안내와 음악교육에 대한 효율적인 적용에 대하여 특히 Helga de la Motte-Haaber의 『Handbuch der Musik-psychologie』¹⁾, 『Musikpsychologie: Ein Handbuch』²⁾ 및 J. David Boyle의 논문³⁾을 주로 참조하였음을 밝힌다. 또한 현재 인쇄되어 출판된 음악테스트에 관한 상세한 안내는 음악 심리학 관련 문헌과, 특히 테스트 목록서로서 테스트의 개관과 관련 문헌 안내와 함께 특수분야 별로 분류된 Test in Print(Vol. I, II)⁴⁾를 인용하였다.

학교현장에서의 핵심적인 문제는 반대의 견지에도 불구하고 여전히 능력의 측정과 평가로 보인다. 따라서 이는 수업에 대한 상호작용의 과정에서 효과를 내는 측면에서 능력의 측정과 평가를 순화하기 위하여 수업내용의 통계적인 새로운 규범으로서 교육과정의 개선에서 다루어질 과제이기도 하다. 교수와 학습과정의 방법적으로 상이한 요소로서 평가의 문제는 개성적 유전인자와 사회적 유전인자에서 드러나게 된다. 그러므로 평가는 결코 재능의 확증과 그에 따른 능력의 검사에 국한되지 않으며, 커리큘럼을 수업의 항구적인 검사의 표시의 한 과정으로 볼 때에 평가는 교안, 교수양식, 교수목표, 방법 및 교재의 개선에 결정적인 가치를 지니게 된다. 이러한 관점에서 학생들의 능력에 관한 정보는 각 학생의 개인차에 따른 교수, 교육과정 및 프로그램에

1) Helger, Motte-Haaber, 1985.

2) Bruhn, H., Oerter, R. & Roesing, H., Reinbek bei Hamburg, Rowolt Taschenbuch Verlag gmbH, 1993.

3) Boyle, J. David, Evaluation of Music Ability. In: Colwell, R.(Ed.), Handbook of Research on Music Teaching and Learning: A Project of the Music Educators National Conference, New York, Schirmer Books, 1992, PP. 247-265.

4) Oscar Krisen Buros(ed), The gryphon Press, Highland Park, New York, 1974.

대한 객관적인 근거가되므로 중요하다. 거의 한 세기동안 교육자와 심리학자들의 연구가 지속되었음에도 불구하고 음악능력의 정의와 평가 그리고 그 내포 성분에 관하여 여전히 많은 논쟁들이 있다. 예컨대, 음악적성과 능력, 그 성취는 어느 정도로 구조적으로 구분될 수 있는가? 이들은 또 음악재능, 음악적 소질, 음악지능, 음악성, 음악청취 및 음악적 감수성과 어떻게 구분될 수 있는가?

실제로 용어에 관한 일치의 부족과 각기 용어마다 어떤 의미로 적용되고 있는지에 관한 문제는 학생들의 음악적 자질, 능력 및 학습의 측정과 평가에서 뿐만 아니라, 음악교사들이 개개 학생들의 학습기회와 음악을 통한 경험을 가지게 하는 중요한 결정에 그와 같은 자료를 일차적인 근거로 하는지에도 문제가 있다. 따라서 이 글에서 의도하는 바는 이러한 용어의 명확한 이해와 음악교육에서 그와 같은 평가적인 자료의 효율적인 적용이 필수적임을 음악심리학적 접근방법으로 밝혀보고자 한다.

가치있는 평가도구는 객관적인 테스트 실행이라는 관점에서 '테스트' *test*와 '측정' *measurement* 및 '평가' *evaluation* 개념의 구분이 요구된다. 이 글에서 의도하는 바에 따라 '테스트'는 한 개인의 특수한 과제나 일련의 과제에 대한 행동을 관찰하는 일체의 체계적 절차를 의미한다. 하나의 과제나 일련의 과제는 전체적으로 또는 부분적으로, 그렇지 않으면 전혀 그 결과가 나타나지 않을 수도 있다. 만일 한 사람이 어느 정도의 범위로 테스트 과제를 입증할 수 있는 양의 체계를 발전시키고자 한다면 이로써 '측정'을 포함하게되는 것이다. 대부분의 음악테스트들은 점수체계나 비율등급을 적용하여 양을 정하고 있다. 그러므로 테스트와 측정이란 용어는 밀접한 관계가 있으며, 또한 일시적으로 서로 교환해서 사용될 수도 있다. 그에 반하여 '평가'는 이 세 가지 용어중에서 가장 넓은 의미이며, 한 음악행동이나 그밖의 다른 시도의 수준과 질에 관한 판단과 결정을 해주는 것을 의미한다. 그러므로 평가는 행동이나 그밖의 시도와 관련된 자료의 체계적인 수집과 분석 그리고 해석을 참작하여야 한다. 대개 테스트와 측정의 자료를 결정하는데 완전한 정보가 제공되지 못하므로 평가의 정확성을 기하하면 테스트와 측정에 확실한 근거가 필요하다. 경험있는 교사와 행정가의 계몽되고 지각있으며 전문성있는 판단과 정확한 테스트와 측정의 자료의 합성은 자료의 확실한 근거가 없는 판단보다는 음악능력의 평가에서 좀더 나은 결정을 쉽게 해준다.

제Ⅱ장에서는 음악능력에 관한 고찰의 전망과 구조적 자질과 논쟁이 다루어진다. 관련자질의 구성에서는 음악능력, 음악적성, 음악지능, 음악적 소질, 음악재능, 음악적 감수성, 음악성 및 음악성취 따위를 규명해보고자 한다. 특수논점으로서 분류호칭의 적합성, 경험과 훈련의 역할, 지능으로서 음악, 테스트 과제의 특성 및 테스트의 타당성을 고찰하고자 한다. 제Ⅲ장에서는 현재 대두되고 있는 음악능력의 다양한 관점을 논평하며, 특히 현재 사용되고 있으며 개선된

음악테스트에 관한 요소의 고찰, 인지적 테스트 과제와 개념적 테스트 과제 및 평가된 연주에 관한 접근을 살펴보고자 한다. 그리고 결론적으로 제IV장에서 음악교육에서 연구하고 실행하기 위한 평가의 적용에 관하여 개관하려 한다.

II. 테스트를 통한 음악능력의 차별화

1. 테스트란 무엇인가?

개인차에 의한 음악능력과 행동을 평가하려는 노력은 서양음악과 심리학의 전통에 뿌리를 둔다. 초기의 시도는 전적으로 일반적인 심리학적 측정 방법론을 차용하였으며, 특히 청각 감각적인 분별의 측정을 사용하고 있다. 개성적인 차이를 설명하기위여 전통적으로 발전된 것으로서 몇 가지의 구조적 자질로 분류하고 있다. 따라서 음악적성, 음악능력, 음악재능 및 음악성취와 같은 구조적 자질들의 측정과 평가는 음악심리학으로 알려진 조사 연구의 평가영역에서 특별한 초점이 되었다. 또한 이러한 구조적 자질에 대한 평가는 음악교육에 실제적으로 중요한 의미를 갖는다.

최초로 널리 사용되는 음악적성을 평가하는 테스트가 미국에서 출판된지 거의 75년이 경과되었다. 악음을 듣는 여섯 가지 소질 또는 능력의 측정을 목표로한 Carl E. Seashore의 음악재능을 평가하려는 테스트는 많은 이러한 작업을 위한 모델과 기동력을 제공하였으며, 20세기 전반 동안에 다양한 자질을 측정하는 것으로서 이에 관련되는 이론과 방법론을 끌어내었다. 또한 이를 평가하는 몇 가지 논쟁으로 발전되었으며, 음악적성의 천성과 교육의 관점, 전체적인 측정에 대비한 특수측정, 용어의 정의, 다양한 자질의 측정 타당성 및 상이한 경험과 심리학적 속성에 대한 여러 가지 자질에 관한 문제는 오랫동안 미해결된채로 남아있다.

일반적으로 음악테스트를 말할 때는 영어의 음악적성 테스트aptitude tests 또는 독일어로 음악재능 테스트Begabungstests를 의미한다. 영어 개념의 achievement tests와 독일의 Leistungstests는 우리말로 '성취테스트'라 하며, 최근에 이르러 감상테스트appreciation tests(Wertungstests)로도 취급되는 오레곤 음악변별테스Oregon-Music-Discrimination 는 이에 대하여 중개적 위치에 있다. 일반적으로 테스트의 분류는 '적성테스트'와 '성취테스트'로 구분한다. Lundin(1967)은 그밖에 감상평가 테스트의 범주로서 음악취향 테스트tests of musical taste와 감상테스트 tests of appreciation 그리고 태도 테스트tests of attitude

로 분류하였으며 독일에서 기악테스트Instrumentaltests라하는 음악연주 테스트tests of musical performance를 도입하였다. 음악연주 테스트 또는 기악테스트는 음악문헌에서 음악 재능과 관련하여 보편화되기도 하였으나 이는 테스트 이론적 입장에서 볼 때에 성취테스트에 해당되므로 지나친 것이다. 테스트는 미국 음악교육이 확정한 규범에 의하여 능력을 판단하는 일정한 검사과정을 말한다. 그러므로 다른 문화권으로의 무제한 적용은 어렵다. 그밖에 여러 가지 각양 각색의 비교적 형식화되지 못한 테스트들이 많으며 이들은 대부분 성취테스트 부류로 간주된다. 음악대학의 입학시험도 이에 속한다. '감상테스트'는 그 기능으로 보아 '적성테스트'와 그 경계를 짓기가 명확하지 못하다. 오래곤 음악변별 테스트와 고든테스트Gordon-Tests의 세번째 부분을 비교하면 다양한 종류로의 명확한 구분이 어려움을 알 수 있다. 1954년의 Kyme의 테스트에서는 프레이징과 속도, 인토네이션과 다이내믹의 차이뿐만 아니라 형용사에 의한 계열에 의한 심미적 특성화도 요구하므로 '적성테스트'와 '감상테스트'가 서로 혼화되었다고 볼 수 있다. 타고난 재능과 지식의 습득의 서로 뒤엉킨 상호작용의 관점에서 '적성테스트'와 '성취테스트' 사이의 구별이 특수한 문제로 남는다. 이들은 서로 모순적으로 작용하여 윙테스트Wing-Test와 오래곤 음악변별 테스트Oregon-Music-Discrimination-Test를 비교해보면 그 반대를 예상하게 함에도 불구하고 현저하게 강한 주위환경의 영향을 나타낸다.

음악능력의 평가에 대한 최근의 발전은 다양한 자질의 특성, 발전 및 평가에 관한 현대사고에 영향을 끼친 여러 가지 이론들을 내포하고 있다. 이들중 가장 두드러진 이론으로 Gordon⁵⁾의 발전적 음악적성론, Gardner⁶⁾의 음악지능은 여러 가지 각각 관련된 다수의 지능중의 하나라는 이론, Karma⁷⁾의 음악적성은 음향적 소재의 지각력과 인식력있는 구조를 내포하고 있다는 견해를 들 수 있다.

Farnsworth⁸⁾, Shuter-Dyson 및 Gabriel⁹⁾이 사용한 것 처럼 '음악능력'은 과제의 성격에 상관없이 학생들이 주어진 음악테스트 과제에서 나타내는 것에 관한 포괄적인 용어로 쓰여질 것이다. 우리가 추정할 수 있는 바로는 테스트 과제들에는 넓은 범위의 청취, 연주, 분석 및 창작의 행동이 포함될 수 있다. Farnsworth는 음악능력이 행동력을 제공하지만 이 힘이 어떤 타고난 잠재력의 결과인지, 특수한 학습의 반응인지 또는 어떤 장래의 잠재력의 예보인지는 추정이

5) Gordon, E., Primary Measures of Music Audiation, Chicago, G.O.A.Publications., 1979a: The nature, description, measurement, and evaluation of music aptitude, Chicago, G.I.A.Publications. 1987.

6) Gardner, H., Frames of mind: The thory of multiple intelligences, New York, Basic Books, 1983.

7) Karma, K., Components of auditive structuring-Towards a theory of musical aptitude, Bulletin of the Council for Research in Music Education, 1985, 82, 1-13.

8) Farnsworth, P. R., The social psychology of music(2nd ed.), Ames, The Iowa State University Press, 1969, p. 151.

9) Shuter-Dyson, R., & Gabriel, C., The psychology of musical ability(2nd ed.), London, Methuen, 1981, p. 3-7.

어려우므로 여러 가지 가능한 용어중에서 가장 폭넓고 무난하다고 본다.

Shuter-Dyson 및 Gabriel¹⁰⁾은 음악적성 테스트와 음악성취테스트는 중첩 구성(Overlapping constructs)임을 알아내었다. “모든 성취테스트가 필연적으로 개인의 최초의 적성에 반응하므로 모든 적성테스트들은 성취테스트를 확대한 것이다.” 음악적성, 음악능력, 음악성, 음악재능, 음악성취의 평가로서 다양하게 호칭된 테스트 문항의 유형을 검토할 경우에 이 논점은 혼동을 가져온다. 테스트에서 흔히 사용되는 변별과제는 선택된 음악과제를 구사하는 수준의 차이를 보여주는 개인과 집단사이를 구별짓는 다양한 자질을 평가하는데 사용된다. 일반적으로 우리는 주어진 과제를 더 잘 표현하는 사람을 더 ‘재능이 있는’ 또는 주어진 과제를 잘 표현하지 못하는 사람보다 더 높은 수준의 ‘적성’이나 ‘능력’이 있는 것으로 취급한다.

이러한 용어들은 많은 음악교사나 연주가들에게 의미가 깊을 수도 있으나 문헌에서는 여전히 용어의 의미에 관한 일치를 이루지 못하고 있는 것으로 지적되고 있다. 용어들이 서로 실제로 자질을 중복시키는지 구분시켜주는지 간에 이들은 개념의 정의와 이해가 빈약하여 불분명하다. 사람이 음악에 의하여 상호작용하는 방법의 차이는 이러한 용어들을 정의하고 이해하는 복합적인 문제로 보인다.

다양한 자질 또는 George 및 Hadoges¹¹⁾가 제시한 ‘음악적 특질’은 사람들의 특성으로 생각할 경우에 음악적 행동으로부터 추론된다고 볼 수 있다. 개개 자질 또는 특성을 추론해 볼 수 있는 음악적 행동들은 아주 다양한 양상으로 변할 수 있으며, 이는 관찰자나 테스트 고안자의 철학적 전망, 이론적 확신, 음악적 경험, 음악적 사색 및 자질에 대한 일반적인 이해와 관련된 논점과 같은 많은 요소에 따라 달라진다. 교사와 연주자가 어떤 사람을 ‘재능있다’, ‘매우 음악적이다’, ‘아주 대단한 잠재력이 있다’고 평가하는데 사용하는 추론된 근거는 음악적성과 능력 또는 성취의 표준화된 평가에 따라 매우 달라질 수 있다.

