

# 제7차 교육과정에 근거한 준거지향적 수행평가 문항의 개발과 평가 -고등학교 과학 「생식」과 「생물 농축」단원을 중심으로-

정영란 · 박진주  
(이화여자대학교)

## Development and Evaluation of Criterion-Referenced Performance Assessment Items Based on the 7th National Science Curriculum -Subject Unit of Reproduction and Biological Accumulation-

Chung, Young-Lan · Park, Jin-Joo  
(Ewha Womans University)

### ABSTRACT

In recent years, there has been an increased emphasis on performance assessment to evaluate students' abilities. Our nation has introduced a change in testing and assessment. Additional work on the efficacy, reliability, and comparability in order to develop the performance assessment item has been needed in the enforcement of the 7th National Science Curriculum. Also, criteria for professional and technical standards has been needed to be developed.

The purpose of this study was to draw out various key concepts and to develop achievement standards, assessment standards and performance assessment items based on the 7th National Science Curriculum on the subject matter of reproduction(chapter 13) and biological accumulation(chapter 17). And also, this study examined the validity of completed performance assessment items based on classical test theory and polytomous item response theory.

Twelve key concepts in chapter 13(reproduction) and four from chapter 17(biological accumulation) were abstracted. Twenty-six achievement standards in chapter 13(reproduction), and nine in chapter 17(biological accumulation) were developed. The achievement standards were determined in terms of knowledge(K), process skill(P) and attitude(A). Twenty-five assessment standards in chapter 13(reproduction) and nine in chapter 17(biological accumulation) were developed. Based on the developed achievement standards and assessment standards, twenty-two performance assessment items(seventeen open-ended questions, three essays, and two portfolios) with concrete grading criteria were developed.

Eight open-ended items were applied to 240 10th graders to evaluate reliabilities of the test which consisted of four items per each chapter. The results would be suggested that the applied items were valid for performance assessment because item difficulties and item discriminations were proper. There was not much differences in item discrimination between interpretation from classical test theory and that from polytomous item response

theory. However, there were some differences in item difficulties between the interpretations of two theories because the characteristics of examinees were reflected in classical test theory.

**Key words:** key concepts, achievement standards, assessment standards, criterion-referenced performance assessment items, classical test theory, polytomous item response theory, science education

## I. 서론

Tyler(1963)는 교육 목표의 달성 여부를 판단하는 행위를 교육 평가라 정의하였고, Cronbach(1984)는 평가의 목적이 새로운 교육과정을 검사하고 그에 근거해서 교육과정을 개선하는 것이라 하였다. 현대 교육에서 교육 평가는 교육과 관련된 모든 것에 대한 가치 판단으로 의미가 확대되고 있지만 결과적으로 학습 과정에 도움을 주어 학습을 극대화하는 것을 그 목적으로 한다(김원중 등, 1995).

현재 학교 현장에서 가장 보편적으로 사용되고 있는 교육 평가 도구들은 Tyler(1963)의 정의에 입각하여 교육과정과 교육 프로그램에 기초한 교육 목적의 달성 여부를 평가하는 것으로, 대부분이 선다형 문항들로 이루어진 규준지향(norm-referenced) 평가이다. 그러나 제7차 교육과정이 도입되면서 수행 평가와 같은 대안적인 평가가 강조 되어졌고, 기존의 규준지향 평가만으로는 교육 평가의 목적을 제대로 수행하기가 힘들어지고 있는 실정이다. 선다형 평가는 단순한 지식이나 정보의 습득여부에 대한 평가를 하기에는 좋은 방법일지 모르나, 학습자로 하여금 정답을 찾도록 유도함으로써 창의적인 사고와 적극적인 반응을 억제하고 수동적인 학습자로 만든다(Whitehead, 1967). 또한 선다형 평가는 학생들의 고등 사고 기능을 평가하기보다 지식 위주의 단편적인 사고 능력을 평가함으로써 학습 과정에 대한 평가가 어렵고(Glaser, et al., 1987), 학생들의 인지, 정의, 심동 영역에 대한 종합적인 평가에 소홀해 지며 학생 개개인의 전인적인 발달을 평가하기에는 미흡하다(백순근 등, 1998).

따라서 학생들이 사물을 이해하는 과정과 사고하는 과정을 추론할 수 있고, 인지적 구조의 변화나 이해 수준에 대한 정확한 진단을 내릴 수 있는 새로운 평가 방식이 필요하게 되었다. 기존의 평가 방식을 벗어난 다양한 대안적인 평가 방법으로 수행 평가 방식이 주목을 받기 시작하였고, 이를 통해 학생들의 인지 과정과 이해수준을 파

악하려는 검사 이론들도 등장하게 되었다(박정, 2001).

Glaser(1963)가 규준지향 평가(norm-referenced evaluation)와 준거지향 평가(criterion-referenced evaluation)에 대한 개념을 구분한 이후로 준거지향 평가는 새로운 평가 모형으로 과학교육 분야에도 도입되었다. 규준지향 평가가 개인과 개인을 상호 비교하는 것이었다면, 준거지향 평가는 개인의 학습 결과와 준거를 비교하는 데 더 관심을 두고 평가의 교수 기능을 강조한다. 준거지향 평가는 절대적인 준거인 학습 목표의 달성 여부나 그 도달에 비추어 학생들의 성취 정도를 확인하는 평가로, 그 점수는 학생이 무엇을 알고 무엇을 모르는가에 대한 확실한 정보를 제공한다(Glaser, 1963). 제7차 교육과정의 시행과 함께 우리나라의 교육 평가는 준거지향 평가를 지향하고 있으므로 학교 교육 현장에서 준거지향 평가가 시행되기 위해서는 우선 그것에 맞는 타당한 기준을 마련할 필요가 있다.

따라서 본 연구에서는 제7차 교육과정에 근거하여 고등학교 과학의 '생식'과 '생물 농축' 단원에 포함된 중요 개념들을 대상으로 필수 학습 요소를 추출하고 성취 기준과 평가 기준을 마련한 후, 이를 바탕으로 다양한 형태의 준거 지향적인 수행 평가 문항들을 개발하였다. 또한, 개발한 평가 문항을 학생들에게 적용하여 고전 검사 이론과 다분 문항 반응 이론으로 분석하였다.

