



과학사에 근거한 진화개념검사도구의 타당도 확인 및 맥락에 따른 진화개념 발달 탐색

하민수*
강원대학교

Examining the Validity of History-of-Science-Based Evolution Concept Assessment and Exploring Conceptual Progressions by Contexts

Minsu Ha*
Kangwon National University

ARTICLE INFO

Article history:

Received 3 June 2016

Received in revised form

16 June 2016

27 June 2016

Accepted 27 June 2016

Keywords:

evolution, history of science, ordered multiple choice, Rasch analysis, conceptual progression

ABSTRACT

Previous studies have investigated the similarity between the development of evolutionary explanations and students' conceptual developments on evolution. However, the validity and reliability of the assessment method reflecting the similarity have not been quantitatively examined yet. In addition, no study has examined the conceptual progressions of evolution concept based on contexts although literature has addressed the contextual difference of evolutionary explanation in the history of science. This study examined the validity and reliability of history-of-science-based evolution concept assessment using ordered multiple choice (OMC) methods and Rasch analysis and explored conceptual progression by three contexts (e.g., human, animal, and plant). The evolution concept assessment developed by Ha (2007) was used to examine 1711 elementary, middle, and high school students, and pre- and in-service science teachers' (biology majors and non-majors) evolution concepts. Internal consistency reliability and item response fitness of the OMC method that provide 0- to 4-point scores to creationism, teleology, intentionality, use/disuse, and natural selection respectively met the benchmark based on the Cronbach alpha and MNSQ indices of Rasch analysis. The level of elementary and middle school students' evolution concepts were located between intentionality and use/disuse while the level of high school and non-biology science teachers' evolution concepts were located between use/disuse and natural selection. The conceptual progressions of evolution concepts were differentiated according to three contexts. This study provided the quantitative evidence for the similarity between the development of evolutionary explanations and students' conceptual developments on evolution and suggest new analysis methods (i.e., OMC) of evolution concept assessment.

1. 서론

학생들의 과학개념이 어떻게 발달하는지 살펴보면 과학사에서 과학개념이 발달한 과정과 상당한 유사성을 발견하게 된다(Clement 1983; Ha, Nehm, 2014; Van Driel *et al.*, 1998; Wisser, Carey, 1983). 학생들의 역학 개념의 발달을 조사한 Clement (1983)의 연구에서 뉴턴 이전과 이후의 사고의 변화과정과 학생들의 개념발달 과정이 상당히 유사하다는 것을 확인하였고, 화학 평형에 대한 학생들의 개념을 조사한 Van Driel *et al.* (1998)의 연구에서도 학생들의 개념과 과학사적 발달과의 유사성을 확인하였다. 과학사와 학생들의 개념발달의 유사성이 반드시 나타나는 것이 아니라는 연구 결과도 있다. 예를 들어서 열과 온도에 대한 학생들의 개념을 연구한 Wisser, Carey (1983), 광합성 개념을 연구한 Wandersee(1985), 지구의 모양에 대한 학생들의 개념을 연구한 Vosniadou, Brewer(1987)에서는 과학사와 개념발달의 유사성이 단순히 표면적으로 비슷하거나 의미가 없을 수 있음을 지적하였다. Nersessian(1989)이 지적한 바와 같이 인식론적

측면, 사회적 측면 등이 과거와 현재가 다르기 때문에 과학사와 현재 학생들의 개념 수준을 직접 비교하는 것은 한계가 있을 수도 있다. 그럼에도 불구하고 과학사와 학생들의 개념의 유사성이 연구되어야 하는 것은 과학사에서 이론의 발전에서 나타나는 다양한 논증의 요소들이 현대 학생들의 학습에 유용하게 활용될 수 있기 때문이다.