테스트의 호칭문제를 간소화하기 위하여 Boyle 및 Radocy¹²⁾는 주어진 테스트 과제는 각기 다른 기능을 할 수 있으며 평가와 연관된 이들의 실제 논쟁은 주어진 테스트에 대한 타당성의 적절한 유형을 확립하는 것이라고 주장하였다. 그러므로 주어진 선율형이나 리듬형 변별을 요구하는 과제에 대한 학생의 실행은 일반적으로 특수한 변별기술의 학습(성취기능), 강함과 취약함의 영역(진단적 기능), 또는 음악 학습의 장래의 성공에 대한 잠재력(적성기능)을 나타낼 수 있다. 그러나 테스트 사용자는 테스트의 타당성의 적절한 유형이 다양한 테스트 기능에 대하여 확립되었는지를 확인하여야만 한다. 즉, 테스트가 실증적으로 만족스럽고, 비교적 표준화되었으며, 예보적이 될 수 있는지를 조사하여야 한다.

10) op. cit., p. 7.

11) op. cit., p. 401-403.

12) op. cit. pp. 8-12.

2. 음악능력의 구조적 관련 자질

주지하는 바와 같이 지금까지 많은 음악심리학자들이 음악능력의 개념을 규명하고자 노력하였으며 이것이 오늘날 개인차 음악심리학의 중요한 부분을 이루고 있다. 일반적으로 '음악능력'을 하나의 총체적 용어로 사용함으로써 음악능력의 평가에 관련된 여러 가지 용어들의 확산과 상이한 사용은 다양한 용어로 표현되는 자질에 대한 개관을 필요로 한다. 그러나 음악능력의 구조적 관련 자질을 명확하게 구분짓는 일은 매우 어려우며 경우에 따라서는 서로 분리되기 보다는 상호 관련되는 편이다.

1) 음악능력

Boyle¹³⁾은 그의 논문에서 음악능력을 다음과 같이 정의하였다. "음악능력은 어떤 사람이 음악적으로 무엇인가 '할 수 있다'는 것을 말한다. 이 용어는 일반적으로 특별한 자료에 대한 관련이나 그것이 어떻게 발전되었는지에 상관없이 총체적인 의미로 쓰인다. 그러나 이것은 한 개인이 가졌을 수도 있는 형식적 구조가 주어진 요소라 할지라도 유전적이며, 원숙한 잠재력, 문화 발전적 영향과 환경적 영향으로 다루어진다."

음악능력이 일방적인 개념인지 다양한 능력의 범위를 나타내는지는 음악능력 평가에서 초기의 주요 접근방법의 철학적 차이에서 파생된 논쟁이다. 이 논쟁은 과거의 Carl Seashore와 James L. Mursell의 논쟁 즉, Mursell의 Seashore 종합 테스트 Measures of Musical Talent¹⁴⁾ 대한 논평에서 시작된 것으로 보인다. Seashore는 음악능력이 몇 개의 특정하지만 막연히 관련되는 감각적 변별과제를 통해서 나타난다고 주장한 반면에, Mursell¹⁵⁾은 음악능력은 전체적으로 파급되는 능력all-pervasive ability으로 구성된다는 '총괄Omnibus'이론으로 반론을 제기하였다. Shuter-Dyson 및 Gabriel¹⁶⁾은 이 문제에 관한 요소 분석 연구에서 확대된 논점을 제기하였으나 명확한 해결이 없었다. "시간이 지나면서 음악능력이 하나의 막연하게 관련된 감각적 기술만은 아니라는 사실은 명백하게 드러났으나 일반적인 음악능력의 요소는 규명하기 어려운 문제로 남는다. 음악적으로 능력이 있는 사람은 특별한 음악적 상황에서 다양한 기술을 발휘할 수 있다. 그러한 기술이 앞으로의 전망의 관점이다.¹⁷⁾

E. Gordond¹⁸⁾은 음악능력을 기억력, 구별력, 인식력 및 판단력의 네 가지로 구분하였다.

13) op. cit., p. 248.

14) Seachore, C. E., The psychology of musical talent, New York, Silver Burdett, 1919.

15) ursell, J. L., The psychology of music, New York, w. W. Norton, 1937.

16) op. cit., pp. 51-76.

17) Radocy, R. E., & Boyles, J. D., Psychological foundations of musical behavior(2nd ed.), Springfield, charles C Thomas, 1988, p. 298.

즉 들은 음관계를 어느 정도나 기억하는가, 기억력을 전제로 하여 다른 두개 또는 그 이상의 음관계를 구별할 수 있는지, 주어진 음관계가 어떤 것인지 말할 수 있는가의 가장 후천적 성취가 요구되는 인식 및 여러 음관계 중에서 보다 좋은 음을 판단할 수 있는가의 가치기준의 능력을 말한다. 이와 같이 음악의 여러 가지 매개변수(파라미터)에 의하여 나타나는 음악능력은 청음에 의하여 상당 부분이 평가되기도 하지만 실제에서는 여전히 음고, 리듬 및 화성의 파라미터에 치우친 것이 사실이다.

2) 음악 적성

일반적으로 '음악 적성'은 음악학습을 위한 특히, 발전적 음악 기술의 잠재력을 가리키는 데 사용된다. 전통적으로 학자들은 적성을 '타고난 능력' 또는 '소질' 중의 한 기능으로 보아 왔다. 그러나 근래의 평가에서는 일반적으로 음악기술(감수성)이 정식 음악교육이 없이도 발전될 수 있는 것에 더하여 '유전적 천부의 자질과 숙성의 결과'라는 인식이 통설이다.

음악 적성 또한 음악 적성 테스트에 의해 측정되는 능력에 관한 용어로 개념지워진다. 그러나 음악 적성 테스트의 내용에 관한 평가들¹⁹⁾에서는 음악 적성을 예견해주는 음악과제의 성격에서 약간의 차이를 보여주고 있다. 그들의 논평에서는 음악 적성 테스트는 한 쌍의 음고, 강약, 음색, 음가를 변별하는 몇 개의 선율(음악적 또는 조성적이라고 부르기도함) 기억 테스트와 리듬과 템포, 박자, 패턴과 같은 다양한 관점과 연관되는 약간의 변별과제로 구성된다고 보고 있다. 음악적 인용의 두 가지 연출에 관한 우선 반응을 요구하고 있는 두개의 테스트는 그 프레이징, 리듬, 화성, 균형 또는 양식에 관한 용어면에서 더 우수하다.

더구나 그러한 테스트 과제들은 음악적 잠재력의 예견에 유효하여 적절하다고 평가된다. 음악테스트에 관한 많은 논쟁은 음악적 잠재력의 예견을 위한 이러한 과제들의 유용성에 관한 문제들에서 특히, 연구에서 일정한 비음악적 변수들이 예견의 정확성을 크게 증가시켰다는 데이터를 제시한 이후부터 비롯되었다²⁰⁾. Rainbow의 데이터에서는 학문적 성취, 학구적 지능,

18) The psychologz of music teaching, Prentice Hall INC, 1971, Chapter II, The discription of musical aptitude.

19) Lehman, P. R., tests and measurements in music, Englewood Cliffs, Prentice-Hall, 1968, pp. 37-56; Gordon, E., The psychology of music teaching, Englewood Cliffs, Prentice-Hall, 1970, pp. 12-36; George, op. cit., pp. 296-315; Boyle, J. D., A study of the validity of three published, standardised measures of music preference, Psychology of Music, Special Issue, 1982, pp. 14-21; Webster, P. R. New perspectives on music aptitude and achievement, Psychomusicology, 1988, 7, pp.177-182 참조.

20) Rainbow, E. L., a pilot study to investigate the constructs of musical aptitude, Journal of Research in Music Education, 1965, 13, pp. 3-14; Whellams, F. S., The relative efficiency of aural-musical and non-musical tests as predictors of achievement in instrumental music, Bulletin of the Council for research in Music Education, 1970, 21, pp. 15- 21; Hedden, S. K., Prediction of music achievement in the elementary school, Journal of Research in Music Education, 1982, 30, pp. 61-68; Karma, K.,

사회경제적 배경이 음악능력의 중요한 예보자로 나타난다. 중요한 음악 예보자에는 조성 기억, 음악성취, 음악에 대한 관심이 포함된다. Hedden은 몇 가지 음악적 변수와 비음악적 변수중에서 학구적 성취가 음악성취의 가장 좋은 유일한 예보자로 보았다. 명백히 성공은 음악적이던 학문적이던간에 음악에서 성공은 미래의 좋은 예보자이다.

음악적성이 사람들이 지니고 있는 다양한 수준의 특성이라고 볼 때에, 이는 음악적성이 띠는 성격은 '행동적' 이기보다는 '심리적'인 성격의 강한 부대적 의미를 갖는다. 아마도 이는 대다수의 음악적성 테스트들이 주로 한 쌍의 음악 반응중에서 지각과 인지의 차이에 의존하는 변별과제로 구성되었기 때문이다.

Karma는 음악적성이 지각과 인지 과정으로 구성되었다는데에 긍정하고 있으나 음악적성을 평가하는 전통적인 접근의 가치 에는 의문을 제기하고 있다. 그는 음악능력의 측정으로서 변별과제에 대한 지속적인 신뢰와 성취테스트와의 상호관련에 의한 이러한 테스트들이 수반하는 확증은 온당치못하고 순환적이라고 주장하였다²¹⁾.

Karma는 음악적성을 음악의 지각적 인지적 추구로 보고, '강렬한 형태gestalt'에 관한 형식화 기대와 인식, 강렬한 형태에 의한 구조화, 강렬한 형태에 대항하는 구조화, 변하는 기대, 시기, 강렬한 형태의 내적 구조의 분석"으로 구성된 정보 처리전략information-processing strategies을 만들었다²²⁾. 본질적으로 그의 구조화 과제structuring tasks는 수험자에게 세번에 걸쳐 반복되는 짧은 주제를 듣고 반응하기를 요구한다. 이 주제들은 음고, 음가 또는 강약이 다른 음들로 구성되었으나 단 한 가지 요소가 주어진 한 가지 문항에 의하여 변화된다. 음고의 구조(선율)를 주시할 경우에는 음의 강약과 음가는 변하지 않고 그대로 유지된다. 짧은 휴지후에 "응답"이 제공되고 수험자는 응답이 '같은지' '다른지'를 판단해야 한다. 다음의 예1)은 음고와 강약 문항을 보여준다.

(예 1) 음고와 강약의 구조화를 위한 Karma의 주제

(동일하다)

(다르다)

components of auditive structuring - Towards a theory of musical aptitude, Bulletin of the council for Research in Music Education, 1983, 82, pp. 1-13.

21) op. cit. 1982.

22) op. cit. 1985, p. 11.

Karma는 이러한 구조화 전략은 전통적인 테스트 과제들이 실행하는 것 보다도 더 심리학적 구조로서 추론된 음악적성에 대한 많은 강력한 근거를 제공한다고 주장하고 있다. 음악의 지각적 인지적 구조화에 대한 평가를 위한 Karma 테스트 과제들이 여전히 발전 과정에 놓여 있으나 인지적 구조화라는 용어로 음악적성을 고찰하려는 설득력있는 논점을 제공한다.

음악적성에 관한 문헌에서 최근의 다른 진전은 Gordon²³⁾에 의하여 주창된 발전적 음악적성에 관한 서술로서 아동의 음악적성이 확립되는 9세경까지의 아동 음악적성의 변동을 기술하고 있다. 그의 음악적성 테스트에서 나타나는 변별과제에 대한 어린 아동들의 응답의 불안정에 관한 그의 명백한 인식에도 불구하고 Gordon은 “아동의 음악적성은 태어났을 때보다 더 높은 수준에 결코 이르지 못한다”고 말한다²⁴⁾. 또한 발전기 동안의 유리하거나 불리한 비공식적 환경의 경험들이 남녀 아동이 태어날 때 지니는 음악적성의 수준을 음악성취에서 한 아동의 실현 여부를 가름한다고 주장한다²⁵⁾. 하나의 표면상의 모순으로서 발전적 음악적성이라는 개념은 연구의 관심있는 분야를 제시해준다.

3) 음악지능

수년동안 음악 지능이라는 개념은 음악능력의 평가를 위한 Seashore의 접근에 대한 반대급부로서 알려진 종합테스트test battery, 음악지능 표준테스트Standardised Tests of Musical Intelligence와 연관되어 사용되었다. 이 테스트는 음악능력의 일반요소에서 Wing의 확신을 반영하며, 합성되거나 종합된 측정으로 합쳐질 수도 있는 음악능력의 부속테스트들로 구성되었다. 이와 반대로 Seashore는 그의 음악재능 테스트(Mesures of Musical Talents²⁶⁾)의 부속테스트들은 6가지 다른 측정의 측면을 제공한다고 주장하였다. 몇 개의 음악적 인용의 두 가지 연출에 관한 판단을 해주는 테스트는 Wing의 접근이 Seashore보다 더 ‘음악적’으로 취급된다. Wing이 음악능력이 대부분 타고난다고 주장하지만 그의 테스트 특히, 네 가지 ‘감상’ 부속테스트는 필수로 서양음악 실체에 대한 감수성을 반영하는 반응을 요구한다. 그러므로 정확한 응답을 하는 데는 Seashore 테스트의 응답에서 보다 더 음악을 통한 경험에 달려 있다. 음악적성이나 능력 대신에 음악지능 측정이라 부르는 Wing의 호칭은 용어론에서 혼동을 더해준다.

보다 최근에 와서 음악 지능이라는 용어는 Gardner²⁷⁾(1983)의 다지능이론theory of

23) op. cit. 1979a: Learning sequences in music, Chicago, G.I.A. Publications, 1984; op. cit., 1987.

24) op. cit. 1984, p. 25.

25) ibid.

26) Seashore, C. E., Lewis, E., & saetveit, J. G., Seashore Measures of Musical Talents, New York, The Psychological Corporation, 1939/1960.

27) op. cit.

multiple intelligences에서 상세히 설명된 몇 개의 각각 연관되는 다지능 중의 하나로 인식되게 되었다. 그는 지능에 대한 8 가지 범주를 제시하였다. 즉, 뇌의 손상에 의한 잠재적 분리, 바보 학자와 천재의 존재, 음고와 리듬의 관계에 대한 예민성과 일치시키는 있는 핵심 조작, 전문 연주로 이끌어주는 발전의 역사, 진전의 역사, 경험적 심리학적 과제에 의한 보조, 정신적 발전의 보조, 및 상징체계를 취급하는 지능에 의한 정보를 암호로 바꾸어 쓰는 보조를 말한다. 다른 지능들은 Gardner에 의하면 언어, 논리와 수리, 공간, 육체-근육 운동 감각, 개인 지능으로 구분된다. Gardner가 지능으로서 음악능력의 평가에 아무런 특정한 방법을 제공하지 않지만 '특정한' 방법의 적합성에 의문을 제기하면서 아마도 '중간 영역'과 어느 정도 절충된 접근법이 가장 유용할 것이라고 제안하였다. 그는 보통 수준의 사람은 음악을 듣는 '도식'이나 '틀'을 가진 것으로 보인다고 하였으며, 잘 구조된 프레이즈나 한 악곡의 부분에 대한 기대와 같이 음악적 센스를 만들어주는 방법으로 부분을 완성시키는 최소한 발생 능력이 있다고 계속해서 주장하였다²⁸⁾.