## II. 연구 방법 및 절차

### 1. 필수 학습 요소의 추출과 성취기준, 평가 기준의 개발

고등학교 교육과정 해설(교육부, 1999)과 과학과 교육과정(교육부, 1998)을 토대로 필수 학습 요소를 선정하였다. 교육과정만으로는 그 내용을 구체화하기가 곤란하므로 고등학교 7종 교과서(김찬중 등, 2001; 성민용 등, 2001; 오규환 등, 2001; 이규석 등, 2001; 이문원 등,

2001; 이연우 등, 2001; 정완호 등, 2001)의 해당 단원에서 다루고 있는 개념을 모두 추출하여 공통적으로 다루고 있는 개념들을 필수 학습 요소로 최종적으로 선정하였다.

성취 기준은 교육과정의 '지식'과 '탐구'에 명시된 내용을 학생들이 달성해야할 능력의 범위와 수준 또는 특성이 드러나도록 진술한 것으로(이양락 등, 1998; 강순희, 2001) 상세하게 명시적으로 진술하였다. 제7차 교육과정에서는 과학의 기본 개념 이해, 탐구 능력, 그리고 과학에 대한 흥미와 호기심 및 과학적 태도를 강조하므로 성취 기준을 지식(K), 탐구(P), 태도(A)의 3개의 영역으로 분류하였다.

평가기준은 성취 기준을 바탕으로 학생들이 도달해야할 능력수준으로 상, 중, 하로 구분하여 각 수준에서 기대되는 성취정도를 구체적으로 진술하였다. '상' 수준은 10학년 학생이 충실한 교수·학습 활동으로 도달해야할 것으로 기대되는 이상적인 능력 수준이고, '중' 수준은 '상' 수준의 성취가 양적 혹은 질적으로 미흡한 경우이며, '하' 수준은 보통의 10학년 학생이 성취하도록 기대되는 필수적인 내용을 충분히 성취하지 못한 수준이라고 할 수 있다. 심화·보충과정은 교육과정에서 평가하지 않는다고 제시되어 있으므로 그 평가 기준을 개발하지 않았다. 필수 학습 요소, 성취기준, 평가기준에 대한 타당도는 현장 교사 2인과 과학교육전문가 2인의 검증을 받았다.

## 2. 수행 평가 문항의 개발

개발된 성취 기준과 평가 기준에 맞추어 학생들의 성취 수준을 측정할 수 있도록 각 평가 영역별로 수행 평가 문항을 개발하였고 채점 기준을 작성하였다. 평가 문항은 기존의 선다형 규준지향 평가를 탈피하고 실생활과 밀접한 주제를 사용하여 서술형, 논술형, 포트폴리오 등 다양한 형식으로 개발하였다. 수행 평가 문항은 제7차 교육과정에 근거하여 개발된 성취 기준 및 평가 기준과 일관성을 가지도록 하였고 평가 기준에서 제시한 수준별 진술문을 참고하여 학생의 성취 수준을 판정하였다. 성취 기준 및 평가 기준에서는 지식, 탐구 및 태도로 각 평가 영역을 구분하여 진술했지만, 수행 평가 문항에서는 각 영역간의 통합을 도모하였다. 평가 문항은 성취 기준과 평가 기준을 단독 또는 복합적으로 사용하여 다양하게 제시하였다. 평가문항에 대해 2차에 걸쳐 파일럿 테스트를 실시하여 학생들이 문항의 의도를 정확히 파악하지 못하는 경우는

그 내용을 수정 보완하였다.

## 3. 수행 평가 문항의 분석

수행평가의 검사 도구를 실제 운영하기 위해서는 여러 가지 측정학적 문제들을 먼저 해결하여야 하고(남명호, 1995; 남현우, 1998) 교육현장에서 학생들의 성취도를 제대로 평가할 수 있는 타당한 검사 도구라는 것이 입증되어야만 한다. 개발한 수행평가 문항의 양호도를 검증하기 위하여 서술형 문항 중 각 단원 별로 4문항씩을 선정해 검사도구를 구성하여 경기도 소재 고등학교 1학년 7개 학급에 적용하였다. 피험자가 피험자 자신에 대한 정보를 포함하여 아무 것도 표기하지 않은 경우는 무응답으로 처리하였으므로, '생식' 단원 검사의 총 응답자는 231명, '생물 농축' 단원 검사의 총 응답자는 196명이었다. 검사 시간은 20분으로 학생들이 문제를 해결하는 데는 충분하였다. '생식' 단원의 검사는 학생들이 그 단원을 배운 후 2002년 10월 말에서 11월 초에 형성평가로 이루어졌다. '생물 농축' 단원은 대상 학생들이 중3 때 제6차 교육과정에서 이미 배운 내용이므로 제 7차 교육과정에서 학습하기 전에 진단평가로 이루어졌다. 문항에 대한 내용 타당도는 평가 도구의 개발 과정에서 과학 교육 전문가 2인과 현직 교사 2인의 검증을 받았다.

수행 평가 도구의 신뢰도, 난이도와 변별도 등을 분석하기 위해 학생들의 응답을 바탕으로 검사 이론 중에서 가장 보편적으로 사용되고 있는 고전 검사 이론과 수행형 문항의 분석을 위해 최근 도입되고 있는 다분 문항 반응 이론을 사용하였다.

### 1) 신뢰도

신뢰도는 검사가 측정해야할 내용을 얼마나 일관되게 측정하고 있는가를 나타내는 것으로 .7 이상의 내적 일관성 신뢰도 지수를 보이는 검사를 좋은 검사 도구로 본다(성태제, 1996b). 본 연구에서는 검사 도구의 신뢰도를 분석하기 위하여 SPSS 10.0 프로그램을 이용하여 내적 일관성 신뢰도(Cronbach  $\alpha$ )를 산출하였다.

### 2) 난이도

고전 검사 이론에서 문항 난이도는 전체 피험자 중에서 정답자의 비율로 계산을 하므로, 난이도는 얼마나 어려운가에 대한 지수가 아니라, 얼마나 쉬운가에 대한 지수로

나타난다. 난이도 지수가 .25 이하이면 어려운 문항, .75 이상이면 쉬운 문항으로 해석된다(성태제, 1996). 본 연구의 수행 평가 문항에서는 학생들의 점수가 0점, 1점 혹은 2점으로 구분되므로 각 점수간의 경계 난이도를 따로 구하였다. 따라서 총 피험자 수 중 2점을 얻은 피험자 수의 비율을 2점 경계 난이도, 1점 혹은 2점을 얻은 피험자의 비율을 1점 경계 난이도로 계산하였다.