과학사와 학생들의 개념발달의 유사성을 확인하는 연구에서 진화개념은 가장 널리 연구된 영역이라고 할 수 있다. Ha, Nehm(2014)은 퇴화형질의 진화를 설명하는 과정에서 학생들의 개념이해의 어려움이 다윈을 포함한 19세기 자연과학자들이 가진 어려움과 유사함을 확인하였다. Lee, Lee (2006)는 과학사에서 나타나는 진화개념의 발달과정을 통하여 검사지를 구성하고 학생들의 진화 개념을 확인하기도 하였다. Kampourakis, Zogza(2007)는 과학사에서 나타나는 라마르크의 사고와 학생들이 가지고 있는 사고를 비교하여 현대 학생들이 가지고 있는 사고는 라마르크의 사고에 비하여 보다 낮은 수준의 과학적 사고임을 주장하였다.

많은 연구에서 과학사에서 나타나는 진화개념의 발달과정과 학생

* 교신저자 : 하민수 (msha@kangwon.ac.kr)
http://dx.doi.org/10.14697/jkase.2016.36.3.0509

들이 진화개념을 이해하는 방식이 유사하다고 하였고, 그것을 토대로 개념평가도구를 개발하였다. 하지만 과학사적 발달과정을 반영하여 개발된 개념평가 검사도구의 점수를 분석할 때 과학사적 발달과정을 반영하지 않았다. 예를 들어서 이 연구에서 분석하고자 하는 Ha(2007)이 개발한 도구는 여러 연구에서 활용이 되었으나 그 분석법이 실제로 과학사적 발달과정을 따르지 않는다. Ha *et al.*(2010), Ha *et al.*(2012)의 연구에서 Ha(2007)의 도구를 사용하였는데, 창조, 목적, 의도, 용불용설의 과학사에서 나타나는 4가지 대안개념은 과학적 오개념으로 구분하여 0점을 주고, 다윈의 자연선택만 정답으로 간주하여 채점하였다. 이와 같은 방법은 학생들의 자연선택에 대한 과학적 개념의 수준만 측정할 뿐 그들의 사고가 과학사적으로 어느 단계에 있는지 발달적 측면에서는 분석할 수 없다. Ha *et al.*(2006)의 연구에서도 5가지 보기의 비율을 토대로 분석하는 방법을 사용하였는데 이 경우에는 5가지 설명을 동일한 수준으로 고려하여 분석한 것이다. 다시 말하면 5가지 설명이 생물형질의 출현에 관한 독립된 설명이라는 전제를 바탕으로 한 분석이다. 이와 같은 분석이 타당하지 않은 것은 아니다. 하지만 창조, 목적, 의지, 용불용설, 자연선택은 과학사적으로 선후관계가 명확하고, 여러 문헌에서 사고의 발달적인 측면을 강조하고 있기 때문에 이 점을 평가점수에 반영하기 위해서는 빈도분석의 방법으로는 불가능하다. 대신에 개념발달의 수준을 확인하기 위해서는 순진한(*naive*) 사고에서 과학적 사고까지 단계적으로 점수를 부여하는 순위선다형 방법으로 구별할 수 있다. 앞서 소개한 Lee, Lee (2006) 연구뿐만 아니라 목적론, 의도, 용불용설과 같은 대표적인 과학적 오개념을 포함한 진화개념 측정도구인 *Conceptual Inventory of Natural Selection*(Anderson *et al.*, 2002), *Measure of Understanding of Macroevolution*(Nadelson, Southerland, 2009)에서도 목적론과 용불용설 두 설명을 동급으로 취급하고 있다. 이 연구는 창조, 목적, 의도, 용불용설, 자연선택으로 이어지는 과학사적 발달 과정을 반영하여 순위선다형방법으로 분석하는 것이 타당하지 문항반응이론의 라쉬분석과 다양한 통계기법을 활용하여 확인할 것이다. 또한 이와 같은 방법을 통하여 학생들의 개념발달과정을 점검하고, 과학교육에서의 시사점을 확인할 것이다.