4) 음악 소질

'소질capacity'은 전통적으로 유전적인 천부의 자질과 성숙의 결과로 개인의 능력의 부분으로 인정되었다. Seashore²⁹⁾는 그의 테스트가 타고난 또는 천부적인 소질을 측정하였다고 주장하였다. 그가 어느 정도 소질의 특성에 관한 정의를 하기위한 노력을 기울이기도 하였으나 그의 저술에서는 소질을 정신의 특질로 규정하였다. Lundin³⁰⁾은 소질의 더 포괄적인 의미를 인식하였으며 "우리가 음악적 행동을 발전시키는 기본틀로서 하나의 생물학적 잠재적 부여이다"라고 하였다. 그가 '타고난inborn'이란 말보다는 '생물학적biological'이란 용어를 사용한 것은 소질의 일부분으로 음악적 잠재력의 효과를 낼 수도 있는 성숙의 변수를 인정하는 경향을 띤 것이다.

Boyle³¹⁾은 "소질은 평가의 목적에 유용한 자질은 아니다. 그것은 환경적인 영향으로부터의 유전적으로 결정된 음악적 행동의 만족할만한 분리는 여전히 성취되어야 하기 때문이다."라고 지적하였다.

5) 음악 감수성

음악인들이 사용하는 것처럼 '음악 감수성Music Sensibility'은 다른 무엇보다도 악음고, 음색, 강약에 대한 예민한 감각적 변별을 해주는 능력을 의미한다. 초점은 미세한 음향적인 차

28) Gardner, op. cit. p. 108 참조.

29) Psychology of music, New York, McGraw-Hill, 1938, p. 332.

30) Lundin, R. W., An objective psychology of music(2nd ed.), 1969, p. 206.

31) op. cit. 1992, p. 250.

이에 대한 감수성보다는 음악적 감수성에 있다. 감수성은 변별과 반응 감정에 반영되는 것처럼, 음악에서 민감한 차이에 대한 반응과 지각을 내포하고 있는 것으로 보인다. 음악 감수성은 음악 표현에서 뿐만 아니라 음악 청취에도 반영될 수 있다.

감정이나 느낌, 차원은 음악능력을 평가하는 사람에게 문제를 만들어주지만 Gordon³²⁾은 그의 음악적성 파일 Musical Aptitude Profile(MAP)의 한 부분으로서 부속테스트 '음악민 감성 Musical Sensitivity'를 위한 케이스를 만들면서 다음과 같이 주장하였다. "음악창조성과 표현이 악구를 구성하는 음들의 조성과 리듬 관계를 지각하는 능력에서와 마찬가지로 최소한 음악교육의 성공에도 중요하다고 인식된 이후에, 음악적 감수성 측정은 기본 음악능력의 평가를 포괄적으로 만드는데 공헌하였다." Gordon의 MAP 외에도 감수성 부분이 포함된 두 가지 테스트가 출판되었다: Wing의 음악지능 표준테스트 Standardised Tests of Musical Intelligences³³⁾와 Long의 인디애나 오레곤 음악 변별 테스트 Indian - Oregon Music Discrimination Test³⁴⁾가 있다. 그러나 이 세 가지 테스트의 음악감수성 부분의 점수와 상호 관계는 매우 낮다³⁵⁾. 그러나 이 세 가지 테스트의 이러한 실질적인 상호 관계의 부족은 방법론적인 고려나 실질적인 예의 친숙함에 기인할 수도 있다. 자질로서의 음악 감수성은 일반적으로 인식되어 있으나 이를 평가하려고 고안된 선택된 측정은 분리된 측정으로서 잘 받아들여지지 못하였다.

6) 음악성

마찬가지로 불명확하지만 '음악성 musicality(독 Musikalitaet)'은 복합적인 특성으로 '음악적'이라는 상태를 가리키는 말로서 음악의 뉴앙스에 민감하다는 의미이다. 음악감수성과 같이 이 용어는 인지적이며 감정적인 차원을 내포하고 있다. 일반적으로 지능지수와 상호관련된 즉 흥연주능력은 작곡의 천재성과는 전혀 상관이 없는 것으로 드러났으며, 자동적으로 연결되지는 않지만 음악가의 생산력과 재생산력도 다차원적인 특성에 관련되고 있음을 보여준다. Seashore테스트와 같이 음악청취에서 무엇보다도 변별력을 판별해주는 음악성테스트도 다차원적으로 고안되었으며 표준화의 모델에서 출발하였다. 유일하게 단 한 개의 테스트, Gaston

32) Musical aptitude Profile, Boston, Houghton Muffin, 1965, p. 5.

33) Wing, H. D., Sstandardised Tests of Musical Intelligence, The Mere, England, National foundation for Educational Research, 1939/1961.

34) Long, N. H., Indiana-Oregon Music Discrimination Test, Bloomington, Midwest Music Tests, 1965.

35) Gallagher, F. d., a study of the relationships between the Gordon Musical aptitude Profile, the Colwell Music achievement tests, and the Indiana-Oregon Music Discrimination Test, Unpublished doctoral dissertation, Indiana University, Bloomington, 1971; Sampsell, S. A. , A study of the comparative validity of three measures of music sensitivity, Unpublished master's thesis, Pennsylvania State University, University Park, 1980.; Boyle, op. cit. 1982..

의 음악성 테스트 Test of Musicality³⁶⁾만이 음악성 측정을 목표로 하였다. 이 테스트의 명칭으로 미루어 음악감수성을 반응하는 우선 과제를 내포하고 있을 것으로 기대할 수 있지만 그러나 그렇지 않다. 이 테스트의 음악 과제는 하나를 제외하고는 변별과제들이다. 이 단 하나 예외는 수험자에게 반복 진행된 음들의 최종음의 높고 낮음을 가리는 질문을 하는 과제이다.

7) 음악 재능

우리는 다만 인간의 능력과 특성의 변화를 관찰할 수 있으며 이를 확인할 수 있을 뿐이다. 그러나 보다 더 중요한 것은 이들이 어디에 근거하는지를 아는 일이다. 그것은 이를 통하여 판단을 용이하게 할 수 있기 때문이다. 예컨대, 한 아동이 천재성을 보여 악기교육의 투자가 가치가 있는지에 대한 문제는 유전된 재능을 확증할 수 있다면 대답은 명백해지며, 마찬가지로 능력이 습득될 수 있다는 것을 안다면 학습의 지연을 개탄하게 될 것이다. 지역사회의 체계변화가 요구가 대두될 때에는 언제나 사람사이의 드러난 차이는 환경의 영향으로 간주되어 학습을 통하여 고르게 조정할 수 있다고 보는 것이다. 1970년대에 독일에서는 교육시스템의 개편과 관련하여 음악성의 차이가 생활환경이나 소질에 달려있는지에 대한 논쟁이 극치를 이루었다. 미국과 구소련에서는 행동주의와 반사론이 이미 사십년전에 개인적인 특성의 절대적인 조작에 대한 가능성으로서 두각을 나타내었다. 그후 모든 인종론이 하나의 운명적인 결정론을 추종하는 시대에 접어들어서는 다시금 유전된 재능이 강조되었다. 다시금 오늘날의 문화정책적인 풍토에서 전형적으로 되어버린 것처럼 타고난 특성에 의한 개인차로의 환원은 어떤 상태유지를 위한 이념론에 지나지 않으며 이는 바람직하거나 가능한 변화로 볼 수 없다고 Motte-Haaber³⁷⁾는 주장하였다.

능력이 유전되는지 학습에 의하여 습득되는지는 학문에서도 한동안 명백한 대답을 제시하지 못하였으나 이념론이 쉽게 정주할 수도 있는 이들 둘의 합성에 대한 관계를 다양한 짜임조식으로 확대하였다. 특히 음악성과 같이 재능은 유전적인 결정을 뚜렷하게 해주기 위하여 선호되는 파라미터이다. 방법적으로 다양한 반론들이 제기되고 있으나, 유전적 천재성을 확증하기 위한 실례로서 Bach 가문의 음악가들이 세대별로 정리한 족보³⁸⁾가 가장 선호되고 있다. 그러나 여기서는 넓게 분포된 과거의 세대에 대한 정확한 정보를 얻기가 어려우며 여성가족에 대한 고려가 배제되었으며 이에 관하여는 알려진 바가 적다. 그 밖에 환경의 영향으로서 아들들이 아버지의 직업을

36) Gaston, E. T., A test of Musicality(4th ed.), Lawrence, O'Dell's Instrumental Service, 1957.

37) op. cit., pp. 261f.

38) Bach가문의 칠대에 걸친 100명의 자손중에서 열한명의 우수한 음악가가 배출되어 유전적인 음악재능을 나타내는 예를 보여주고 있다. 그밖에도 Couperin의 가문도 음악가왕통을 자랑하고 있으며 유명한 수학자 Bernoulli가도 재능의 유전성을 보여주는 좋은 예로서 알려졌다. 그러나 학문적 방법상의 어려움은 여성가족에 대한 정보의 부족이다.

계승하고 그에 상응하는 조기교육과 집중적인 음악훈련을 받았다는 사실을 배제할 수 없다.

‘재능’(영talent, 독Begabung)이라는 용어는 흔히 음악연주와 연관되어 사용되지만 음악 연주 기술의 높은 수준은 또한 학습의 높은 수준을 반영한다는 사실도 시사하고 있다. 그러나 한 개인이 음악연주 기술이 높은 수준을 보일 때 재능이 있다고 말한다.

이러한 개념 정의와 실제 사용의 모순외에도 음악인이나 일반인들에 의한 용어의 일반적 적용은 평가를 위한 사용면에서 복잡하게 한다. 개성은 지정된 재능이라는 기준은 다양하고 또 부정확하다. 어쩌면 평가자는 이 용어의 사용을 피하고 단순히 한 개인의 음악재능이 그의 동년배를 능가할 경우에 하나의 불명확한 용어 정도로 알아야 할 것이다. 누구나 거의 일정한 음악재능을 지녔으나 그 중에 소수만이 재능있는 사람으로 인식된다. 재능있는 사람에 대한 호칭의 기준에 관련된 일치가 있을 지는 의심스럽다. 그럼에도 불구하고 우리가 어떤 사람이 재능 있다고 말하는 의미를 이해하는 것은 모호하기도 하지만 어느 정도 광범위하기도 하다.

8) 음악 성취

음악 성취는 음악을 통한 경험, 음악적 현상, 또는 음악과 관련된 자료들의 성과로서 음악의 완성을 말한다. 음악성취는 그러한 경험의 결과로서 무엇이 학습되었는지를 반영한다.

무엇이 성취되는지의 범위는 크고 시간의 다양한 시기를 거친 완성에서 반영될 수 있을 것이다. 음악의 완성은 매우 특별할 수도 있고 아주 보편적일 수도 있으며 일반 음악 지식, 기보 지식, 시청각 기술, 연주 기술, 작곡으로 이루어질 수도 있다.

음악 성취는 학교나 다른 교육 시설에서의 형식적인 학습의 결과일 수도 있으며 혹은 덜 구조적이며 비형식적인 시설에 의한 학습의 결과를 반영할 수도 있다. 음악교사는 주어진 교육의 성과를 내는 음악성취의 종류와 가장 밀접한 관련이 있다.

비교적 많은 음악교육의 특수한 성격의 차이로 대부분의 음악성취는 내용이나 과제가 특수하다. 학생들의 음악성취는 보통 그들의 특정한 학습 경험을 반영한다. 성취의 평가는 학습자와 교사에게 무엇이 학습되었으며 학습되지 못하였는지에 관한 교수법적 경험의 효과에 관련되는 피이드백feedback을 제공해주므로 중요하다.

3. 구조적 관련 자질에 관한 논쟁

음악능력 평가에 대한 몇 가지 논쟁이 위의 개관에서 대두되었으며, 이에 대한 약간의 근거로 분류호칭의 적합성, 경험과 훈련의 역할, Gardner의 음악지능론과의 관련, 음악능력 테스트의 검사 과제의 특성 및 테스트 타당성의 중요성의 관점을 다음에서 논술하고자 한다.

1) 분류 호칭의 적절함

우리는 테스트 결과에 관한 학생의 특히 다양한 자질에 대한 높게 추정된 특성에 관하여는 그 분류 호칭labeling을 삼가하여야 한다. 분류 호칭은 교사와 학생간의 상호작용에 잠재적으로 해로운 효과를 내는 것으로 알려졌다³⁹⁾. 음악능력을 검사하는 데에는 교육적인 타당한 근거가 있어야 하며 검사 대상 개성이나 개성을 위한 음악교육을 안내하고 강화시키는 책임을 질 사람에게 정보를 제공되어야 한다. 테스트의 선택과 개발은 교육적인 결정에 의해 만들어지는 것과 관련된 정보를 이끌어내는데 초점을 맞추어야 한다. 음악능력 테스트의 범위에 대한 정보는 매우 유용하다.

2) 경험과 훈련의 역할

대부분의 음악교사들은 음악적 경험과 훈련이 음악능력의 발전에 도움이 된다는 사실에 이의가 없다. 유전된 잠재적 개성이 무엇을 소유하였던지 음악과 서로 상호작용을 하기 훨씬 오래 전에 이미 결정된 것이다. 그러므로 음악교사는 그들이 어떤 영향을 미칠수 있는 요소에 초점을 맞추어야 한다.

음악적 발전에 초기 어린시절이 중요하다는 인식이 증가하면서(음악적 발전을 위한 비판적인 시기들이 제시되었던 범위까지도) 결코 알아낼 수 없는 개인의 타고난 적성에 대한 토론이 계속되어 거의 초점이 흐려질 정도에 이르렀다. 그러나 중요한 것은 형식적이거나 비형식적 음악경험의 성격과 질이 대다수 아동의 음악적 발전에 깊은 영향을 미치는 것으로 나타난다.

3) 음악 지능

Gardner의 유일한 지능으로서의 음악능력에 대한 인식은 많은 음악교사들이 오랫동안 믿어왔던 신뢰와 일치된다. 일정한 연주매체를 통하여 조작되고 개념화된 것과 소리로 들리는 음높이와 리듬형에 응답하는 지각능력은 전통적인 학구적(언어와 수리) 지능외에 어느 정도의 인지능력을 내포하는 것으로 보인다.

어쨌건간에 Gardner의 음악 지능에서 발전된 특수 측정을 위한 새로운 노력이나 아무 시도가 아직까지 보이지 않는 것은 놀라운 일이다. 그것이 무관심의 소치인지 현상태로의 만족에 서인지 또는 구조적 이해의 부족에서 오는 것인지는 분명하지 않다⁴⁰⁾.

39) Boyle, J. D. & Radocy, R. E., op. cit., pp. 27-28.

40) Boyle, J. David, op. cit., pp. 251-252.

4) 테스트 과제의 특성

Aristoteles 이후 영혼의 작용으로 이르는 '능력'을 더 정확하게 확증하기 위하여 20세기 전환기에 들어서 만들어진 테스트들은 다양한 영혼의 기본작용을 측정하는 과제를 개발하였다. Binet와 Henri(1895)가 발전시킨 지능의 특징에 대한 제안들과 마찬가지로 Seashore(1919)의 음악성을 측정하는 표준화테스트도 개개 요소들의 관념에서 출발한 것이다. 다양한 측정에 대한 테스트 과제의 대다수는 몇 개의 음악적 자극에 관한 것을 제외하고는 여전히 변별과제들이 주류를 이루고 있다. 근래 수년동안의 경향은 과제를 더 음악적으로 만들려하며 몇 개의 테스트들은 음악적 감수성을 평가하려는 의도에서 연주 부분을 더 확대시켰으나 테스트의 근본적인 구조는 미세하게 변화되었다.