다분 문항 반응 이론에 의한 문항 모수와 피험자 능력 모수치를 얻기 위해서는 박정(2001)의 선행연구를 참조하여 PARSCALE 컴퓨터 프로그램(Muraki & Bock, 1998)을 사용하였고, 문항 난이도는 Baker(1985)의 언어적 표현에 따른 난이도 범위를 기준으로 해석하였다. 다분 문항 반응 이론에서는 문항 모수들이 각 점수의 경계에서 하나씩 나오게 된다. Baker(1985)에 따르면 문항 난이도 지수가 - .5에서 + .5의 범위에 있을 때, 중간 정도의 능력 수준을 가진 학생들에게 가장 적합한 문항이며, 범위보다 낮은 난이도 지수를 나타내면 쉬운 문항, 범위보다 높은 난이도 지수를 나타내면 어려운 문항으로 판단한다.

### 3) 변별도

고전 검사 이론에 의한 변별도는 총점과 문항 점수간의 상관 계수를 산출하여 추정한다. 고전 검사 이론에 대한 Ebel과 Frisbie(1986)의 언어적 해석 기준에 따르면, 문항 변별도 지수가 .30 이상이면 변별력이 있는 문항, .40 이

상이면 변별력이 높은 문항으로 해석된다. 본 연구에서는 고전 검사 이론에 의한 변별도 산출을 위해 SPSS 10.0 프로그램을 사용하였다.

다분 문항 반응 이론을 적용한 문항 분석 결과에서 문항 변별도는 성태제(1991)의 언어적 표현에 따른 변별도 범위를 기준으로 해석하였다. 본 연구에서 사용한 2-모수 로지스틱 모형의 경우, 변별도 지수가 .35 이상이면 변별력이 있는 문항이며, .65 이상의 경우 적절한 변별력을 가진 문항이라고 해석한다(성태제, 1991).

## III. 연구 결과 및 논의

### 1. '생식', '생물 농축' 중단원의 필수 학습 요소

과학과 교육과정에서 명시하고 있는 생식 영역에 대한 내용은 사람의 생식 기관의 구조와 생식 세포의 형성, 여성의 생식 주기와 발생 과정의 이해, 모체가 태아에 미치는 영향에 대한 조사이다. 그리고 인공 수정에 대하여 조사와 토의하기는 심화 과정으로 되어 있다. 환경 단원에서는 생물 농축, 산성비, 온실 효과, 소음 등과 같은 환경 문제가 발생하는 원인을 알고 그 해결 방법을 찾는 것을 주요 내용으로 하고 있다(교육부, 1998).

추출된 필수 학습 요소는 Table 1과 같다. 생식 단원에서 추출된 필수 학습 요소는 사람의 생식 기관의 구조와

Table 1. Key concepts of chapter13(reproduction) and chapter 17(biological accumulation)

Chapter	Key concepts
Chapter13. (reproduction)	<ul style="list-style-type: none"> <li>· The structure and function of human reproductory system</li> <li>· Spermatogenesis and oogenesis</li> <li>· Structure of a sperm and an egg</li> <li>· Regulation of female reproductory cycle by hormones</li> <li>· Change of an ovary and an uterus during a menstrual cycle</li> <li>· The process of fertilization and implantation</li> <li>· Development and growth of fetus</li> <li>· Material exchange between a mother and a fetus</li> <li>· The effect of a mother's condition to a fetus</li> <li>· The process of parturition</li> <li>· The mechanisms and methods of contraception</li> <li>* Investigation and discussion on artificial insemination</li> </ul>
Chapter 17. (biological accumulation)	<ul style="list-style-type: none"> <li>· The definition of biological accumulation</li> <li>· The process of biological accumulation</li> <li>· Characteristics of the materials generated by biological accumulation</li> <li>· Damages by the biological accumulation</li> </ul>

\* Key concepts of intensive course

기능을 비롯하여 모두 12개이며 인공 수정에 대하여 조사와 토의하기는 심화 과정으로 되어있다. 생물 농축 단원에서는 생물 농축의 정의와 과정, 생물 농축 물질의 특성, 생물 농축의 피해로 모두 4개의 필수 학습 요소가 추출되었다.

## 2. '생식', '생물 농축' 중단원의 성취 기준

성취 기준은 필수 학습 요소를 바탕으로, 지식(Knowledge; K), 탐구(Process skill; P), 태도(Attitude; A)의 세 가지 영역에 대하여 학습 목표로 사용할 수 있도록 개발하였다. 각 필수 학습 요소가 세 영역에 모두 적용된 것은 아니지만, 필수 학습 요소가 하나 이상의 영역에 적용된 경우도 있었다. 심화 과정은 '생식' 단원에서 조사와 토의 활동으로 제시되었으므로 탐구(P) 영역의 성취 기준을 적용하였다. 지식(K) 영역에서는 필수 학습 요소에 해당하는 개념과 그 특성을 비교하거나 설명할 수 있도록 성취 기준을 설정하였으며, 탐구(P) 영역에서는 SAPA(Science-A Process Approach)에서 제시한 탐구 과정 기능의 기초적 탐구와 통합적 탐구 기능을 모두 포함할 수 있도록 하였다. 태도(A) 영역은 정의적 요소로서 비교적 장기간의 학습 결과로 내면화되는 속성을 지니므로 단원 수준에서 포

괄적으로 개발하였으며, 과학적인 태도를 기르고, 과학적 지식과 탐구 능력을 실생활에 활용할 수 있도록 하는 마음가짐을 기르는 것을 목표로 하였다.

'생식' 단원에서는 심화 과정을 포함하여 지식영역 12개, 탐구영역 12개, 태도영역 2개로 총 26개의 성취 기준이 개발되었고 '생물 농축' 단원에서는 지식영역 4개, 탐구영역 4개, 태도영역 1개로 총 9개의 성취기준이 개발되었다. '생물농축' 단원의 성취기준을 Table 2에 제시하였다.

## 3. '생식', '생물 농축' 중단원의 평가 기준

개발된 성취 기준에 대하여 각각의 성취 정도를 어떻게 평가할 것인가에 대한 평가 준거로서 평가 기준을 개발하였다. 평가 기준은 성취 기준에 일대일 대응으로 개발하였으므로 성취 기준과 마찬가지로 지식, 탐구, 태도의 3개 하위 영역으로 나누었으며 상/중/하의 세 수준으로 개발하였다. 평가에서 상/중/하로 구분하기 위해서 양적인 진술과 질적인 진술의 위계적 진술, 병렬적 진술을 사용하거나 양적, 질적 진술을 혼용하여 사용하였다.