II. 진화 개념의 발달 과정에 관한 이론적 배경

이주 오래전부터 사람들은 모든 생명체는 신에 의하여 창조되었다고 믿었다(Johnson, Peeples, 1987). 생물형질의 존재 이유에 대한 의문을 제기하기보다 모든 생명체는 신에 의지에 의하여 태초에 그런 모습으로 창조되었다고 믿었다. 생물의 다양한 형질이 존재하게 된 이유에 관한 철학적 설명은 아리스토텔레스부터이다(Greene, 1992). 아리스토텔레스는 생물들이 특정형질을 가지는 이유는 목적인(*final cause*)라 불리는 형질마다의 특정한 목적 때문이라 강조하였다(Opfer *et al.*, 2012). 대부분의 생물형질들은 기능이 있으며, 기능은 생명현상에 영향을 미친다. 예를 들어서 동물은 눈이라는 기관 없이 앓을 볼 수 없고 생존이 위태로울 수 있다. 그러므로 기능을 위해서 형질이 존재할 것 같은 믿음이 생길 수 있다. 목적론적 사고에 관한 최근 연구에 의하면 목적론은 인지편향이라는 것이 보다 명확해지고 있다(Kelemen *et al.*, 2013; Willard, Norenzayan, 2013). Kelemen(1999a)의 연구에서 살펴보면 목적론은 어릴 때 학습과 관계없이 자연스럽게

는 형성되며, Kelemen(1999b)의 연구에서도 생명체뿐만 아니라 가상의 인공물에서도 학생들은 목적론적 사고를 통하여 형질의 존재이유를 설명하는 것을 관찰하였다. 목적론적 설명은 그럴 듯한 설명이지만, 원인이 결과보다 시간적으로 후에 존재할 수 없다는 철학적 논리에 의하여 기각된다(Jungwirth, 1975). 아리스토텔레스에서 시작된 목적론적 설명은 2천년 가까이 이어지다 서양근대철학의 발달로 논리적 허점을 발견하게 되었다(Osler, 2001). 서양근대철학은 목적론이 논리적으로 바람직한 설명이 될 수 없음을 지적하고, 과학적으로 합리적인 인과관계 추론을 강조하였다. 이와 같은 노력으로 인하여 생물의 형질의 존재 이유 역시 목적과 같은 논리적으로 타당하지 않은 원인이 아닌 다른 원인에 의하여 발생할 것으로 과학자는 믿게 되었다.

창조적 사고는 개인의 신념이며, 목적론적 사고는 과학적 인과관계가 성립하지 않는 비논리적 추론이다. 하지만 18세기 생물학자들은 생물이 환경에 적합한 형태로 진화하고 있으며, 새로운 형질들이 출현하고 있음을 인식하고는 있었으나(Bowler 2009, p. 78), 어떻게 그런 일이 발생하는지 답을 제시하지는 못하였다. 그때 제시된 이론들 중 하나가 생물의 의도에 의한 형질의 변화이다(Bowler 2009, p. 86). 에라스무스 다윈(Erasmus Darwin)은 생물의 노력이나 의지가 형질변화를 일으킬 수 있다고 생각하였다(Bowler 2009, p. 86). 이와 같은 설명은 기존의 목적론에 비하여 논리적으로 타당한 과학적 인과관계를 가지고 있다. 하지만 생물의 의지와 노력이라는 추상적인 요소가 형질 변화의 직접적인 원인이 될 수 있는지에 대한 의문은 당대에도 많았다(Bowler 2009). 이와 같은 과정에서 출현한 이론이 라마르크의 용불용설이다. 현행 교육과정에서 라마르크의 용불용설은 대표적인 과학적 오개념으로 인식되고 있으나 당대에는 혁명적인 사고였다.

라마르크의 용불용설이 상당히 과학적 사고였음은 여러 연구에서 확인되었다. Kampourakis, Zogza (2007)는 학생들이 생각하는 용불용설과 라마르크가 생각하는 용불용설은 달랐음을 강조하고 있다. 그리스 학생들을 대상으로 조사한 결과 학생들이 가지고 있는 용불용설에 대한 일반적인 생각은 환경의 변화로 인한 새로운 먹잇감을 획득하기 위하여 형질을 사용할 필요성이 증가되고 그에 따라 형질이 변화하였다는 목적론적 설명에 가까운 반면, 라마르크의 생각은 지속적인 사용과 사용하지 않음이 신체의 형질을 변화시켰다는 것이다. 또한 라마르크의 과학적 추론에는 상당한 수준의 증거가 있었다(Kampourakis, Zogza, 2007). 다윈이 종의 기원을 발표하고 난 이후 라마르크의 생각을 가지고 있던 과학자들과 다윈의 생각을 가지고 있던 과학자들의 논쟁은 상당히 길었으며, 퇴화형질에 대한 설명을 다윈의 생각으로는 납득이 가능한 수준의 설명을 만들지 못한다는 이유로 19세기 후반까지 미국의 자연과학자들은 거의 대부분 라마르크의 생각을 가지고 있었다(Ha, Nehm, 2014). 돌연변이와 같은 구체적인 유전지식과 생태학과 집단유전학과 같은 거시적인 수준에서 생물학을 이해하지 못하는 경우 다윈의 사고를 받아들이기 힘들며, 그런 상황에서 용불용설은 인과관계가 명확한 상당히 그럴듯한 과학적 설명이었다.