아마도 테스트 고안자들이 새로운 접근을 시도할 시기가 도래한 것 같다. Karma의 지각적 인지적 구조의 테스트들은 전통적인 테스트 과제와 매우 다르며 유용한 새로운 방향을 제시해줄 수도 있다고 본다.

국내에서는 90년대 후반부터 기존의 시각적 테스트에서 청각을 통한 음악테스트로 이전하려는 노력⁴¹⁾이 시작되었으며, 최근에는 음악대학 입학시험을 위한 청음을 통한 음악테스트⁴²⁾가 개발되기도 하였다.

5) 측정의 타당성

음악능력 테스트의 타당성에 관한 관심은 음악테스트가 만들어진 이후 초기부터 지속되어왔으나 그 관심은 여기서 그 중요성을 강조하고 강조속에서의 가벼운 변화를 요구하기 위하여 되풀이되고 있다. Gronlund 및 Linn⁴³⁾이 지적한 것처럼 '테스트의 타당성'에 대하여 말하는 것보다 테스트 결과에 대한 해석의 타당성을 말하는 것이 더 명백하다. 어떤 테스트든지 일정한 특수 용도나 해석에 유효하다. 그들은 타당성은 테스트 고안자가 제공하는 다양한 종류의 증거에 대한 일방적인 개념이라고 주장한다. 목적하는 바에 따라서 테스트가 사용되고, 테스트 고안자는 관련 내용과 표준 그리고 자질을 제공하여야 한다. 요점은 타당성의 증거가 테스트의 목적에 대하여 명백하여야 한다.

41) 예컨대, 정세문 외 7명이 제작한 '국민학교용 표준화 음악능력검사' (서울출판사, 1970).

42) 예컨대, 이진용의 '보통 및 전문 교육에 있어서의 음악적 능력평가 시스템 개발'(1986).

43) Gronlund, N. E., & Linn, R. L., Measurement and evaluation in teaching(6th ed.), New York, Macmillan, 1990, pp. 48-50.

Ⅲ. 음악테스트 과제의 구조

앞장에서는 음악테스트에 의하여 추론된 용어와 자질을 논하였으나 여기서는 개인의 음악능력의 다양한 전망을 평가하는데 적용될 수 있는 테스트의 구조에 관하여 살펴 보고자 한다. 따라서 논의의 요점은 테스트 과제에서 요구되는 음악행동에 있다.

세 부분으로 나누어서 첫째 음악능력의 선택된 측정과 측정의 발전을 위한 요인들을 고찰하고, 둘째로는 음악능력의 지각적 개념적 차원의 평가에 사용되었던 접근 방법들을 개관하며, 마지막 부분에서는 음악연주의 평가에 관한 접근 방법을 논평해 보고자 한다.

많은 조사된 테스트 과제들이 집단형으로 실행되고 있으나 논술의 초점은 집단평가보다는 개인평가에 두었다. 그러므로 종이와 연필이나 에세이 질문에 의한 정식으로 평가되는 표준화된 평가요소에 관한 정보를 위한 테스트 문항에 관하여는 제외된다. 이에 관하여 자세히 알고자 하면 이 문항을 다룬 연구물⁴⁴⁾들을 참조하기 바란다.

1. 음악테스트 과제의 선택과 개발

음악능력의 측정을 선택하고 개발하기에 앞서 평가자는 기구의 사용과 디자인에 관련된 많은 요소를 생각하여야만 한다. 이는 경험있는 테스트 제작자나 사용자에게는 명백하겠으나 충분하게 그 기준을 재확인하는 것이다.

1) 테스트의 목적

어떤 음악능력 테스트를 개발하거나 선택하는데 가장 먼저 고려해야 할 점은 그것이 사용될 기능의 타당성을 확고히 하는 일이다. '타당성'과 '신뢰도'는 어떤 테스트를 선택하는 두 가지 주된 기준이다. 본질적으로 신뢰도는 테스트가 측정하는 것의 일관성에 관련되며, 타당성은 테스트가 측정하고자 하는 것을 어느 정도 잘 측정하는가를 의미한다. 신뢰도가 없이는 어느 테스트도 가치가 있을 수 없으나, 테스트가 사용되는 특수한 목적에 따라서 타당성이 없어도 테스트가 유효할 수 있다. 다기능을 목적으로 하는 테스트는 각 기능에 대한 타당성 데이터를 제공하여야 한다. 그러므로 테스트의 첫 단계는 테스트가 사용될 기능을 명백하게 규정하는 일이다.

44) Gronbach, E. G., *Essentials of psychological Testing*(4th ed.), New York, Harper & Row, 1984; Ebel, R. L. & Frisbie, D. A., *Essencial of educational measurement*(5th ed.), Englewood Cliffs, Prentice-Hall, 1990; Gronlund & Linn, op. cit.; Boyle & Radocy, op. cit., 1987.

학생들의 음악 학습에 대한 잠재력을 예견해주는 테스트는 얼마나 잘 장래의 음악 학습을 진단하는가에 관한 증거를 제공해야 한다. 그러한 증거는 장래 음악 학습의 일정한 기준에 관련되는 테스트 상의 평점과 연관된 기준이다. 한 학생의 현재 능력을 평가하는 음악테스트는 그에 일치되는 능력의 증거와 관련된 기준을 제공하여야 한다. 다시 말하자면 현재 음악능력의 일정한 기준에 연관된 테스트에서 평점되는 증거를 뜻한다. 학생의 음악성취를 평가하는 테스트는 내용 관련 타당성에 대한 증거를 제공하여야 한다. 이러한 테스트는 교육 목표와 학습 경험의 내용을 반영하여야 한다.

대다수의 출판된 테스트들은 표준화되었다. 그들은 대표적인 규범적 집단에 의한 데이터를 제공하므로 교사는 그의 학생의 연주를 비슷한 연령이나 수준의 다른 학생들의 연주와 비교할 수 있다. 표준화된 테스트에는 신뢰도와 타당성에 대한 기술적인 증거가 제공된다. 수년에 걸쳐 많은 테스트가 개발되고 개선된 이후, 보통 테스트 편람이 그러한 데이터를 제공한다. 그러나 테스트 사용자는 기술적 데이터를 검색하여야 하며, 테스트의 타당성과 테스트가 사용될 목적에 대한 전체적인 적절함에 관련되는 판정을 해야 한다.

표준테스트는 개별 학생의 연주를 다른 학생과 비교할 경우에 특히 적절하다. 학생 선택의 토대로서 객관적 비교 데이터가 요구될 경우에는 표준 음악적성 테스트가 매우 유용할 것이다. 그러나 관심사가 음악성취의 평가에 있다면 표준 음악테스트는 그보다 적합하지 못하다.

성취테스트는 학생이 받았던 교육의 내용을 측정하는데 유효하므로, 성취테스트를 개발하거나 선택하는 사람은 그 교육 내용을 반영하여야 한다. 단순히 출판된 표준화한 음악성취 측정 선택하는 것은 특수한 음악 기술이나 지식을 의 성취를 평가하는 데는 적합치 못하다. 음악성취 평가가 일반 실재와 거리가 멀면 멀수록 규범연관 측정보다는 기준연관 측정이 나을 것이다.

표준연관 테스트는 한 학생이 무엇을 할 수 있는가의 용어로 학생의 성취를 기술하며 연주의 기준이나 절대적 표준에 대비하여 평가될 수 있다. 따라서 한 학생의 성취는 다른 학생의 연주와 비교하여 얼마나 잘 했나보다는 개별적 연주의 의미에서 평가된다.

표준테스트는 교사 제작 테스트에 비하여 다음과 같은 장점이 있다. 즉, 수차례에 걸쳐 수정 보완되었으며, 더 신뢰성이 있을 수 있으며, 기술적으로 더 우수하며, 규범을 제공하고, 예비 시간을 요하지 않는다. 교사 제작 테스트의 장점은 특별한 요구나 목적에 적합할 수 있으며, 믿을만한 표준테스트가 없는 영역의 내용을 평가할 수 있고, 일반적으로 비용이 적게 들며, 다양한 문항의 유형에 대한 더 많은 여유를 가질 수 있다⁴⁵⁾

교사 제작 테스트는 규범 연관 방법이나 기준 연관 방법, 양자에 의해서 분석될 수 있다. 한 학생의 성취를 다른 학생과 비교할 경우(예컨대, 그래프로 등급을 나타낼 때)에 교사는 규범

45) Lehman, op. cit. pp. 28-29.

관련 의미에서 성취를 평가하고 있는 것이다. 그러나 각 학생을 기준과 대조하여 평가할 경우 (주어진 과제에 의하여 작성될 수 있는)에 교사는 기준 관련 의미로 평가하는 것이다. 어느 접근법이 더 '좋은지'는 평가의 목적에 달려 있다.

2) 수험자

테스트 과제는 수험자의 보편적 발달 수준, 학문적 수준, 테스트 응시 경험에 부합되도록 고안되어야 한다. 어떤 한 테스트의 과제 구조가 너무 복잡할 경우에는 과제 구조의 높은 복잡성이 테스트가 평가하고자 하는 학생이 가지고 있는 음악 기술이나 지식의 표출을 저해할 수 있다. 아동들의 테스트 과제를 읽는 수준은 같은 연령이나 수준에 있는 학생들의 테스트에서 읽는 수준에 맞추거나 오히려 더 낮아야 할 것이다. 정답의 피드백을 통한 사례 문항의 자유로운 사용에는 학생이 주어진 테스트 과제를 이해하는 지를 확증하여야 한다.

테스트 과제상의 음악행동도 또한 수험자의 음악경험의 정도를 참작하여 고안되어야 한다. 특별히 주의를 기울여야 할 점은 수험자의 음악행동에 의한 경험의 보편적 정도 뿐만 아니라 테스트의 기능에 테스트 과제가 적합한지를 확증하는 일이다. 예컨대, 수험자의 음정 경험에 관하여는 매번 다른 가정이 있을 수 있으므로 음정 변별과 인지 과제에서는 몇 가지 다른 수준이 검증되어야 할 것이다. 학생들은 음정의 명칭에 관한 지식이 필요치 않은 단순히 청각적으로 연출된 두 음정이 '다른지' '같은지'를 응답할 수도 있으며 또는 음정의 명칭에 관한 경험을 요구하는 청각적으로 연출된 음정에 명칭을 붙이는 응답을 할 수 있거나, 기보된 음정에 관한 지식을 요하는 주어진 악보상의 낮은 음위에 청각적으로 연출된 음정을 기보하는 문제를 다룰 수도 있다.

3) 범위와 내용

대다수의 교사 제작 음악테스트가 성취를 평가하려는 경향을 보인 이후, 교사들이 교육 내용을 명확히 반영하는 지가 특히 중요해졌다. 이러한 과정에서 유의할 점은 표출될 행동, 행동이 표출될 조건, 수궁할만한 연주에 대한 기준 또는 기준들을 뚜렷하게 지정하는 규정된 지시 지침을 명확하게 설정하는 일이다. 테스트의 지시 지침은 학생들의 목표 달성을 촉진하기 위하여 만들어진다는 가정은 각 목표에 대한 테스트 과제 제작의 문제를 비교적 간단하게 해준다.

특히 테스트의 기능에 적합하고 균등한 행동 사례가 테스트에 반영되는 지는 매우 중요하다. 성취테스트의 경우에 있어서는 그 교육 목표와 그에 따르는 부수적인 중요성이 반영되어야 한다. 그와 같은 균형을 유지하지 못한 성취테스트는 신뢰도에 있어서 믿을 수가 없게 된다.

4) 능력

테스트는 가장 경제적인 방법으로 필요한 정보를 이끌어내도록 고안되어야만 한다. 대다수 테스트가 개인을 대상으로하는 표현 과제로 구성될 수 있으나, 이는 현실적으로 그와 같은 테스트를 위한 비실용적으로 지출될 실시 비용을 초래할 수 있다. 마찬가지로 집단으로 시험지와 연필로 실시되는 테스트도 기능한한 능력적으로 실행되도록 해야한다. 유효한 테스트를 입증하는 충분한 행동 사례를 취하기 위하여 얼마나 많은 절충을 해야 하며 실질적인 문제들을 고려해야 하는 지는 말하기 어렵다. 하나의 '완전한' 테스트라도 실시하는데 매우 비경제적이며 시간 소모적이라면 이는 결코 사용될 수 없는 비능률적인 것이다. 또 한편, 테스트가 의도하는 기능에 대한 정보를 산출해내는데 실패했다면 이 또한 비능률적이되고 만다.

5) 테스트 작성 원칙

음악 테스트 디자인과 행동의 결정을 위한 '원칙'으로서 위에서 몇 가지를 살펴보았다. 이러한 원칙들이 음악테스트의 발전이나 이 분야를 이해하고자 하는 이에게 도움이 될 수 있다.

Boyle은 음악테스트 디자인에 관한 원칙으로 다음의 일곱 가지를 제시하였다.

- (1) 음악테스트의 내용과 디자인은 "테스트는 서비스를 목적으로 한다"는 특수한 기능을 반영하여야 한다.
- (2) 음악테스트 내용의 범위와 특수성은 테스트의 기능에 달려있으며 그에 부합되어야 한다.
- (3) 음악테스트 내용은 일반적 발전 수준, 학문적 수준, 수험자의 테스트 경험을 참작하여야만 한다.
- (4) 인지 행동, 감정적 행동, 정신운동적 행동의 수준은 테스트 과제에서 테스트 기능에 부합되어야 하며, 수험자의 음악적 행동을 통한 경험의 일반 수준을 고려하여야 한다.
- (5) 음악테스트는 테스트 기능과 연관된 행동의 균형잡힌 표본을 제공하여야 한다.
- (6) 음악테스트는 능률적이어야 한다.
- (7) 음악테스트는 기술적으로 적합하여야 한다.

또한 그는 음악테스트 행위결정에는 다음의 세가지 원칙을 들었다.

- (1) 테스트 과제에 요구되는 반응 행동은 가능한한 직접적으로 지식과 기술 또는 가치있는 것이 평가되도록 반영하여야 한다.
- (2) 가능하든지 불가능하든지 간에 음악적 행동의 평가는 청각적 음악 자극에 대한 응답을 포함해야 한다.
- (3) 테스트가 요구하는 응답 방법론은 아무런 상관이 없는 기술과 지식의 영향을 과소 평가해야 한다.

2. 지각과 개념 과제

지각과 개념 음악과제는 한 개인이 수용하고 식별하거나 음악반응을 진행할 수 있는 범위를 평가하는데 주력한다. 테스트 제작자들이 주로 관계해온 과제들은 넓게 세 가지 방향으로 분류된다. 청각 변별, 청각 인식, 청각-시각 변별이 그것이다. 청각 변별과제는 보통 학생들에게 청각적으로 연출된 둘 또는 그 이상의 음악 자극들의 차이를 변별하는데 반하여, 청각 인식 과제는 청각적으로 연출된 음악 자극의 특성이나 다양한 성질을 감정하는 것을 강조한다. 청각-시각 변별과제는 일반적으로 악보를 통하여 소리로 들리는 것과 연상시키는 학생들의 능력을 평가하기 위하여 제작된다.