'생식' 단원에서는 지식영역의 평가기준(EK) 12개, 탐구영역의 평가 기준(EP) 11개, 태도영역의 평가 기준(EA)

Table 2. Achievement standards of chapter 17(biological accumulation)

Knowledge(K) domain	
K17-1	Students can explain the meaning of biological accumulation related to food chain.
K17-2	Students can list the names of materials generated by biological accumulation.
K17-3	Students can explain the common characteristics of biologically accumulated materials.
K17-4	Students can explain the damages by the biological accumulation.
Process skill(P) domain	
P17-1	Students can interpret the materials displaying the process of biological accumulation. (Interpretation)
P17-2	Students can deduce the pathway of materials biologically accumulated to ecosystem by investigating the damage cases. (Deduction)
P17-3	Students can make a report about biological accumulation by diagnosis and spot investigation. (Data collection and arrangement)
P17-4	Students can consider a counterplan to diminish the damage of biological accumulation. (Prediction)
Attitude(A) domain	
A17-1	Students are interested in the damage by biological accumulation and practice the plan for preventing it.

2개로 총 25개가 개발되었으며, '생물 농축' 단원에서 서는 지식영역 4개, 탐구영역 4개, 태도 영역 1개로 총 9개가 개발되었다. '생식' 단원의 탐구영역의 평가 기준의 일부를 Table 3에 예시하였다.

#### 4. 수행 평가 문항의 개발

개발된 성취 기준과 평가 기준에 근거하여 수행 평가 문항을 개발하였다. 본 연구에서는 다양한 수행 평가 방법들 중에서 서술형, 논술형, 포트폴리오의 세 가지 형태의 수행 평가 문항들을 개발하였는데 개발한 평가 문항은 생식 단원에서 서술형 9개, 논술형 1개, 포트폴리오 1개이

었고, 생물농축 단원에서 서술형 8개, 논술형 2개, 포트폴리오 1개로, 총 22문항이었다.

서술형 평가 문항은 실제 학교 현장에서 가장 많이 활용되고 있는 수행 평가 방법이며, 일반적으로 수행 평가 문항의 개발과 시행 및 관리에 드는 시간과 비용의 부담이 크다는 점을 감안하면, 비교적 적용이 용이한 평가 형식이라고 할 수 있다. 개발된 평가문항 중 문항분석에 사용했던 서술형 문항 중 하나를 Table 4에 예시하였다.

예시문항은 복제양 '돌리'를 소재로 하여 수정란의 착상 과정에서 필요한 호르몬의 이름과 역할을 묻는 문항이다. 이 문제에 대한 필수 학습 요소는 호르몬에 의한 여성의 생식 주기 조절과 수정과 착상의 과정이었다. 지식(K)

Table 3. Five among eleven assessment standards of process skill domain in chapter 13(reproduction)

	Assessment standards		
	High	Middle	Low
EP 13-1	Students can compare the structures and the functions of male reproductive systems to female one by analyzing materials.	Students can compare the structures and the functions of male reproductive systems to female one by analyzing materials partly. However, it is qualitatively insufficient.	Students can not compare the structures and the functions of male reproductive systems to female one by analyzing materials.
EP 13-2	Students know that germ cells are from meiosis. They can compare spermatogenesis to oogenesis.	Students know that germ cells are from meiosis. However, they can not compare spermatogenesis to oogenesis.	Students do not know that germ cells are from meiosis. And they can not compare spermatogenesis to oogenesis.
EP 13-3	Students can compare the characteristics of sperm and eggs such as structure, chromosome number, size, and mobility.	Students can compare the characteristics of sperm and eggs such as structure, chromosome number, size, and mobility partly.	Students can not compare the characteristics of sperm and eggs such as structure, chromosome number, size, and mobility.
EP 13-4	From the female hormonal graph, students can deduce the change of ovary and endometrium, and the regulation of female reproductive cycle by hormones.	From the female hormonal graph, students can deduce the change of ovary and endometrium, and the regulation of female reproductive cycle by hormones partly. However, qualitatively insufficient.	From the female hormonal graph, students can not deduce the change of ovary and endometrium, and the regulation of female reproductive cycle by hormones.
EP 13-5	Students can expect the site of fertilization, the pathway of fertilized egg, and the site of implantation.	Students can expect the site of fertilization, the pathway of fertilized egg, and the site of implantation partly.	Students can not expect the site of fertilization, the pathway of fertilized egg, and the site of implantation

EP: assessment standards of domain of process skill  
13-1: the first standard of chapter13(reproduction)

영역의 성취 기준은 여성 호르몬에 의한 여성의 생식 주기 조절을 설명하는 것과 수정란이 수정 후 발생하여 착상되기까지의 과정을 기술하는 것이었다. 그리고 탐구(P) 영역의 성취 기준은 자료를 통하여 수정란이 착상하기까지의 이동 경로와 발생 과정을 유추하는 것이었다. 예시 문항에 대한 평가 기준은 상/중/하의 진술로 Table 5에 제시하였다. 이 문항은 여성 호르몬의 종류와 생식 주기

조절 과정에서의 역할, 그리고 수정란의 이동 경로 및 발생 과정에 대한 지식과 추론 능력을 통합적으로 묻는 문항이므로, 평가 기준 EK13-4, EK13-7, EP13-6, 세 개가 모두 적용되었다. 채점기준은 Table 6과 같이 호르몬의 이름과 역할을 제대로 설명한 경우에 2점, 호르몬의 이름은 알고 있으나 역할의 설명이 미흡한 경우에 1점을 부여하고 무응답이나, 오답은 0점으로 처리하였다.

Table 4. An example of performance assessment item(Item 3)

\* Following processes indicate how the cloned sheep 'Dolly' was born.

1. Take unfertilized egg from a sheep(A), and remove the nucleus from it
2. From a 6-year-old sheep(B), take a nucleus from the mammary gland cell and transplant it into the egg from prepared by step 1.
3. Implant the cell from prepared by step 2 to the uterus of the other ewe(C).
4. The lamb which (C) delivered is the clone of (B).