라마르크의 용불용설 이후에 등장한 이론이 다윈의 자연선택이론이다. 자연선택의 핵심은 변이, 유전성, 차별적 생존과 생식이다. 비록 다윈이 변이와 유전성의 기작에 관한 설명을 이해하고 있지 않았으나 경험적으로 변이와 유전성을 이해하고 있었다. 또한 다윈은 과잉생산과 제한된 자원을 통한 경쟁, 그리고 궁극적으로 차별적인 생존과

생식을 통한 개체군의 변화를 생물 형질의 주요 원인으로 제시하였다 (Ha, Nehm, 2014). 유전학의 발달로 인하여 다윈의 자연선택은 진화론의 중요 개념으로 인정된다. 다윈의 자연선택이론이 제안된 이후에 라마르크의 용불용설과 상당히 많은 논쟁이 진행되었다. 그 논쟁의 중심에는 퇴화 형질을 어떻게 설명할 것인가 하는 문제도 포함되었다 (Ha, Nehm, 2014). 유전학적 지식의 부족으로 인하여 최초의 변이가 어떻게 형성되는지에 대한 이해 없이 용불용설은 그럴듯한 설명일 수밖에 없었다. 이 과정에서 독일 생물학자인 아우구스트 바이스만 (August Weismann)이 생식세포에서 일어나는 유전자의 변화만 자손으로 유전된다는 점을 확인하면서 획득형질의 유전을 전제로 한 용불용설은 쇠퇴하게 된다(Churchill, 1968).

이상의 설명에서 생물형질의 존재이유에 관한 설명이 창조, 목적, 의도, 용불용설, 자연선택으로 이어짐을 확인할 수 있다. 종교적 신념인 창조론을 제외한 목적, 의도, 용불용설의 세 가지는 과학사에서 나타나는 대표적인 비과학적 개념으로 진화개념평가도구에서 오답으로 사용되고 있다(Anderson *et al.*, 2002; Nadelson, Southerland, 2009). 하지만 철학적 추론인 목적론, 인과관계에 대한 초기 사고인 내부 의도, 명확한 인과관계와 증거를 동반한 용불용설로 이어지면서 과학적 사고는 점차 확장되었다. 현대 진화론을 제외한 모든 개념들이 비과학적 오개념이지만, 사고와 논리 수준이라는 점에서 보았을 때 목적론에 비하여 용불용설은 상당히 진보된 과학적 설명이다. 그러므로 학생들의 진화적 사고의 수준을 과학사적인 발달과정을 고려하여 확인할 필요가 있다. 따라서 각 설명별로 차등적으로 점수를 부여하는 방법을 고려할 수 있을 것이다.

각 설명단계별로 점수를 차등적으로 부여하면, 학습자가 어느 단계에 위치하고 있는지도 확인할 수 있다. 이럴 경우 교수학습이라는 관점에서 여러 가지로 유용하다. 그 이유는 학습자의 사고의 발달이 과학사와 유사하며, 각 과학사적 시점에 있었던 많은 논쟁들이 학습자의 개념발달에 유용하게 사용될 수 있기 때문이다. 예를 들어 목적론이 극복되면 시점은 근대철학이 발생하던 시점이었으며, 근대철학자들이 강조한 목적론의 철학적 비논리성이 학생들의 개념변화를 위한 중요한 학습 자료가 될 수 있다. 학습자가 용불용설 수준에 있다면 라마르크를 추종하던 용불용설 학파들과 자연선택 학파들 간의 있었던 수십 년간의 과학적 논쟁이 개념발달을 극복하는 중요한 학습소재가 될 수 있다. 진화론적 사고가 어느 한 순간 생성된 것이 아니라 목적론에서 내부의도, 내부의도에서 용불용설, 용불용설에서 자연선택으로 단계적으로 이어져 왔으므로 진화 교육이 이와 같은 사고발달의 순서에 맞게 개발되면 보다 효율적일 수 있다.