지각과 개념 과제를 위한 음악 자극들의 정도는 단순한 것부터 복잡한 것, 개성을 발휘한 특유한 것일 수도 있으나 선율 형태, 리듬 형태, 화성 형태의 음향적 자극으로 구성되된다. 또한 초점을 좁게 또는 폭넓게 둘 수 있으며, 자극적으로나 반복적으로 연출될 수 있고, 다양한 음가를 취할 수 있다. 자극의 변화는 과제의 난이도에 따라서 과감하게 변경시킬 수 있다. 마찬가지로 지각 과제들도 청각 자극들을 변별하는데 주력하는 반면에, 개념 과제는 청각 자극의 분석, 기억, 그밖의 다른 개념적 작용에 초점을 맞춘다.

음악테스트의 대다수의 과제들은 둘 또는 그 이상의 선택에서 '정확한' 또는 '가장 나은' 응답을 도출해내려고 고안되었다. Webster⁴⁶⁾는 최상의 응답에 대한 여러 개의 가능성과 '집중적' 고려 사항으로 구성된 기본적으로 수렴성 사고를 요하는 접근방법에 주목하였다. 또한 그는 부과된 과제에 대한 진전된 많은 가능한 해결들로 구성된 분산적 사고 과제도 만들었으며, 특히 음악적성을 평가하는 유력한 사례를 만들었다⁴⁷⁾. 여기서 그는 음악 창조성의 중요한 세 가지 관점으로 대규모, 유연성 및 독창성을 들었는데, 이는 명백하게 분산적 사고와 연결된다. Webster의 입장은 음악능력의 평가에 대한 하나의 흥미있는 새로운 전망을 제공하지만 이 분야의 전문가들은 여전히 음악능력의 지각과 개념적 차원의 평가를 기본으로하는 수렴성 사고를 요하는 과제를 적용하고 있다. 따라서 본 논문에서는 이에 관하여 살펴보고자 한다.

1) 청각 변별 과제

대부분의 청각 변별 과제들은 학생들에게 두 개의 연속해서 표현된 청각 자극들의 차이를 인지하도록 요구한다. 위에서 언급된 바와 같이 자극들은 몇 개의 음이나 조성형, 또는 악곡이나 인용구로 구성될 수 있다. 과거의 음악테스트의 가장 통상적인 자극들은 한 쌍의 음으로 구

46) op. cit.

47) op. cit., p. 186.

성된 테스트 과제로서 음고의 조성적 속성, 강약, 음색, 음가를 변별하는 것이었다. 최근에 와서 음악테스트 과제는 일반적으로 선율과 리듬의 형태를 변별하는 데 주력하고 있다.

조성적 속성을 위한 청각 변별 과제는 주로 '똑같다' 또는 '다르다'와 같은 유형의 응답을 요구하는데 이는 기본적으로 일단계 과정이며, 어떤 변별과제들은 부수적 단계가 첨가된 것도 있다. 예컨대, 두 개의 음정이 단순히 '같은지' '다른지'를 확정하는 과제는 두번째 음이 첫음보다 '더 높은지' '더 낮은지'를 확정하기, 또는 첫음의 높이와 둘째음의 높이가 얼마나 차이가 나는지를 변별하기보다 훨씬 쉽다. '똑같다'거나 '다르다'라는 응답보다 더 많은 것을 요구하는 변별 과제의 구조는 난이도를 크게 증폭시킨다.

음악 변별과제형 특히, 선율형을 변별하는 과제들은 여러 가지 이름으로 불려졌다. 리듬형 변별과제들은 연속적인 연출에 대한 기억을 필요함에도 불구하고 일반적으로 리듬 변별과제라고 한다. Gordon의 유명한 MAP(1965) 테스트의 리듬 이미지 부분은 각각 '템포'와 '박자'로 명명되었으며 리듬의 각기 주어진 차원에만 초점이 맞추어졌으며 학생들에게 선율적 전후 관계에 불구하고 단순히 리듬 형태가 '똑같다' '다르다'만을 가리키도록 되어 있다.

그러나 선율형을 변별하는 능력을 평가하는 대부분의 접근 방법들은 기억 테스트의 일종으로서 보통 조성 기억이나 선율 기억 테스트로 불려졌다. 이러한 접근 방법들이 일반적으로 널리 사용되는 유형이다. 조성 기억이라 불리는 유형은 학생들에게 보통 두번째 연출된 조성적 씨퀀스에서 어느 음이 변화되었는지를 묻는다. 연속 씨퀀스의 음의 숫자는 셋에서 열개 사이로 이루어질 수 있다. 변화된 특별한 음을 알아내기는 단순히 선율형이 같은지 다른지를 가리키는 것보다 더 어려운 과제이다. Seashore⁴⁸⁾, Wing⁴⁹⁾, Bentley⁵⁰⁾의 테스트들은 부속테스트가 '기억', '음가기억', '조성기억'이라는 제각기 다른 명칭이 붙었음에도 불구하고 이 접근방법을 적용하였다. Colwell⁵¹⁾의 음악성취테스트 Music Achievement Tests(MAT)는 이러한 변별과제에 변형을 가하였다. 한 코드가 연출되고 이어서 arpeggio로 연주되어서 한 음이 변화될 수도 있고 변화되지 않을 수도 있다.

Drake⁵²⁾와 Gaston⁵³⁾의 과제들은 다른 접근 방법으로 선율형을 변별하도록 요구한다. 그들은 수험자가 씨퀀스 연주중의 부속 씨퀀스(여섯 개 이상)를 비교하도록 하는 것에 반대되는 모델로서 음의 씨퀀스를 제공한다. 씨퀀스 연주는 조성, 음표, 또는 리듬의 용어에서 모델과 동일할 수도 있으며 차이가 날 수도 있다. Drake의 과제는 '음악 기억 musical memory'테스트

48) op. cit. 1939.

49) op. cit. 1961.

50) Bentley, A., Measures of musical abilities, London, George G. Harrap & Co., 1966.

51) Colwell, R., Music Achievement Tests, Chicago, Follett Educational Corporation, 1969-70.

52) Drake, R. L., Drake Musical Aptitude Tests, Chicago, Science Research Associates, 1957.

53) op. cit.

라 부르며 Gaston의 과제는 '조성 기억tonal memory'테스트라 한다. Seashore, Wing, Bentley의 테스트보다 이 두 테스트를 '기억' 테스트로 만드는 것은 수험자가 몇 가지 방법중에 하나에 의해서 변화될 수 있는 일련의 버전을 듣고 원형 모델을 기억하게 하는 것이다.

1965년에 만들어진 Gordon의 MAP의 부속테스트 '조성 이미지'는 선율변별과제의 확대로 보인다. 그의 선율변별과제는 모델선율의 장식된 버전이나 다른 선율로 이어지는 하나의 모델 선율을 제공한다. 테스트는 장식(첨가된 음들)이 이동되었을 경우에 두번째 선율이 첫번 것과 '같은지' '다른지'를 결정하는 것이다. Gordon은 이 과제를 '화성' 테스트로 확대하였는데 이는 기본적으로 선율 테스트 과제와 동일한 유형이다. 그러나 장식된 버전은 저성부는 첼로가, 상성부는 바이올린이 연주한다. Gordon의 선율변별과제는 명백하게 두 선율의 변별하는 단순한 과제보다 더 복잡하며, de la Motte-Haaber⁵⁴⁾는 고도의 인지적 기능과 심미적 판단을 거의 상실하고 있을뿐만 아니라 객관적인 측정의 한계를 보이고 있다고 지적하였다.

청각변별과제들은 예견도구로서, 교육성과의 측정으로서 음악교육에 유용하게 보인다. Colwell의 음악성취테스트MAP⁵⁵⁾는 내용이 몇개의 기초적 음악 씨리즈의 교육적 객관성을 토대로 하며, 음고, 음정, 박자, 장단조 모드의 변별 평가를 제공해준다.

2) 청각 인지 과제

이런 성격의 과제들은 대부분 흔히 성취테스트 문항으로 사용된다. 기본적으로 청각 인지 과제는 수험자에게 그들이 이전에 우연히 마주친 음악 자극의 일정한 특성이나 질을 동일시하도록 요구한다. 이러한 마주침은 주로 테스트 시간이거나 테스트 형식의 한 부분이었을 수도 있다. 그 이전에 경험한 유형의 마주침은 공식적인 교육 기간에 있었을 수도 있다. 그와 같은 인지 과제의 사례는 음정이나 악기의 음색을 동일시하는 지시일 수 있다. 청각적으로 연출된 음정이나 악기의 소리와 연관되는 이름이나 호칭이 붙은 지시에 따라서 학생들이 각기 음정이나 악기의 호칭을 인지할 수 있는지를 결정하는 것을 테스트한다.

다른 유형의 인지 과제는 테스트의 한 부분으로서 하나의 청각 모델의 연출로 이루어지며 수험자에게 이를 선택하거나 동일시하라고 요구하는데, 보통 이를 다시 듣고 셋이나 넷중에서 택일한다. 변별과제의 초점이 자극들의 '차이'를 묻는데 반하여, 위의 두 가지 인지 과제의 초점은 사전에 들은 음악 자극들의 '기억'과 '동일시하는 것'이다.

음악 청각 인지 과제의 정도는 실제로 한정이 없다. 그러므로 평가자는 학생들의 평가에 사용되는 청각 인지 과제가 적절하고 적절한 복잡성 수준인지를 확증하는 모든 노력을 기울여야

54) op. cit., p. 279.

55) op. cit. 1969-70.

한다. Colwell(1969-70)의 MAT와 같은 인지 과제가 몇개의 인지 부속 과제로 구성되었으나 교사와 연구자들이 그들의 평가 목적에 부합되는 청각 인지 과제를 발전시킬 필요가 있는 것으로 보인다.

3) 청각적 구조화 과제

비록 단 한 사람의 학자 Kai Karma⁵⁶⁾(1983, 1985)의 이론과 연구에 기초하는 것이지만, 청각 구조화auditive structuring는 음악능력 평가를 위한 잠재력을 지닌 것으로 보인다. 그것은 아마도 전통적인 변별과제와 인지 과제보다 더 음악 지각과 인지의 현대 이론들과 더 일치되는데 있다. 그의 이론에 의하면 음악적성을 '음향적 소재를 구조화시키는 능력'⁵⁷⁾이라고 정의하였으며, Karma의 청각 구조화 과제는 기본적으로 수험자에게 '연속적으로 반복된 것에서 작은 동기 하나를 찾아내라고 요구한다'⁵⁸⁾. 수험자는 세번에 걸쳐 연속적으로반복되는 동기와 비교하여 한 부속 씨퀀스 동기subsequent motif가 '똑같은지' '다른지'를 가리켜야 한다. 앞에서 서술한 바와 같이 오직 한 조성적 차원만이 원형 동기에서 변화된다(한 음의 높이와 음가 또는 강약). 하나의 씨퀀스의 음의 숫자는 변경될 수 있으나 대개는 셋이나 넷 또는 다섯 음으로 구성된다. Karma는 학생들이 둘에서 여섯 음으로 구성된 씨퀀스를 할 수 있다고 제안하였다.

Karma는 '청각적 소재의 씨퀀스 구조화로 이루어진 몇 개의 인지 작용의 존재에 대한 실증을 제시하였다: 예상의 형식화, 인지, 형태gestalt를 통한 구조화, 예상의 변경, 타이밍, 우세한 형태의 내적 구조의 분석'⁵⁹⁾. Karma는 한 개인이 이러한 작용을 소리의 구조화에 사용할 수 있는 범위가 음악적성의 반영이라고 주장하였다. 그외에도 테스트 과제는 일반(비음악적) 지능의 효과를 최소화하고 어린 아동들의 사용에 불편이 없어야 하며, 음악적성에 대한 교사의 기악평가와 밀착되어야 한다고 그는 주장하였다⁶⁰⁾.

4) 시청각 변별 과제

시청각 변별 기술은 듣기와 보기의 상호작용이다. 대다수의 일상형식은 기보, 선율, 리듬 및 화성받아 쓰기에서 잘못을 발견해내는 것이다. 받아쓰기 과제가 들은 바를 기억하고 악보에 기록된 것을 유지하는 첨가된 단계를 요구한 이후부터는 받아쓰기 과제는 어느 면에서는 시청각

56) op. cit., 1983; op. cit., 1985.

57) op. cit. 1985, p. 4.

58) ibid., p. 30.

59) op. cit. 1985, p. 11.

60) op. cit. 1983, p. 31.

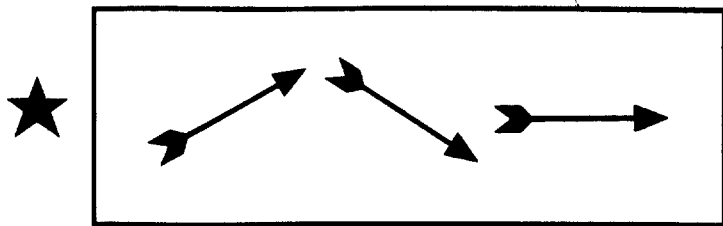
표현법에 적용되는 것을 능가한다. 악보의 사용으로 일정한 학습은 시청각 변별과제를 완성하는데 필요하다. 그러한 과제는 대개 학생들의 성취를 평가하고 장단점을 진단하는데 적성 테스트 과제보다 더 유용하다.

전형적인 시청각 변별과제는 하나 또는 두 개의 예가 견본과 동일한지 다른지를 가리는 음악 사례의 악보와 사례의 반주 레코딩으로 구성되어 출판된 테스트에서 찾아볼 수 있다. 수험자는 녹음된 버전과 다른 부분을 악보에 표시하여야 한다. 일반적으로 그러한 과제는 불일치가 생길 수도 있는 리듬, 선율 또는 화음의 악보에 관한 관점을 특수화한다. 이 과제의 변형 다문항 선택형은 수험자가 들은 것을 선택해야만 하는 몇 가지 양자 택일과 견본에 ‘굵은 종선’을 표시하는 것이다.

다른 시청각 변별과제도 마찬가지로 사례의 기보된 버전과 귀로들은 버전 사이에 생길 수 있는 불일치의 유형이 나열된 몇 가지 양자 택일과 악보에 한 선을 표시하는 다문항 선택형 형식을 사용할 수 있다. 전형적으로 선택은 ‘음가 오류’, ‘리듬 오류’, ‘음가와 리듬의 오류’, ‘오류가 없음’으로 구성될 수 있다.

시청각 변별과제는 실용을 감안하여 난이도와 기교면에서 아주 폭넓은 수준으로 고안될 수 있다. 어린 아동들에게 청취한 사례의 선택이나 그림으로 나타낸 악보의 분류를 요구하는 과제들은 시청각적인 상징의 연상 작용을 포함한다. ‘예비 악보 표상prenotation icon’은 선율적 지시, 소리로 들리는 음의 숫자, 선율의 형태, 속도 또는 강약의 변화와 같은 음악 소리를 시각적으로 일정한 상황을 표현한다. 예컨대 Simons⁶¹⁾은 초보 연령의 학생들에게 한 선율이 움직이는 방향(상진/하진/동일)에 따라 화살표에 원을 그리라고 지시하였다.