〈Question〉

What are the names and the roles of the hormones that the surrogate mother(C) needs for a normal growth of the fertilized egg after implantation?

Table 5. Assessment standards of performance assessment item 3

	Assessment standards		
	High	Average	Low
EK13-4	Students can explain the names and functions of female sex hormones controlling menstrual cycle.	Students know the names and functions of female sex hormones. However, they cannot explain them related to menstrual cycle.	Students do not know the names and functions of female sex hormones. And they cannot explain them related to menstrual cycle.
EK13-7	Students can explain the process of implantation of fertilized egg related to its moving pathway.	Students can explain the process of implantation of fertilized egg related to its moving pathway partly. However, it is qualitatively insufficient	Students can not explain the process of implantation of fertilized egg related to its moving pathway.
EP13-6	Students can deduce the developmental process of fertilized egg considering its moving pathway and implantation time.	Students can deduce the developmental process of fertilized egg considering its moving pathway and implantation time partly. However, it is qualitatively insufficient	Students can not deduce the developmental process of fertilized egg considering its moving pathway and implantation time.

Table 6. Grading criteria of performance assessment item 3

Score	Answers
2	Surrogate mother(C) needs estrogen and progesterone to prevent more ovulation and to keep uterine lining thick.
1	Students know that estrogen and progesterone are needed. However, they can not explain the roles of the hormones enough.
0	No responses or wrong answers.

Table 7. Reliabilities of the tests

		'Reproduction'	'Biological accumulation'
$\alpha$ value when each item was excluded	Item 1	.511	.472
	Item 2	.480	.408
	Item 3	.456	.423
	Item 4	.540	.390
Cronbach $\alpha$		.570	.497

## 5. 서술형 평가 문항의 분석

### 1) 신뢰도

서술형 수행 평가 문항에 대한 학생들의 응답을 채점한 결과를 근거로 검사 도구의 신뢰도를 분석하였다. 각 검사도구의 신뢰도 지수는 Table. 7과 같다. '생식' 단위 검사의 신뢰도 계수는 .570, '생물 농축' 단위의 신뢰도 계수는 .497로 다소 낮은 편이었다. 이러한 결과는 검사 도구의 문항 수가 각기 4 문항으로 한정되었고, 각 문항마다 필수 학습 요소 및 성취 기준과 평가 기준이 각기 달라 서로 독립된 문항이라고 할 수 있으므로 검사 도구의 내적 일관성 신뢰도가 낮게 나온 것 같다. 하지만, 문항 수가 좀 더 많았거나, 같은 필수 학습 요소와 성취 기준에 대한 문항으로 구성된 검사 도구였다면 신뢰도 계수는 더 높아졌을 것이라 생각된다.

일반적인 준거지향 평가의 경우 문항의 수가 많지 않기 때문에 고전검사 이론에 근거한 신뢰도 지수들을 적용하면 정확한 신뢰도 지수 값을 추정할 수 없는 문제점이 있다. 그러므로 고전 검사 이론에 따른 신뢰도 지수대신 검사결과에 근거한 절대 기준의 일관성을 강조한 새로운 개념의 신뢰도 지수의 사용이 필요하다(성태제, 1996).

### 2) 난이도와 변별도

#### (1) 고전 검사 이론에 의한 문항 분석

고전 검사 이론으로 8개 문항을 분석하였을 때 평균 난이도(1점 경계)는 .520이었고 평균 변별도는 .322이었으므로 개발된 문항들이 수행평가 문항으로 적절하다고 생각된다. 단위 별로 자세히 분석해 보면 '생식' 단위의 검사에 대한 문항 평균 점수와 변별도 및 난이도는 Table 8과 같다. 문항 평균 점수가 가장 높은 문항은 문항 1(1.186)로, 남자와 여자의 생식 세포 형성의 공통점과 차이점을 설명하는 문제였으며, 평균 점수가 가장 낮았던 문항은 문항 3(.455)으로 임신을 지속시키는 역할을 하는 호르몬에 대한 문제이었다.

'생식' 단위에서 4개 문항의 평균 난이도는 1점 경계에서 .543, 2점 경계에서 .232이었다. 이것은 1점 이상을 받은 학생들이 절반 정도 되었다는 뜻이므로, 피험자 집단의 수준에 1점 경계의 난이도는 적정 수준인 것으로 보이지만, 2점 경계의 난이도는 약간 높다고 해석할 수 있다. 1점 경계에서 문항 1(.788)은 쉬웠고, 문항 2(.472)와 문항 4(.667)는 적절한 난이도를 보였고, 문항 3(.247)은 약간 어려웠다. 그러나 2점 경계에서는 문항 1(.398)만 적절했고 문항 2(.113), 문항 3(.208), 문항 4(.208)는 어려운 문항으로 분석되었다. '생식' 단위 검사 문항의 평균 변별도는 .354로 변별력이 있었고 문항 1, 2, 3은 변별력이 있는 문항이었고 문항 4는 거의 판별 경계점수인 .298이었다.

'생물 농축' 단위 검사의 문항 분석 결과는 Table 9와



Table 8. Difficulties and discriminations of the items included in 'reproduction' test in terms of classical test theory

Item	Average score	Discriminations	Difficulties	
			1 point boundary	2 point boundary
Item 1	1.186	.337	.788	.398
Item 2	.584	.379	.472	.113
Item 3	.455	.401	.247	.208
Item 4	.875	.298	.667	.208
Average	.775	.354	.543	.232

Table 9. Difficulties and discriminations of the items included in 'biological accumulation' test in terms of classical test theory

Item	Average score	Discriminations	Difficulties	
			1 point boundary	2 point boundary
Item 1	1.454	.234	.862	.592
Item 2	.546	.309	.306	.240
Item 3	.709	.293	.459	.250
Item 4	.592	.326	.357	.235
Average	.825	.291	.496	.329

같다. 검사의 4개 문항에 대한 평균 난이도는 1점 경계에서 .496, 2점 경계에서 .329이었다. 따라서 전반적으로 볼 때, '생물 농축' 단원의 검사는 1점 경계와 2점 경계에서 피험자 집단에 적절한 난이도 수준인 것으로 나타났다. 1점 경계에서는 문항1(.862)은 쉬웠고 문항2(.306), 문항3(.459), 문항4(.357)은 적절한 난이도를 가졌다. 2점 경계에서는 문항1(.592), 문항3(.250)은 적절한 문항이었고, 문항2(.240)와 문항4(.235)은 어려운 문항으로 분석되었다. '생물 농축' 단원의 검사는 학생들에게 제7차 교육과정에 따른 학습이 일어나기 이전에 진단 평가의 형식으로 실시하였는데 학생들에게 먹이 연쇄와 관련한 선행 학습이 비교적 잘 이루어졌음을 알 수 있었다. '생물 농축' 단원 검사의 평균 문항 변별도는 .291로 약간 낮다고 해석된다. 문항1이 지나치게 쉬웠으므로, 문항의 변별력이 .234로 낮았던 것이 전체 검사의 변별도를 떨어트린 주요 원인이 된 것 같다. 문항2(.309), 문항3(.293), 문항4(.326)는 변별력이 비교적 양호한 문항이었다.