위의 필요성을 근거로 이 연구에서는 창조론적 사고, 목적론적 사고, 의도적 사고, 용불용설적 사고, 자연선택적 사고에 과학사적 발달을 고려하여 순위를 적용한 방법이 타당하지 확인하였다. 두 번째로

초·중등 학생들의 진화 개념이 과학사적으로 어느 단계에 위치하고 있는지 확인하였다. 이 연구문제를 달성하기 위하여 방법적으로 순위선다방식(ordered multiple choice)을 활용하였고, 부분점수모델(partial credit model)을 사용한 라쉬분석을 활용하였다.

자세한 연구문제는 다음과 같다.

1. 진화 개념의 과학사적 발달 과정을 반영한 순위선다방식의 문항 반응 타당도는 어떠한가?
2. 라쉬분석을 통해 확인한 문항의 수준과 학생들의 능력을 근거로 추정된 초·중·고 학생들의 진화 개념 수준은 어떠한가?

III. 연구 방법

1. 진화개념 검사도구

이 연구에서 사용한 진화개념 검사도구는 Ha(2007)이 개발한 검사도구이다. 이 검사도구는 종교적 신념을 비롯한 과학사에서 나타나는 대표적인 대안개념인 목적, 의도, 용불용설의 세 가지 비과학적 개념과 돌연변이와 자연선택의 현대적 진화 개념의 5가지 보기를 통하여 생물형질의 존재 이유에 대한 질문에 답하는 형태를 취하고 있다. 역사적으로 가장 오래된 사고인 창조론적 사고를 1단계로 하고, 목적론을 2단계, 의도를 3단계, 용불용설을 4단계, 그리고 마지막으로 자연선택적 사고를 5단계로 구분하였다. 이 검사도구는 학생들의 진화 개념이 문항의 맥락(인간, 동물, 식물)에 따라 차별적이라는 것을 강조하고 있다. 문항의 맥락과 학생들의 개념에 관한 연구는 이 연구 이후 Nehm, Ha(2011), Opfer *et al.*(2012)에서 계속 강조되었다. 그래서 이 연구에서도 인간, 동물, 식물의 세 가지 상황적 맥락을 구분하여 분석하였다.

2. 참여자

이 연구는 5개의 집단 1711명의 응답을 바탕으로 진행하였다. 두 개의 초등학교에서 수집된 4·5학년의 초등학생 자료는 373명(21.8%, 남학생 52.5%, 여학생 47.5%), 두 개의 중학교에서 수집된 중학생 자료(2학년-3학년)는 537명(31.4%, 남학생 54.0%, 여학생 46.0%), 세 개의 고등학교에서 수집된 고등학생(2학년) 자료는 733명(42.8%, 남학생 34.7%, 여학생 65.3%)이었다. 이 연구의 주된 대상은 초등학생에서 고등학생이나 개념의 발달 단계를 면밀히 확인하기 위해서는 전문가 집단(이 연구에서는 생물전공교사)의 자료와 비교할 필요가 있기 때문에 자료수는 상대적으로 적으나 생물교사자료를 추가하였다. 생물비전공교사는 예비교사가 32명과 현직이 3명으로 35명(2.0%),

Table 1. Example of evolution concept assessment

Level	Ideas	Answers on “The giraffe has a longer neck than other animals. Why does the giraffe had long neck?”
Level 1	Creationism	In the beginning, God created the giraffe’s long neck.
Level 2	Teleology	The giraffe had long neck in order to pick and eat the leaves of tall trees.
Level 3	Intentionality	The giraffe’s desire to pick and eat the leaves of the tall trees made their neck long.
Level 4	Use/Disuse	Giraffes have used their necks continuously to reach the leaves of the tall trees and eat them, so the neck became longer.
Level 5	Natural selection	A long-necked giraffe was born because of a sudden mutation and they multiplied.