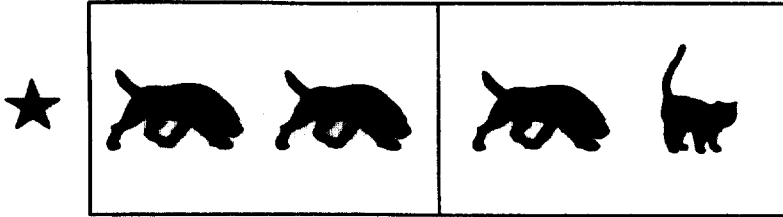
상진 선율/하진 선율/동일 선율



(예 2) Simons의 음악 듣기기능 측정의 ‘선율’ 응답 형식

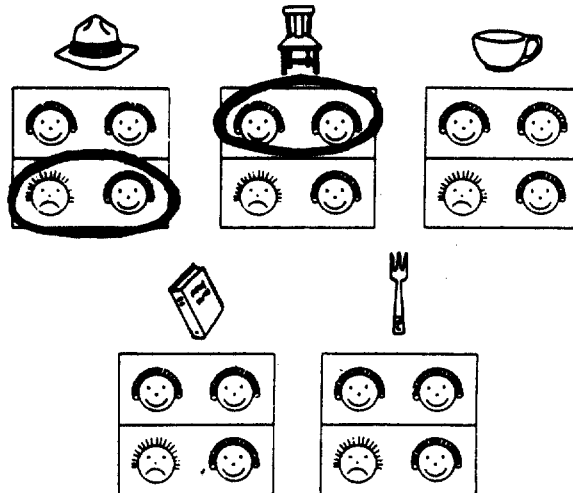
61) Simons, G. M., Measurements of musical listening skills for young children, Chicago, Stoetling, 1974.

조성 형태와 비교해볼 때에 그는 두 마리 개와 한 마리 고양이와 개 한 마리를 예시하였다. 수험자는 들은 두 형태가 같을 경우에는 두 마리 개에, 다를 때는 한 마리 개와 고양이에 원을 그려야 한다.



(예 3) Simon의 '형식과 조성 형태' 응답모형

아직 글을 읽지 못하는 5세 아동에게 적용할 수 있는 Gordon의 PMMA⁶²⁾는 아동들이 조성이나 리듬의 형태가 동일할 경우에는 한 쌍의 웃는 얼굴에, 다른 형태일 때는 웃는 얼굴 하나와 눈살을 찌푸린 얼굴 하나가 함께 있는데 원을 그리게 하였다. 응답과제 위에 주어진 모자, 의자, 책과 같은 그림으로된 심볼은 과제의 번호와 유사한 것이다. 이러한 심볼은 각 문항이 주어질 때마다 진술된다(예 4). 그와 같은 과제를 구분해주는 주요 관심사는 아동들의 성숙과 교육 그리고 음악경험의 정도에 따른 적절하고 의미있는 상징을 확증한다.



(예 4) Gordon의 PMMA의 '조성 테스트'의 응답모형

62) op. cit. 1979a.

다른 한편으로는 기교의 극치를 보이는 것으로서 악보 읽기 연구에서 오류발견 과제가 있다. 이러한 과제들은 수험자들이 기보된 버전과 들은 버전 사이의 차이를 표시하는 것으로서 사례가 두서너개 선율의 악보부터 총보에 이르는 단계를 이루고 있다. 오류의 유형과 예민성은 이러한 과제에서 크게 변화될 수 있어서 청취자의 경험과 기교의 극치 그리고 테스트의 목적에 달려있다. 악보의 상황과 소리로 들리는 모든 상황 사이의 오류를 찾아내는데 연관된 안내자가 과제와 더 가깝게 접근할 수록 과제가 더 어려워진다. 총보 읽기 기술의 발전에 관한 연구의 한 부분으로서 오류 발견 테스트를 발전시킨 학자로는 Costanza⁶³⁾, DeCarbo⁶⁴⁾, Grunow⁶⁵⁾ 및 Ramsey⁶⁶⁾가 있다.

아마도 변별과제보다 더많은 인지 과제로 이루어진 다른 유형의 시청각 기술은 음악적 주제, 작품, 또는 악보의 양식의 인지로 구성되었다. Gordon⁶⁷⁾은 그러한 기술을 음악을 소리없이 듣는 즉, 소리로 들리지 않는 청취의 한 유형이라 부른다. 본질적으로 청각적 심상을 내포하기 위하여 과제는 수험자가 기보된 사례가 어떻게 소리나는지를 상상하는 그들의 능력을 반영함으로써 긴 기간의 기억의 유형을 요구한다. 그리고 나서 주어진 주제나 작품 또는 사례의 유형적 또는 다른 음악적 특성과 연관시킨다. 이러한 과제의 난이도는 마찬가지로 주제와 작품이나 유형적 또는 다른 특성이 이름의 명단에 의하여 선택되는지 또는 장기간 기억에 의해 회생되는지에 따라 변화될 것이다.

5) 음악 감수성 선별 과제

심미적 판단을 위주로한 다음의 세 가지 출판된 테스트는 주어진 음악 실예의 두 가지 연출 사이의 연주의 질을 변별하는 문항과 어떤 버전이 '더 나은지'에 연관되는 하나의 정해진 선별 반응으로 구성된다. 이들 테스트의 기본 과제는 동일한데도 그 명칭이 각기 '감상 부속테스트 appreciation' subtests(The Standardized Tests of Musical Intelligence)⁶⁸⁾, '음악 감수성 부속테스트 musical sensitivity' subtests (MAP),⁶⁹⁾ '변별 테스트 discrimination'

63) Costanza, A. P., Programmed instruction in score reading skills, Journal of Research in Music Education, 1971, 19, pp. 453-459.

64) DeCarbo, N. J., The effects of conducting experience and programed materials on error detection scores of college conduction students, Unpublished doctoral dissertation, Kent State University, Kent, 1981.

65) Grunow, R. F., An investigation of the relative effectiveness of four modes of score preparation on visual-aural discrimination skills development, Unpublished doctoral dissertation, University of Michigan, Ann Arbor, 1980.

66) Ramsey, D., Programed instruction using full-score band literature to teach pitch and rhythm error detection skill to college music education students, Unpublished doctoral dissertation, University of Iowa City, 1978.

67) op. cit., 1987, pp. 13-15.

68) Wing, op. cit., 1961.

test(Indiana-Oregon-Music- Discrimination-Test)⁷⁰⁾ 따위로 각기 다르게 불리워지고 있다.

특히 문화적인 음악평가를 목적으로하는 Wing의 부속테스트⁷¹⁾들을 각각 리듬의 강세, 화성, 강도 및 프레이징 감상테스트로 호칭을 붙였으며, '연주에 대한 감수성'을 측정하기 위하여 고안되었다. 감각적인 변별력을 측정하는 Wing테스트는 그 음악실예가 구식으로 들리는 점이 없지 않으나 오늘날까지도 사용되고 있으며 다른 모든 음악성테스트처럼 일정한 시간내에 집단 형으로 실시된다. Gordon 테스트의 한 부분인 음악감수성 'musical sensitivity'은 프레이징, 밸런스, 스타일이라는 세 개의 부속테스트로 구성되었다. 밸런스 부속테스트는 기본적으로 두개의 인용구의 종결 악구의 타당성과 어느 것이 더 나은지를 결정하는 사이의 변별을 요구한다. 스타일 부속테스트는 본질적으로 인용구의 두 가지 연출에서 변화된 차이로서 템포를 사용한다. Oregon테스트도 마찬가지로 음악실예의 두 가지 연출중에 어느 것이 더 나은지에 관한 판단을 요구한다. 그에 더하여 수험자는 그러나 리듬, 선율, 화성중의 하나가 변화되었는지도 표시하여야 한다(물론 음악 사례의 두 가지 연출은 '똑같지 않다'). 이것이 이 과제를 약간 차이나게 만들며, Wing이나 Gordon의 테스트보다 더 복잡하게 보이게 한다. 또한 한 선율에 대하여 어느 것이 더 좋은지 나쁜지에 대한 오학년생에 의한 심미적판단은 결실하다기 보다는 우연적이며 아동의 연령이 높아질수록 신뢰성도 높아진다. Wing이나 Long의 테스트는 친숙한 선율을 사용하는데 비하여 Gordon의 감수성 테스트의 문항은 친숙하지 않은 선율이며, 음악지식을 전제로 하는 감각적 판단력과 기억력에 치중하여 Bentley 테스트와 유사하다는 지적을 받고 있다⁷²⁾.

기본적으로 이러한 테스트들의 각 문항에서 '더 나은' 버전은 '전문가들'의 좋은 연주 실기나 전통적인 서양 예술 음악의 뼈대 구조에 대한 해석을 반영한다. 음악적 경험이 있으며 음악에 관한 지식이 있는 수험자는 그러한 경험이 없는 사람보다 이 테스트에서 더 좋은 점수를 낼 수 있다. 실제로 이 측정 방법은 성취도 평가에 사용될 수 있다.

6) 테스트 실시의 새로운 방식

앞장에서는 변별, 인지, 음악의 다양한 상황에 대한 감수성 평가에 대하여 서술하였으나 학생에게 테스트 과제를 실시하는 방식에 관하여는 거의 언급되지 못하였다. 대개 지금까지의 테

69) Gordon, op. cit., 1965.

70) Long, op. cit., 1965.

71) Tests of musical ability and appreciation, British Journal of Psychology Monographs, Supplement, No. 27(2nd ed.), London, Cambridge University Press, 1968. p. 83.

72) de la Motte-Haaber, op. cit., p. 279

스트 과제들은 녹음된 음악 자극에 의한 집단형으로 실시되었으며, 요약어와 둘 또는 그이상의 선택중에서 적절한 응답을 가리키는 검증 표시나 번호를 요구한다. 답지는 손으로 또는 전자기기에 의하여 기록될 수 있다. 이것이 음악테스트 역사를 통해 그와 같은 기술과 능력을 검사하는 가장 통상적인 테스트 실시 방식이다.

컴퓨터 기자재 교육의 발전과 현실적인 접근 용이성은 음악에서 학생의 성취도를 항구적으로 평가할 수 있는 '컴퓨터에 의한 개별 상호 작용 테스트 실시 individualized interactive testing'라는 새로운 실시 방식이 나왔다. 여러 가지 변별과제 대다수는 이제 새로운 체제로 만들어지고 학생은 헤드폰과 컴퓨터 화면을 통하여 테스트를 받게 할 수 있다. 일반적으로 학생은 간단한 키보드 반응으로 응답한다. 가장 통상적으로 사용되는 컴퓨터 소프트웨어 프로그램에서는 컴퓨터가 각기 학생의 응답한 기록을 보관한다. 많은 프로그램들이 일정한 '하위branching' 처치 방법에 의한 교육 프로그램을 개별화하는 응답 데이터를 사용하고 있다. 본질적으로 그러한 처치 방법들은 '정확한' 응답을 수반하는 더 어려운 교육 프로그램으로 이동하거나 '부정확한' 응답을 수반하는 덜 어려운 과제를 수반하는 교육 프로그램으로 이동하게 해준다.

그러한 하위 처치방법은 개별학생을 위한 더 효율적인 프로그램을 만드는데 공헌한다. 이와 유사한 처치 방법이 '인용테스트 실시adaptive testing'라는 이름으로 테스트 실시에 적용되었다. 이러한 테스트는 과거에 실행된 개인의 응답에 일치하는 각기 성공적인 문항의 난이도를 적용한다. '정답'은 더 어려운 문항으로 구성되는데 반하여 '오답'은 덜 어려운 문항으로 이루어진다. 컴퓨터는 즉각적인 채점을 가능하게 해주며 문항 응답 이론에 일치되는 복잡한 산술적인 계산에 의한 각각 부속 문항의 선택을 해준다. 알려진 난이도와 변별 수준의 커다란 문항풀이는 '인용 테스트' 실시에 필요하다.

Vispoel⁷³⁾은 Seashore의 '조성 기억' 부속테스트와 과거에 개발된 278개 문항 인용 테스트풀로 구성된 다른 세 개의 전통적인 유형의 테스트에서 조성 기억을 위한 '인용 테스트'의 측정의 정확성, 효율성 그리고 타당성을 비교하였다⁷⁴⁾. 468명의 고등학생과 대학생의 응답으로 이루어진 그의 데이터는 인용 테스트가 34-69퍼센트의 더 적은 문항을 사용하는 전래의 테스트들보다 더 측정의 정확성면에서 우수하다는 것을 나타내었다. 테스트의 신뢰도와 타당성에서도 마찬가지로 전통적인 테스트를 능가한다.

많은 컴퓨터 기자재 교육 프로그램이 필수 부분으로 되어있는 이러한 테스트 실행 방식 외에도 Seashore의 '피치' 부속테스트⁷⁵⁾, Drake의 Musical Aptitude Test⁷⁶⁾, Gordon의

73) Vispoel, W. P., Computerized adaptive music tests: A new solution to three old problems, Paper presented at the Music Educators National Conference, Washington, 3. 1990.

74) Vispoel, An adaptive test of musical memory: An application of item response theory to the assessment of musical ability, Unpublished doctoral dissertation, University of Illinois, Urbana, 1987.

Primary Measures of Musical Audiation⁷⁷⁾과 같은 여러 개의 전통적인 테스트가 마이크로 컴퓨터 기자재를 개별 테스트실시에 적용하였다. McCarthy와 Robinson이 컴퓨터화한 버전의 개별화, 피드백, 기록 보존에 대한 우수성을 간파하였으나 원본 테스트와 컴퓨터화한 테스트 사이에 상호 관계의 증거를 제시하지는 못했다. Gordon의 PMMA에 컴퓨터화한 Forsyth의 인용 테스트는 PMMA(85에서 88)와 깊은 상호 관계를 나타내었다.

음악 지각과 인지를 위한 테스트 실시의 모델로서 컴퓨터의 잠재력은 아직 유년기에 있으나 최근의 저널 심리음악학Psychomusicology의 특수 논점⁷⁸⁾에서 몇 개의 논문이 음악 지각과 인지의 연구에 마이크로 컴퓨터가 어떻게 사용되어왔으며, 또 사용될 수 있는 지에 대한 전망을 제시하고 있다. 이에 관한 더 자세한 내용은 Amus⁷⁹⁾, Dowling⁸⁰⁾ 및 Gibson⁸¹⁾의 논문을 참조하기 바란다.

3. 연주 과제

일반적으로 음악연주의 개념은 '복합적이며 다차원적인 행동'으로 이해되며, 평가자의 주관성, 음악활동 자체의 정의의 부족 및 음악 행동의 합성 성분에 대한 일관된 정의로 인하여 평가가 쉽지 않았다. 연주평가는 모든 성악이나 기악레슨 또는 앙상블 리허설에서는 필수 부분이며, 그와 같은 연주평가의 대다수가 주관적이며, 의식적이던 무의식적이던 개별 평가자가 중요하다고 생각하는 특별한 기준을 반영한다. 주관성은 언제나 연주 평가의 중요한 부분이지만 연주평가의 정확성은 과정에 더 객관성을 가져다 주는 확실한 절차를 적용함으로써 크게 진전되었다.

다음에서는 연주평가의 객관화된 접근 방법, 연주 평가의 광역 접근 방법, 연주 평가의 특수 접근 방법, 평가와 관련된 연주의 중요성 및 연주 평가에서 전자 분류의 역할로 나누어 고찰하여 보고자 한다.