(2) 다분 문항 반응 이론에 의한 문항 분석

다분 문항 반응 이론으로 8개 문항을 분석하였을 때 평

균 난이도(1점 경계)는 -.368이었고 평균 변별도는 .742이었으므로 개발된 문항들이 수행평가 문항으로 적절하다고 생각된다. 단원 별로 자세히 분석해 보면 '생식' 단원 검사의 문항 모수 값은 Table 10과 같다.

'생식' 단원에서 4개 문항의 평균 난이도는 1점 경계에서 -.508로 약간 낮은 수준의 학습자에게 적합한 난이도였으며 2점 경계에서는 1.116으로 높은 수준의 능력을 가진 학습자에게 적합하다고 분석되었다. 1점 경계에서 문항1(-1.414)과 문항4(-.943)는 비교적 쉬운 문항이라고 해석할 수 있었으며, 문항2(.153)와 문항3(.170)은 적절한 난이도를 갖는다 할 수 있다. 그러나 2점 경계에서는 4문항 모두가 비교적 어려운 문항으로 해석되었다. 변별도의 경우, 1점 경계에서는 문항1, 문항2와 문항3이 전부 .65 이상의 적절한 변별도를 보였다. 문항4는 .494로 약한 변별력을 보였다. 2점 경계에서는 모든 문항에서 변별도가 매우 낮았다.

Table 11은 '생물 농축' 단원의 검사 문항들을 다분 문항 반응 이론을 사용하여 분석한 결과이다. '생물 농축' 단원에서 4개 문항의 평균 난이도는 1점 경계에서 -.228로 적절하였으나 2점 경계에서는 .706으로 어렵다고 분석

**Table 10.** Difficulties and discriminations of the items included in 'reproduction' test in terms of item response theory

Item	Discriminations		Difficulties	
	1 point boundary	2 point boundary	1 point boundary	2 point boundary
Item 1	.690	.062	-1.414	1.041
Item 2	.700	.068	.153	1.116
Item 3	1.417	.187	.170	.845
Item 4	.494	.043	-.943	1.463
Average	.825	.090	-.508	1.116

**Table 11.** Difficulties and discriminations of the items included in 'biological accumulation' test in terms of item response theory

Item	Discriminations		Difficulties	
	1 point boundary	2 point boundary	1 point boundary	2 point boundary
Item 1	.460	.054	-2.588	.508
Item 2	.719	.110	.883	1.229
Item 3	.625	.075	.202	.681
Item 4	.834	.109	.409	.588
Average	.659	.087	-.228	.706

되었다. 1점 경계에서는 문항 1의 경우 -2.588로 쉬웠으나, 문항 2의 경우는 .883으로 어려웠다. 문항 3(.202)과 문항 4(.409)는 중간 정도의 능력 수준을 가진 피험자에게 알맞은 문항이었다. 2점 경계에서는 모든 문항이 적당히 높은 수준의 학습자에게 적합한 문항이라고 할 수 있다. '생물 농축' 단원에서 4개 문항의 평균 변별도는 1점 경계에서 .659로 높았고, 2점 경계에서는 .087로 낮았다. 1점 경계에서 문항 1과 문항 3은 낮은 변별도를 보였고, 문항 2와 문항 4는 적절한 변별도를 보였다. 그리고 2점 경계에서는 모든 문항의 변별도가 매우 낮았다.

(3) 고전 검사 이론과 문항 반응 이론에 따른 문항 분석의 비교

고전 검사 이론은 계산이 용이하므로 현재까지 평가도구의 문항을 분석하는데 보편적으로 사용되어 왔고 지필 검사의 질적 관리를 위해 많은 공헌을 하였다. 이 이론은 우리나라에서 아직까지 평가도구의 질을 분석하는데 절대적으로 사용되고 있다. 반면에 문항 반응 이론은 1980년대 컴퓨터의 발달과 함께 수리적 계산과정이 용이해지면

서 소개되기 시작하였으나 아직은 그 이용이 초기 단계라 할 수 있다(성태제, 1996a). 각 검사 이론을 적용한 문항의 난이도와 변별도를 해석할 때에 각기 다른 기준을 사용하였으므로, 검사 이론에 따른 문항 분석의 차이는 수량적인 것이 아니라, 언어적 해석의 차이이다. 다분 문항 반응 이론을 적용하면 각 점수의 경계에서 난이도와 변별도가 각기 따로 산출되나 고전 검사 이론을 적용하면 난이도는 각 점수의 득점자를 따로 확인하여 계산할 수 있으나 변별도는 점수 경계마다 따로 제공되지 않는다. 그러므로 본 연구에서 고전 검사 이론에 따른 변별도 해석 결과는 다분 문항 반응 이론을 적용했을 때 가장 뚜렷한 변별도를 보인 1점 경계 변별도와 비교하였다.