남자 40.0%, 여자 60.0%), 생물전공교사는 예비교사가 19명, 현직이 4명으로 23명(1.3%, 남자 34.8%, 여자 65.2%)이었다. 조사과정에서 집단에 응답하지 않은 참여자가 10명이었다.

3. 연구 방법

과학사에 근거한 진화개념검사도구의 응답을 순위선다형점수를 활용하여 분석하는 방법의 타당도는 3차원 부분점수모형 라쉬분석(three dimensional-partial credit model-Rasch analysis)을 사용하여 확인하였다. 18개 문항에 인간, 동물, 식물의 세 가지 맥락에 따라 구분되어 있어서 각 맥락을 분리하여 분석하는 3차원 분석과 맥락 구분 없이 단일 차원으로 하는 방법 중 어느 것이 더 타당한지 확인하였다. 이것을 위하여 Multi-dimensional 라쉬분석이 가능한 ConQuest 4.5.0 프로그램을 사용하였다. 먼저 두 다른 차원을 비교하기 위해서 3차원 라쉬분석과 1차원 라쉬분석에서 생성되는 Final Deviance와 Akaike Information Criterion(AIC)의 값을 비교하는 방법을 사용하였다(Neumann *et al.*, 2011). 피평가자의 분포가 문항의 타당도를 확인하기 위해 적절한지 확인하기 위하여 Item Separation Reliability를 확인하였고, 각 항목별 문항들이 학생들의 능력을 구분할 수 있는지 확인하기 위해 EAP/PV Reliability를 사용하였다.

라쉬분석을 통하여 생성된 MNSQ(mean-square)의 값은 문항에 대한 학생들의 반응을 수치화 할 수 있다. MNSQ가 1은 문항적합도가 최적이며 그 수치가 낮거나 높아질 경우 문항적합도는 나빠진다. MNSQ에 대한 기준은 평가의 특성에 따라 다르게 적용될 수 있는데 이 연구에서는 Wright, Linacre(1994)가 제시한 것을 토대로 분석하였다. 고부담 선택형 평가의 경우 0.8-1.2가 적합한 것으로 PISA(The Programme for International Student Assessment) 평가가 이 기준을 활용하고 있다. 선택형이면서 평범한 평가일 경우에는 0.7-1.3, 설문이면서 평정척도일 경우에는 0.6-1.4, 임상관찰법일 경우에는 0.5-1.7

등을 기준으로 삼고 있다(Wright & Linacre, 1994). 이 연구에서 사용한 검사 도구는 저부담의 일반적인 개념평가도구이면서 부분점수모형을 사용하였으므로 0.6-1.4가 적절할 것으로 판단된다. 다음으로는 3차원 분석을 통해 생성된 3가지 점수(인간, 동물, 식물문항)의 점수를 바탕으로 다변량분산분석(MANOVA, Multivariate analysis of variance)을 하였다. 이 분석에서 생성되는 각 변인별 분산 분석값도 제시하였다. 분산분석에서 효과크기는 partial eta squared(PES)를 사용하였다. 각 연구대상별로 문항별 점수를 대응표본 t-검정을 통하여 비교하였다. t-검정에서는 효과크기를 Cohen' d로 확인하였다. 라쉬 분석을 제외한 분석은 SPSS 22버전을 사용하였다.