75) McCarthy, J. F., The itch test: Creative computing, 1984, 10(3), pp. 211-212, 216-217.

76) Robinson, R. L., Microcomputer adaptation of the Drake Musical aptitude Tests, Gainesville, University of Florida, 1985.

77) Forsythe, R., The development and implementation of a computerized preschool measure of musical audiation, Unpublished doctoral dissertation, Case Western Reserve University, Cleveland, 1984.

78) Vol., 8, Autom 1989.

79) Amus, E. P., computer-based modeling of music concepts for testing, evaluation, and refining theory, Psychomusicology, 1989, 8, pp. 171-182.

80) Dowling, W. J., Programming small computers to produce experiments in music cognition, Psychomusicology, 1989, 8, pp. 183-190.

81) Gibson, D., An effective computer-assisted protocol for music perception experiments, Psychomusicology, 1989, 8, 191-196.

1) 객관적 연주 평가

연주평가는 필수적으로 어떤 표준에 의한 주어진 연주에 대한 판정을 할 평가자를 요한다. 이상적 극치와 다른 사람의 연주수준 또는 교육적 객관성에 대한 일정한 특수 기준이 그 표준이 될 수 있다. 평가과정에서 객관성 또는 '공정함'을 확고히 하기 위하여 가장 일반적으로 수용되고 있는 절차는 평가될 연주에 대한 명확한 기준을 세우는 것과 각 연주자가 갖는 평가 기준의 범위를 가리키는 평가의 잣대의 일정한 유형을 사용하는 일이다.

연주기준은 전통적인 오디션의 연주가 그 총체성으로 평가되는 '광역 접근 방법'과 평가자가 연주의 선택된 상황에 초점을 두는 '특수 접근 방법'에 따라 달라질 수 있다. 연주의 초점을 광역 상황에 둘 것이냐 특수 상황에 둘 것이냐는 평가의 목적이나 평가자의 철학적 전망에 달려 있다. 많은 경우에 평가자가 광역과 특수 접근방법을 혼합하여 사용한다.

라이커트 등급Likert scale이 사람과 사람의 근원적으로 연속되는 행동을 평가 하는데 가장 일반적으로 사용된다. 연속은 넓은 범위일 수도 있고 좁은 범위일 수도 있으나 이에 대한 보편적인 점수는 다섯에서 일곱 점이다. 점수는 '매우 일치함', '일치함', '중간이다', '일치하지 않다' 또는 '매우 일치하지 않다'와 같이 서술적인 호칭으로 표현될 수도 있다. 동그라미를 칠 번호나 체크할 빈칸에도 마찬가지로 라이커트 등급의 방식이 많이 사용된다. 라이커트 등급의 응답은 통상적으로 한 연주나 연주의 한 특수 상황에 관한 일정한 진술에 대한 판정을 반영한다. 전형적인 진술은 다음과 같은 4차원 연속체continuum에 의해 평가될 수 있다⁸²⁾.

	일 치			불일치
1. 이 연주는 질적으로 우수하다	_____	_____	_____	_____
2. 리듬은 정확하게 연주되었다	_____	_____	_____	_____
3. 인트로이션은 정확하였다.	_____	_____	_____	_____
4. 프레이징은 매우 '음악적'이다	_____	_____	_____	_____

제 2, 3, 4 진술이 특수 접근 방법인데 반하여 제1번 진술은 광역 접근 방법이다. 연주의 몇 가지 특수 상황에 대한 등급 평가는 총체적 연주 등급을 예비하기 위하여 요약될 수 있다. 라이커트 등급은 광역 접근 방법보다는 특수 접근 방법에 더 적절하게 보이나, 비교적 많은 숫자의 문항이 필요하고 라이커트 등급에 대한 신뢰성을 확립하는 판단이 있어야 한다.

82) Boyle, J. David, op. cit., pp. 258f.

의미를 측정하는 시도에서 발전된 '어의적 차등semantic differential'⁸³⁾은 한 셋트의 반의어 사이의 연속체상에 연주의 일정한 상황을 차등 평가하는 데 응용되었다. 라이커트 등급과 마찬가지로 반의어 사이의 연속체상의 점수는 변할 수 있으나, 절차는 비슷하다. 음악연주를 기술하는 반의어 셋트 또는 반의어구는 다음과 같다.

	1	2	3	4	5	
음조를 유지	_____	_____	_____	_____	_____	음조를 벗어남
좋은 음질	_____	_____	_____	_____	_____	빈약한 음질
리듬믹	_____	_____	_____	_____	_____	아리드믹
음악적	_____	_____	_____	_____	_____	비음악적
민감함	_____	_____	_____	_____	_____	민감하지 못함
일관적	_____	_____	_____	_____	_____	일관되지 못함

어의적 차등 데이터는 라이커트 등급에 의한 캔 데이터로 통산될 수는 없다. 간단한 '연주 윤곽'이 많은 평가 목표를 충족할 수도 있으나, 어의 차등의 극단적인 적용은 점수와 상관되는 기본적인 주요 요소 또는 차원을 결정하는 분석 요인을 필요로 한다.

비슷한 정도로 높은 수준의 몇 개의 연주에 등급을 매길 필요가 있을 경우나 몇 명의 아주 우수한 연주자가 모두 협주곡과 같은 동일한 작품을 연주할 때, '한쌍 비교법paired-comparisons technique'은 개인 연주의 전통적인 오디오션 보다는 크게 평가의 정확성을 기할 수 있다. 이 방법은 개별적으로 연주를 듣고 등급을 평가하는 것보다 더 성가시고 시간을 요하는 것이지만 한 쌍의 연주를 체계적으로 비교할 수 있게 해준다. 모든 연주가 녹음되고 한 쌍씩 서로 비교되며 선택된 한 쌍중에서 어느 것이 더 우수한 지를 가리는 몇 가지 판정이 내려진다. 비슷한 정도로 높은 수준의 연주들을 비교하는데 내포된 모호한 표현이 있다는 가정과, 각 쌍 사이의 차이의 판정은 일반 분류에 일치되는 데에 따라 변화된다는 가정은 실질적 차이에 치중한 것이다. 이에 관한 더 상세한 절차는 Boyle 및 Radocy⁸⁴⁾의 것을 참조하기 바란다.

많은 수의 연주를 비교하여 '한쌍 비교법'이 비능률적일 경우에는 '연계 차등 successive intervals' 방법이 유용할 것이다. 이 평가 방법에서는 판정자는 연주의 등급을 정하기 보다는 기준을 정한다. 오로지 비교적인 판정만을 위하여 판정자의 기준은 일반 분류에 일치하게 변경

83) Osgood, C. E., Suci, G. J., & Tannenbaum, P. H., The measurement of meaning, Urbana, University of Illinois Press, 1957.

84) op. cit., 1987, pp. 180-183.

될 것이 추정된다. 이 방법은 각 기준마다 그에 따른 응답 숫자의 차이로 이루어진다. 여기에는 다섯 가지 기준이 사용되는데 예컨대, 페스티벌 연주를 차등 평가하는 기준, 한 기준에 대한 각 연주를 담당할 판정자의 비율 정하기, z점수화로 환산, 실질 기준 영역의 확립, 각 연주에 대한 z점수의 가치의 의미 결정이 있다. 이 방법의 사용이 비교적 간단하지만 실제로 사용하는 데는 절차에 관한 상세한 안내가 필요하다⁸⁵⁾. '연계 차등법'은 비교적 유사한 연주를 차등 평가하고 분류하는데 실용적인 방향을 제시한다.

Radocy⁸⁶⁾는 '등급 평가법 magnitude estimation'을 음악연주 평가에 적용하였다. 주로 S. S. Stevens⁸⁷⁾의 방법을 토대로한 '등급 평가법'은 본래 강약, 취향, 피치 또는 선명함과 같은 감각을 측정하는데 사용되었다. 본질적으로 이 방법은 일정한 자극의 크기와 세기 또는 질을 숫자로 합산하는 것으로 이루어진다. 크기와 세기 그리고 질이 더 많거나, 더 나은 데에 따라서 숫자가 더 높게 지정된다. 판정자에게는 주된 자극의 등급을 나타내는 일정한 긍정적 숫자를 선택할 것이 지시된다. 각 판정자는 그 자신의 표준과 규모를 확립할 수 있으며 평가를 받는 사람은 일반 방식 또는 표준을 정하고, 표준에 비례하는 관계를 반영하기 위하여 개개 연주에 대한 각 판정자의 평가를 분류한다. 등급 평가의 계수 알파는 매우 높은 것으로 나타났다⁸⁸⁾. 등비중향geometric mean은 각 연주에 대한 자극의 평균치를 내기 위하여 산출된다. 이 방법은 아주 유사한 연주 보다는 차이나는 연주의 광역 평가를 하는 데 유용할 것이다. 또한 이 방법에는 앞에서 살펴본 '한쌍 비교법'이나 '연계 기준법' 보다 판정을 위하여 듣는 시간이 덜 소요된다.

연주를 평가하는 데 가장 많이 사용되는 공식적 절차는 전통적인 경연대회나 음악회 연주평가형식이다. 본질적으로 판정자는 각 연주마다 평가 등급을 제시하기 위하여 I-V와 같이 하나에서 다섯까지의 기준을 정해야 한다. 전형적인 형태는 음, 인토네이션, 기교, 균형, 해석과 같은 작은 표제와 연주의 각각 상황에 관하여 판정자가 등급을 정하고 논평하는 전체 결과로 구성된다. 성악 독창의 경우에는 디션과 호흡 조절과 같은 다른 기준이 포함될 수 있다. 아마도 작은 표제에 대한 등급과 논평이 I, II, III, IV, V로 이루어진 기준에 의한 총체적인 등급을 위한 근거를 제시해 줄 것이다.

85) Ghiselli, E. E., Campbell, J. P., & Zedeck, S., Measurement theory for the behavioral sciences, San Francisco, W. H. Freeman and Co., 1981, pp.404-408; Boyle & Radocy, op. cit., pp. 183-185.

86) Radocy, r. E., the influence of selected variables on the apparent size of successive pitch intervals, Psychology of Music, 1978, 6(2), pp. 21-29; Magnitude estimation of melodic dissimilarity, Psychology of music, 1982, 10(1), pp. 28-32.

87) Stevens, S. S., Psychophysics, New York, Wiley, 1975.

88) Figgs, L. D., Qualitative differences in trumpet tones as perceived by listeners and acoustical analysis, Psychology of Music, 1981, 9(20), pp. 54-62.

마지막으로 연주 평가에 대한 접근 방식으로서 '앵커 또는 모델(anchors or models)'의 사용을 살펴 보기로 하자. 이 방법은 평가자가 녹음한 연주에 대하여 '두드러진', '보통', '빈약한' 연주라는 판정 표준을 정하는 편의에 도움이 될 수 있다. 그와 같은 모델은 판정자의 기준에 '앵커'역할을 하며 많은 측정법에 적용될 수도 있다.

궁극적으로 어느 연주 측정방법을 사용할 것인 가는 그 '유용성'에 달려 있다. 주지하는 바와 같이 유용성은 테스트의 실행과 평가분류의 산출을 쉽게 해주는 상식으로서 실행과 시간계정 그리고 비용과 같은 요소가 가름의 기본이 된다. 그러나 유용성 고려로 연주의 측정선택이 신뢰도가 있으며 타당한가를 확증하기 위한 테스트 사용자의 책임이 면해지지는 않는다.

2) 광역 접근방법

음악교사들이 연주평가에서 겪는 곤란에도 불구하고 현재 사용할 수 있는 출판된 연주측정은 와트킨스 페넘 연주스케일(Watkins-Farnum Performance Scale (WFPS)⁸⁹⁾과 페넘 현악 스케일(Farnum String Scale(FSS)⁹⁰⁾ 뿐으로 알려졌다. 전자는 현악기와 스네어드럼 snare drum을 위한 것이고 후자는 오케스트라 현악기를 위한 연주평가 방법이다. WFPS는 코트넬 연주측정⁹¹⁾의 한 파생물로서 14개의 연습곡이 한 세트로서 바리톤 음자리표를 사용하고 있는데 대부분 적절한 조옮김과 편곡이 되었다(B플랫, 알토 및 베이스 클라리넷; 색소폰과 오보에; 플루트; 혼; 트롬본, 베이스 바리톤 및 바순; 튜바). 스네어드럼 측정은 12개의 연습곡만으로 구성되어 있다. FSS는 14개 연습곡으로 이루어졌으며 바이올린, 비올라, 첼로 및 베이스 현악기에 적용된다. 일반적으로 시창테스트로서 실행되는 이 방법은 개별적으로 시행되어야 하며 WFPS에서는 특수 기준과 지시가 피치, 음가, 음가의 변화, 악상 표현, 슬러 및 반복 기호에 대한 과실을 단정하는데 주어진다. FSS에서는 슬러과실을 운궁법과실로 대응하였다. 이 측정방법은 점수단위이며, 얼마나 많은 여러 가지 과실이 있었던 간에 각 측정마다 오직 하나의 과실만을 따진다.

1920년대와 1930년대에 몇 가지 테스트가 시창 기법 평가를 위하여 개발되었으나 대다수가 더 이상 사용되지 못하고 있다. 수년동안에 논문 주제나 박사 논문 연구로서 다른 몇 개 테스트⁹²⁾가 개발되기도 했으나 이들도 역시 음악교사들이 출판을 보증하고 WFPS가 기악 연주

89) Watkins, J. G., & Farnum, S. E., The Watkins-Farnum Performance Scale, Winona, Hal Leonard, 1954

90) Farnum, S. E., Farnum String Scale, Winona, Hal Leonard, 1969.

91) Watkins, J. G., Objective measurement of instrumental performance, New York, Teachers' College Bureau of Publications, Columbia University, 1942.

92) Cooper, J. J., the development of a sight-singing achievement test for use with college students, Unpublished doctoral dissertation, University of Colorado, Boulder, 1954; Scofield, W. R., The construction and validation of a method for the measurement of the sight-singing abilities of high school

자에게 실행했던 방법과 같은 전문적인 사용에는 충분치 못한 것으로 보인다. 가장 최근에 출판된 시창테스트 불원 밀의 가창 성취테스트Belwin-Mills Singing Achievement Test⁹³⁾도 마찬가지이다. 이 테스트는 그 표준과 신뢰도 테이타의 부족으로 비판을 받았다⁹⁴⁾.

연주평가에 대한 체계적인 광역 접근법은 전통적인 오디션이나 페스티발 오디션 컨테스트에 사용되는 평가 형식보다 현저하게 부족하다. 전문가들이 이러한 테스트의 사용 필요성을 느끼지 못하는데 그 원인이 있는지, 연주 평가의 광역 측정의 복잡성이 실질적 측정의 발전을 실행불가능하게 만드는 지는 분명치 않다.