고전 검사 이론과 다분 문항 반응 이론을 각기 적용하였을 때, 문항 변별도의 언어적 해석에는 거의 차이가 없었다. 단지 '생물 농축' 단원의 문항 1에서 고전 검사 이론에 의하면 변별력이 없다고 해석되었으나 다분 문항 반응 이론에 의하면 변별력이 있다고 해석되었다. 그러나 문항 난이도의 경우에는 두 이론의 해석에 몇 가지 차이를 보였다. 생식단원의 문항 3(1점 경계)에서 고전검사가 이

론에 의하면 어렵다고 해석되었으나 문항 반응 이론에서는 적절하다고 해석되었다. 문항 4(1점 경계)에 대해 고전 검사 이론은 적절하다고 해석하였으나 다분 문항 반응 이론에서는 쉬운 문항으로 해석하였다. 그리고 생식단원의 문항 1(2점 경계), 생물농축단원의 문항 1(2점 경계), 문항 2(1점 경계)에 대해 고전 검사 이론에서는 난이도가 적절하다고 해석하였으나 다분 문항 반응 이론에서는 어렵다고 해석하였다. 이와 같이 변별도에서는 적게 나타나는 언어적 해석의 차이가 난이도에서 다양하게 나타나는 것은 본 연구에서 사용한 두 가지 검사 이론이 피험자 집단에 의존하는 정도가 다르기 때문이다. 같은 검사를 능력이 높은 집단에 실시하였을 경우, 정답자의 비율이 높아지면 고전검사 이론에 의한 난이도 지수가 높아져 쉬운 문항으로 해석된다. 고전 검사 이론에서는 피험자 집단의 특성이 적극적으로 반영되기 때문에, 검사를 실시하는 집단마다 각기 다른 난이도와 변별도 값을 보이게 된다. 그러므로 개발된 수행 평가 문항을 다양한 집단에서 사용하기 위해서는 문항 고유 특성으로서의 문항 모수를 보여주는 문항 반응 이론에 의한 문항 특성의 검증이 필요하다고 생각된다.

#### IV. 결론 및 제언

본 연구에서는 제7차 교육과정에 근거하여 고등학교 과학 중 '생식' 단원과 '생물 농축' 단원에 대해서 필수 학습 요소를 추출하고, 성취 기준, 평가 기준 및 평가 문항을 개발하였다. 또한 개발한 평가 문항이 수행 평가 도구로 사용하기에 적절한지를 알아보기 위해 피험자 응답에 고전 검사 이론과 다분 문항 반응 이론을 적용해 문항 분석을 실시하였다. 연구의 결론은 다음과 같다.

제7차 교육과정이 제시하고 있는 학습 내용 영역과 7종 교과서의 분석을 기초로 '생식' 단원에서 12개, '생물 농축' 단원에서 4개의 필수 학습 요소를 추출하였다. '생식' 단원에서 개발된 성취 기준은 심화 과정을 포함하여 모두 26개로, 지식(K), 탐구(P), 태도(A)의 각 영역에서 12개, 12개, 2개이었다. '생물 농축' 단원에서는 지식영역에서 4개, 탐구영역에서 4개, 태도영역에서 1개로 모두 9개였다. 개발된 성취 기준에 대하여 성취 정도를 판단하기 위한 평가 준거로서 평가 기준을 개발하였다. 개발된 평가 기준의 수는 '생식' 단원에서 총 25개, '생물 농축' 단원에서 9개였다. 성취 기준과 평가 기준에 따라 총 22개

의 수행평가 문항을 개발하였는데 생식단원에서 서술형 9개, 논술형 1개, 포트폴리오 1개이었고 생물 농축 단원에서는 서술형 8개, 논술형 2개, 포트폴리오 1개이었다. 각 평가 문항은 교육과정에 근거한 객관적인 평가에 사용할 수 있도록 구체적인 채점기준을 제시하였다.

개발된 문항 중에서 서술형 문항을 학생들에게 적용해 본 결과를 바탕으로 검사 이론을 적용한 문항 분석을 실시하였다. 전체 8문항을 적용하여 고전 검사 이론을 사용하였을 때의 1점 경계 평균 난이도는 .520, 평균 변별도는 .322로 검사 문항으로 적절하였다. 다분 문항 반응 이론을 사용하였을 때에도 1점 경계에서의 평균 난이도와 평균 변별도가 각각 -.368과 .742로 문항의 질은 비교적 양호한 편이었으므로 수행 평가 문항으로 사용하기에 비교적 적절하였다.

두 가지 검사 이론에 따른 문항 특성의 언어적 해석이 변별도 면에서는 차이가 크지 않았으나 난이도에서는 다양한 패턴으로 나타났다. 이는 고전 검사 이론의 피험자 집단 의존성에 기인한 것이다. 고전 검사 이론의 문항 특성이 피험자 집단에 따라 달라지는데 문항 반응 이론이 피험자 집단의 특성에 구애받지 않는 문항 모수와 피험자 능력 모수를 함께 제공하므로, 문항의 질을 검증하는 데는 문항 반응 이론을 사용하는 것이 바람직하다. 특히, 부분 점수를 주거나 점수가 여러 개의 채점 요소로 구성되는 수행형 문항의 경우에는 고전 검사 이론보다 다분 문항 반응 이론의 적용이 더 적절하다고 생각된다.

위와 같은 결론을 바탕으로 다음과 같은 제언을 하고자 한다.

첫째, 현장 교사들이 수행 평가 문항의 질을 검증하는 데는 문항 반응 이론을 사용하는 것이 더 적절하지만, 현실적으로 프로그램의 활용이 용이하지 않다. 그러므로 피험자의 문항 반응 결과에 문항 반응 이론을 적용시키는 컴퓨터 프로그램들이 좀 더 사용자 친화적으로 개선되어야 할 것이다.

둘째, 양질의 수행 평가 문항을 판별하기 위해서는 정량적인 문항 분석 뿐 아니라, 정성적인 분석이 함께 이루어져야 하므로, 현장에서 수행 평가 문항을 직접 사용해 본 교사들의 피드백을 수용하고, 이에 대한 학생들의 인식도 조사할 필요가 있다. 그리고 수행 평가 문항의 신뢰도를 높이기 위해서는 채점자간 신뢰도를 바탕으로 한 채점 기준의 지속적인 수정·보완이 이루어져 객관적인 타당성을 확보해야 할 것이다.

국 문 요 약

최근 제7차 교육과정이 실시되면서 준거 지향적인 수행 평가에 대한 요구가 커지고 있으며, 이러한 상황에서 평가의 절대적인 기준을 마련할 필요성 또한 절실해지고 있다. 본 연구는 제7차 교육과정에 근거하여 고등학교 과학 중 '생식' 단원과 '생물 농축' 단원에 대해서 필수 학습 요소, 성취 기준, 평가 기준 및 평가 문항을 개발하고, 이를 바탕으로 개발한 평가 문항을 실제 현장에 적용하여 그 타당성을 검증하는 것을 목적으로 한다.

본 연구에서는 제7차 교육과정과 고등학교 과학의 7종 교과서를 바탕으로 하여, '생식'과 '생물 농축' 단원의 필수 학습 요소를 추출하고, 준거 지향 평가의 기초가 되는 객관적이고 타당한 성취 기준 및 평가 기준을 마련하였다. 추출된 필수 학습 요소는 '생식' 단원에서 12개, '생물 농축' 단원에서 4개이었으며, 개발된 성취 기준은 '생식' 단원에서 총 26개, '생물 농축' 단원에서 총 9개로 각각은 지식(K), 탐구(P), 태도(A)의 세 개 영역으로 구분하여 개발하였다. 이상과 같이 개발된 성취 기준을 바탕으로 평가 기준을 개발하였다. 개발된 평가 기준의 수는 '생식' 단원에서 총 25개, '생물 농축' 단원에서 9개였다. 성취 기준과 평가 기준에 따라 '생식' 단원과 '생물 농축' 단원에서 개발된 평가 문항의 수는 서술형 문항이 17개, 논술형 문항이 3개, 포트폴리오가 2개로 총 22문항이었다. 각 평가 문항은 구체적인 채점 기준을 제시하였으므로, 교육과정에 근거한 객관적인 평가에 사용될 수 있다.

개발된 평가 문항 중, 서술형 문항 8개를 경기도 소재 고등학교 1학년 학생 240명을 대상으로 적용·분석하였다. 서술형 문항에 대한 학생들의 응답을 두 가지 검사 이론을 적용하여 분석하였는데, 평가 문항의 기본적인 양호도 검증은 실시한 결과, 본 연구에서 개발된 서술형 문항들은 난이도와 변별도 면에서 전반적으로 적절한 지수를 보여 수행 평가로 사용하기에 적절한 문항들이었다. 개발한 서술형 수행 평가 문항들은 미리 개발된 성취 기준과 평가 기준을 바탕으로 하였으므로 교수 학습의 목표와 내용에도 적합하였다. 고전 검사 이론과 다분 문항 반응 이론에 의해 분석된 문항의 변별도에는 언어적인 해석의 차이가 없었으나, 문항 난이도의 경우는 고전 검사 이론을 사용하였을 때 집단의 특성이 많이 나타나 다분 문항 반응 이론을 사용한 결과와 비교하여 해석의 차이가 다양하게 나타났다.

참 고 문 헌

강순희(2001). 과학과 수행 평가 문항 - 제7차 교육과정 중학교 1학년. 서울: 자유 아카데미.

교육부(1998). 제7차 과학과 교육과정(제7차 교육과정 교육부 고시 제1997-15호[별책9]). 서울: 대한교과서주식회사.

\_\_\_\_\_ (1999). 고등학교 교육과정 해설(제7차 교육과정 교육부 고시 제1997-15호[별책9]). 서울: 대한교과서주식회사.

김원중, 이강주, 오농수, 정태주, 정용직(1995). 학업성취도 평가 문항 작성 자료-지역영역(교육자료 95-8). 대전: 대전광역시 교육연구원.

김찬중, 서만석, 김희백, 심재호, 현종오, 한인욱, 권성기, 박성식(2001). 고등학교 과학. 서울: (주)도서출판 디딤돌

남명호(1995). 수행평가의 타당성 연구. 교육평가연구, 8(2), 115-143.

남현우(1998). 수행평가의 측정학적인 문제들에 관한 고찰. 교육평가연구, 11(2), 1-21.

박정(2001). 문항반응이론을 활용한 수행형 평가문항 분석 방법. 교육학 연구, 39(2), 215-232.

백순근, 소경희, 김광주(1998). 국가 교육과정에 근거한 평가 기준 및 도구 개발 연구: 총론(연구보고 RRE 98-3-1). 서울: 한국교육과정평가원.

성민웅, 김봉곤, 조성동, 강대훈, 강충호, 구자욱, 노일환, 이용철, 임태훈, 최범선, 한인택(2001). 고등학교 과학. 서울: 문원각.

성태제 편역(1991). 문항반응이론 입문. 서울: 양서원.

성태제(1996). 타당도와 신뢰도. 서울: 양서원.

오규환, 이춘우, 오두환, 김영유, 경제복, 이경훈, 박태운, 이영직, 백수관, 김병인, 김봉래, 이기영(2001). 고등학교 과학. 서울: (주)중앙진흥교육연구소.

이규석, 조희형, 박봉상, 박문수, 심국석, 심종섭, 최진복, 장정찬, 이창진, 이용준(2001). 고등학교 과학. 서울: 대한교과서(주).

이문원, 전성용, 최병수, 권석민, 노태희, 허성일, 김출배, 강석진, 박희송, 김경호, 김규상, 채광표, 김진만, 정대영(2001). 고등학교 과학. 서울: (주)금성출판사.

이양락, 이선경, 홍미영, 홍재식, 이미경(1998). 국가 교육과정에 근거한 평가기준 및 도구 개발 연구: 고등학교

- 교 공통과학(연구보고 RRE 98-3-7). 서울: 한국교육과정평가원.
- 이연우, 강석본, 김인석, 김성진, 이진우, 안종재, 배미정, 전화영(2001). 고등학교 과학. 서울: 서울교육정보.
- 정완호, 권재술, 김대수, 김범기, 신영준, 우종욱, 이길재, 정진우, 최병순, 황원기(2001). 고등학교 과학. 서울: (주)교학사.
- Baker, F. B.(1985). *The basic of item response theory*. Portsmouth, N. H.: Heinemann.
- Cronbach, L. J.(1984). *Essentials of psychological testing*. 4th ed. New York: Harper & Row.
- Ebel, R. L., Frisbie, D. A.(1986). *Essentials of educational measurement*. 4th ed. Englewood Cliffs, N.J.: Prentice-Hall.
- Glaser, R.(1963). Instructional technology and the measurement of learning outcomes: some questions. *American Psychologist*, 18.
- Glaser, R., Lesgold, A., Lajoie, S.(1987). *Toward a cognitive theory for the measurement of achievement*. R. R. Ronning, J. A. Glover, J. C. Conoley, J. C. Witt (Ed.). The influence of cognitive psychology on testing. Hillsdale, NJ: Erlbaum.
- Muraki, E., Bock, R. D.(1998). *PARSCALE: IRT based test scoring and item analysis for graded open-ended exercises and performance tasks*. Chicago, IL: Scientific Software.
- Tyler, L. E.(1963). *Tests and measurements*. Englewood Cliffs: Prentice-Hall.
- Whitehead, A. N.(1967). *The aims of education and other essays*. London: Williams & Norgate.