3) 특수 접근 방법

광역 접근법의 복잡성 때문인지 몇명의 학자들이 연주성취의 다양한 국면 또는 단면을 평가하는 도구로서 연주평가의 개발에 착수하였다. 일정한 평가자들의 관심에 부합되는 연주성취의 국면의 하나는 아동의 음성발달이다. 오랫동안 음악교사들의 관심사였던 아동의 성악적 발달은 최근까지도 대부분 비공식적인 교실평가와 기준에 의한 음성의 분류, 또는 발달의 단계에 한정되었다. Rutkowski⁹⁵⁾는 이 분야에 대하여 포괄적인 개관을 하였으며, 많은 분류기구가 가창과 인토네이션의 정확성과 뒤섞였다고 지적하였다. 그녀는 음성발달이 인토네이션 정확성과의 분리된 평가에 대한 유력한 케이스를 만들었으며, 수년후에는 이를 발전시켜서 '가창음성 발달측정The Singing Voice Development Measure'를 고안하였다. 아동을 예비가수, 표정이 풍부한 단계의 가수, 불분명한 가수, 초보단계의 가수, 가수로 분류하는 이 측정은 비교적 충분한 신뢰도를 반영하며 내부 평가에 대한 신용계수로 96과 97을 산출하였다.

연주평가에 가장 많이 사용되는 특수 접근방법은 단계적 발전을 평가하는 '단면 계승階承 접근법facet-factorial approach'이다. Abeles⁹⁶⁾는 클라리넷 연주의 서술로서 94개 진술을 개발하였으며, 50명의 기악 교사를 통하여 중학생 클라리넷 연주자에게 각각 94개의 진술을 설문하였다. 평가 결과에 대한 요인 분석을 통하여 Abeles은 94개의 진술을 클라리넷 연주의 여섯 가지 요소를 가장 잘 반영하는 30개의 진술로 축소할 수 있었다. 그 여섯 가지 요소에는 해석, 인토네이션, 리듬의 연속성, 템포, 아티큘레이션, 음조가 포함된다. 32명의 기악 교사들

and college students, Unpublished doctoral dissertation, Michigan State University, east Lansing, 1980.

93) Bowles, R. W., The Belwin-Mills singing achievement test, New York, Belwin-Mills, 1971.

94) George, W. E., Measurement and evaluation of musical behavior, In: D. A. Hodges(Ed.), Handbook of musci Psychology, Lawrence, National Association for Music Therapy, 1980, p. 335.

95) Rutkowski, J., The measurement and evaluation of children's singing, The quarterly, 1990, 1, pp. 81-95.

96) Abeles, H. F., Development and validation of a clarinet performance adjudication scale, Journal of Research in Music Education, 1973, 21, pp. 246-255.

에 의한 클라리넷 연주의 차후 평가는 단계적 평가의 상호 신뢰도를 확립하는데 사용되었다. Abeles는 본질적으로 복합적인 행동(클라리넷 연주)의 다양한 단면이 단면계승 분석기술'에 의하여 평가될 수 있다는 것을 보여주었다. Abeles의 접근방법은 집단연주에 대한 평가 단계를 개발하는 두 가지 중요한 단면계승 접근법의 모델이 되었다. 그 하나는 코랄 음악⁹⁷⁾을 위한 것이며, 다른 하나는 기악⁹⁸⁾을 위한 것이다. 또한 출판되지 않은 박사학위 논문에서 유우포onium⁹⁹⁾과 튜바연주의 평가에도 적용되었다¹⁰⁰⁾.

Amus¹⁰¹⁾는 유타주 교육부의 음악핵심 커리큘럼 연구의 일부분으로서 음악교사가 평가될 수 있는 일련의 평가형식을 개발하였다. 그는 여기서 커리큘럼의 객관성과 관련하여 학생의 연주성취를 평가하는데 5점 등급을 사용하였다. 음악연주의 객관성은 일반적 음악으로 취주악, 관현악, 합창의 몇 개의 단계로 개발되었으며, 평가 문항 풀(pool)은 각 단계마다 개발되었다. 문항들은 교사들의 논평에 일치되게 입증되었다. 그것은 첫째, 문항이 객관성을 평가하는지 둘째, 학생 정도의 수준에 적합한지 셋째, 과제가 객관성이 적절하게 나타난 경우 학생의 85퍼센트가 달성할 수 있는 문항으로 구성되었는지 그리고 마지막으로 문항이 교사와 학생이 이해할 수 있는 수준인지 따위이다(p. 3).

그밖에도 최근들어서 미국과 독일의 몇개의 출판되지 않은 박사논문들에서 여러 가지 새로운 시도가 실행되었으나 실용단계에는 이르지 못하고 있으므로 여기서는 그에 대한 고찰을 생략하기로 한다..

4) 전자 장치 고안

개개 음들의 특성의 전자 측정은 이미 접근하기가 쉬워졌으나 음악연주의 실시간의 즉시 평가는 다른 문제이다. 음악연주의 일정한 국면의 컴퓨터 분석은 가능하지만 실제로 연주 평가의 실질적 요구를 충족시키기 위하여 개발된 전자장치는 적다. 현재까지 가망성이 있어 보이는 몇 가지 연구 성과를 개관해보기로 하자.

97) Cooksey, J. M., An application of the facet-factorial approach to scale construction in the development of a rating scale for high school choral performance, Unpublished doctoral dissertation, University of Illinois, Urbana, 1974.

98) DCamp, C. B., An application of the facet-factorial approach to scale construction in the development of a rating scale for high school band performance, Unpublished doctoral dissertation, University of Iowa City, 1980.

99) saxhorn과 비슷한 낫쇠 악기.

100) Bergee, M. J., An application of the facet-factorial approach to scale construction in a rating scale for euphonium and tuba music performance, Unpublished doctoral dissertation, University of Kansas, Lawrence, 1987.

101) Asmus, E. P., Development of forms for in-class assessment of musical performance, Paper presented at the southeastern Music Education Symposium, May 1990, athens, GA, 1990.

음악연주의 실시간 즉시 평가를 위한 가장 유력한 두 가지 전자장치는 간단히 말해서 '오디오테이프 녹음기'와 '비디오테이프 녹음기'이다. 레코딩은 음악연주 녹음으로 학자나 음악교사가 한 연주를 반복해서 들을 수 있게 하여 한번만 듣고 해낼 수 있는 것보다 더 주의를 기울인 분석을 용이하게 해준다. 비디오테이프 녹음은 연주의 일정한 상황을 시각적으로 재음미할 수 있는 부가적인 장점을 가졌다. 오늘날 오디오나 비디오의 녹음 장치는 이미 대다수 음악교사들이 접근하기 용이하며 이러한 장점을 취하는데 태만하지 않을 것이다.

두가지 비컴퓨터 전자'교육기기'(탭마스터 Tap Master, 피치마스터 Pitch Master)가 템포럴 어큐어티 프로덕트 회사 Temporal Acuity Products Inc.에 의해 개발되었으며, 이들은 리듬 경과구와 선율 경과구의 단선연주의 평가를 위한 잠재력을 지녔다. 탭마스터는 리듬 읽기 교육을 위한 도구로 고안되었으나 기계의 응답 버튼에 정확하게 처지는 리듬선의 음의 숫자에 대한 점수를 제공해준다. 평가되는 패턴은 특수화된 허용 오차 수준에 의해 사전에 계획되지만 성과 점수는 객관적이며 신뢰할 만하다.

피치마스터도 마찬가지로 평가되는 선율이 사전에 프로그램되며 연주되는 음조를 토대로한 즉, 미리 조절된 허용오차 제한에 의한 점수를 제공한다. 연주되는 매 25초마다 일점이 주어진다. 두 학생의 연주를 비교하기에 대비한 테이프의 사전 준비는 탭마스터와 피치마스터의 일정한 계획(프로그래밍)을 필요로 하나 객관적으로 리듬 형태와 선율 형태의 특수연주를 평가하는데에 유용하다.

Bengtsson과 Gabrielson¹⁰²⁾이 연주의 컴퓨터 분석을 아마도 현재 가장 유용하게 만들었을 것이다. 음악적 리듬의 분석 및 종합과 관련하여 그들은 리듬의 실시간 연주에서의 기계적 규칙성을 통하여 체계적인 변화를 분석할 수 있었다. 그들은 모든 단계에서 생기는 편차를 측정할 수 있었는데 예컨대, 한음,한박내의 음의 집단, 박의 단계, 단계의 특성 따위들이다. 그들의 SYVARD 분석에 의한 데이터를 사용하여 그들은 음악가의 연주로서 리듬 연주의 민감성을 합성하고 이를 실행 표현할 수 있었다. 현재 그들의 컴퓨터 분석은 오로지 연구 목적에만 사용되며, 그러한 체계가 평가 교육에 실질적인 방법으로 응용될 수 있는지는 생각해볼 숙제로 남는다.

아마도 연주의 전자분석에서 주된 관심사는 컴퓨터를 토대로한 기초적 음파 분석이었다고 생각되며 이는 피치 추출 pitch extraction로 알려졌다. Lorek와 Pembrook¹⁰³⁾는 1960년

102) Bengtsson, I., & Gabrielson, A., Rhythm research in Uppsala, In: Music room acoustics(pp. 19-56), stockholm, royal swedish academy of Music, 1977; Analysis and Synthesis of musical rhythm. In: J. Sundberg(Ed.), Studies of music performance(pp. 27-60), stockholm, Royal Swedish Academy of Music, 1983.

103) Lorek, M. J., & Pembrook, R. G., Present and future applications of a microcomputer-based frequency analysis system, Psychomusicology, 1989, 8, pp. 97-110.

대를 즈음하여 이에 관한 테크놀로지가 발달하였으나 최근에 와서야 마이크로컴퓨터 기술과 신경 장치가 진전되면서 마이크로컴퓨터로 그러한 분석이 가능하였다. 초기의 시도에서는 중심틀이 컴퓨터를 필요로 하였다. Graves¹⁰⁴⁾, Lorek 및 Pembroke¹⁰⁵⁾ 세 사람은 각각 시창 평가를 위한 체계를 개발하였다. 현재 그러한 프로그램은 선율상의 음의 기본 음파를 동일하게 조율된 모델과 비교하는 전형적인 유형으로 보인다. 과실은 동일한 평균율에 의한 센트 편차 cent deviation라는 용어로 기록된다. 분명히 확립되지 않으면 안되는 많은 매개변수가 있으나, 그러한 분석은 여전히 연구 목적에만 사용되어 온 것으로 보인다. 하드웨어와 소프트웨어가 더 개선되어서 교사에 의한 시창 평가에 실질적이며 신뢰도를 충족시킬 수 있을 지는 여전히 확실치 않다.

IV. 맺음말

음악능력의 객관적인 평가는 음악교육 연구에서 필요불가결한 것이다. 다수의 음악교육에 관한 연구가 서술적이던 발전적이던 또는 상호관련적이던 경험적이던간에, 음악능력의 일정한 국면의 측정을 독립된 변수나 의존적 변수로서 내포하고 한다. 음악능력의 개념 정의가 명백하며 그 측정이 신뢰성과 타당성이 있다는 전제아래서는 음악능력은 연구자들에게는 유용한 구조이다. 하지만 잘못된 정의와 부정확한 측정 그리고 음악능력에 관한 데이터의 빈약한 해석은 잘못된 결과를 초래할 수 있으며, 비정상적인 연구 성과와 결론을 이끌어 낼 수도 있다. 음악능력을 진단하고 논증하며 측정하는 방법의 다양성은 음악능력에 관련된 데이터의 적용과 해석의 문제로 귀착된다. 그러한 다양성은 음악능력에 관한 데이터 사용자에게 음악능력 평가에 대한 근거로서 제공되는 테스트 과제의 특성을 명백하게 상술하는 데에 효과적이다. 특히 테스트 과제와 음악능력의 특수한 구성 그리고 음악능력이 평가되는 차원 사이에 높은 추정치가 있는 테스트 과제에 효율적이다. 일반적으로 추정표준은 연주차원보다 음악능력의 지각과 개념적 차원에서 더 높다. 하지만 어떤 경우에는 표면상으로 곧바로 범위가 한정적으로 보이는 음악연주 과제로부터 파생된 고차원의 추정이 있는 것으로 보인다. 테스트 과제와 음악능력의 구성사이의 추정차가 클 수록 무엇이 평가되는지에 관한 일치를 볼 가망이 적어진다.

104) Graves, D. L., the development of an objectivesight singing achievement test employing electronic measurement apparatus, Unpublished doctoral dissertation, University of Georgia, Athens, 1980.

105) *ibid.*

넓은 의미에서 음악능력의 평가 데이터는 세 가지 기능을 한다. 첫째, 장래를 향한 적성 또는 예견 기능 둘째, 한 학생이 테스트 시간에 할 수 있는 것에 초점을 두고 장래의 잠재력에 관한 추정을 하지 않고 표시된 능력이 그 어떤 특수한 교육의 성과인지를 진단하는 진단적 기능 셋째, 표시된 능력이 일반적으로 형식적인 교육의 한 유형으로서 경험의 결과를 의미하는 성취 기능이다.

음악교사가 학교에서의 교육적 절차를 촉진하기 위하여 음악능력 데이터를 적용하고 사용하는 범위는 다음 상황에 따라서 크게 달라질 수 있다. 먼저 교육과정 평가의 중요성에 대한 상이한 철학, 음악능력을 평가하는 다양한 지식과 기술 및 음악능력 평가의 데이터의 기능에 관한 이해 부족은 교육과정 확정에 영향을 미칠 수 있다는 이유를 들 수 있을 것이다. 음악능력 평가의 데이터가 교육과정의 필수 부분이 되고 학교 음악교육의 신선한 충격이 되기 위하여 음악교사는 교육과정에 대한 이의 잠재적 기여에 관한 이해를 도모하며, 그러한 데이터를 적용할 수 있는 지식과 기술을 개발하고 그리고 음악능력 평가에서 객관성을 증가시키기 위한 긍정적인 경향을 촉진시켜야 할 것이다.

개인차에 따른 음악능력의 개념과 그 평가방법에 대한 연구는 음악심리학의 한분야로서 미국을 중심으로 발달하여 금세기 후반에 이르러서는 음악심리학에서만 아니라 음악교육학에서도 개인차심리학을 근거로 매우 중요한 위치를 차지하고 있다. 그러나 이 분야에 대한 국내의 연구는 먼저 학문적인 연구차원보다는 실제교육의 수용에서 시각적 평가방법에서 청각적 평가방안을 강구하는 측면에서 시도되었으며¹⁰⁶⁾, 최근에 들어서야 음악심리학에 관한 번역서¹⁰⁷⁾와 저서¹⁰⁸⁾에서 그 일반적인 개관이 소개되고 있다. 따라서 본 논문에서는 음악능력의 개념정의의 음악심리학과 음악교육학적인 측면에서 접근하였으며, 현재 미국을 중심으로하여 사용되고 있는 음악테스트를 소개하고 그 문제점을 살펴봄으로써 우리 현실에 적합한 음악능력 평가방안을 개발하는 학문적 기초를 마련하고자 하였다. Seashore테스트를 비롯한 미국의 음악테스트들이 주로 세계적으로 사용되었으나 현재는 이에 준한 나라별로 문화적인 차이를 감안한 테스트들이 개발되고 있으므로 앞으로 우리의 실정에 맞는 각 분야의 테스트를 만들어내는 작업이 뒤따라야 할 것이다.

106) 각주 42)와 43)을 참조.

107) 서우석(역), Sloboda, 음악의 심리학, 심철당, 1992.

108) 이석원, 음악심리학, 심철당, 1992.