IV. 연구 결과 및 논의

Final Deviance와 Akaike Information Criterion 점수 비교를 통하여 인간, 동물, 식물 문항을 하나의 구인으로 분석하는 것과 세 개의 구인으로 나누어 분석하는 것 중 어떤 모형이 더 적합한지 확인하였다. 3차원 분석의 경우에는 Final Deviance가 77644, AIC가 77800이었으며, 1차원 분석에서는 Final Deviance가 77612, AIC가 77758으로 각각 비율이 99.959%, 99.946%로 거의 일치하였다. 이 값을 근거로 1차원 모형보다 많은 정보를 취할 수 있는 3차원 모델을 사용하는 것이 타당함을 확인하였다. 이와 같은 결과는 인간, 동물, 식물에 따라 학생들이 진화 개념이 다르다는 Ha *et al.* (2006)의 연구를 토대로 분석이 가능함을 보여준다. 또한 맥락에 따라 학생들의 진화 개념이 다르다는 Nehm, Ha(2011)의 연구와도 일맥상통하다. 학생들의 이 연구에서 사용한 문항의 신뢰도(item reliability)는 0.974, 학생들의 능력치를 구분 짓는 수치인 EAP/PV Reliability는 인간문항이 0.827, 동물문항이 0.859, 식물문항이 0.864으로 모두 0.8 이상의 높은 수준의 신뢰도를 보였다(Fauth *et al.*, 2014). 라쉬분석을 통해 생성된 삼차원모형에서 생성된 세 가지 점수의 상관관계는 인간문항과 동물문항

Table 2. Cronbach's Alpha if items deleted and overall Cronbach's Alpha

	Item	Cronbach's Alpha if Item Deleted	Cronbach's Alpha
Human items	S-shaped backbone	0.620	0.664
	Brain size	0.603	
	Cholesterol resistance	0.666	
	Walking upright	0.645	
	Hand shape	0.599	
	Vocal cords	0.593	
Animal items	Bear fat	0.670	0.727
	Chameleon color	0.682	
	Cheetah speed	0.680	
	Giraffe neck	0.706	
	Hawk eyesight	0.677	
	Leafhopper pesticide resistance	0.720	
Plant items	Bracken vascular bundle	0.747	0.761
	Cactus leafless	0.727	
	Leaf size	0.725	
	Lily stoma	0.715	
	Seeds covered in hooks	0.723	
	Rose scent	0.717	

Table 3. Unweighted and weighted MNSQ of items

Item number	ESTIMATE	Human items		ESTIMATE	Animal items		ESTIMATE	Plant items	
		Unweighted MNSQ	Weighted MNSQ		Unweighted MNSQ	Weighted MNSQ		Unweighted MNSQ	Weighted MNSQ
1	0.083	0.97	0.95	-0.034	0.93	0.94	0.079	1.16	1.13
2	0.000	1.00	0.99	0.111	0.95	0.96	-0.030	1.08	1.09
3	-0.303	1.26	1.19	-0.081	1.05	1.10	-0.058	1.04	1.07
4	0.033	1.17	1.11	0.104	1.11	1.05	0.122	1.02	1.01
5	0.009	0.91	0.93	0.019	0.98	0.99	-0.011	1.03	1.05
6	0.178	0.93	0.93	-0.119	1.25	1.20	-0.101	1.04	1.05

의 점수가 0.900, 인간문항과 식물문항이 0.904, 동물문항과 식물문항이 0.941로 유사하였다.

라쉬분석과 함께 문항의 내적일관성 신뢰도 역시 확인하였다. 내적 일관성신뢰도는 Cronbach alpha를 활용하였다. 인간문항의 alpha값은 0.664, 동물문항의 alpha값은 0.727, 식물문항의 alpha값은 0.761로 내적일관성이 적정수준에 있음을 확인하였고, 인간문항에서 내적 일관성 신뢰도가 가장 낮음을 확인할 수 있었다. 전체 내적 일관성 신뢰도보다 더 중요한 것이 모든 문항이 고루 각 구인에 기여하는지 확인하는 것이다. 학생들의 응답이 특정문항에서 5가지 응답에 대한 다른 경향을 보인다면 학생들의 반응의 일관성에는 문제점이 있다고 볼 수 있다. Table 2를 보면 문항이 제거되었을 때의 alpha값과 전체 alpha값을 비교해 놓았다. 콜레스테롤 저항성 문항의 경우 제외되었을 경우 0.666으로 전체 alpha값인 0.664에 비하여 약간 상승하였으나 의미가 있는 수치는 아니었다. 그것을 제외하고 전체 alpha값에 비하여 인간, 동물, 식물 각 항목 당 문항은 전체 구인에서 나타나는 일관성에 영향을 주지 않는 것을 확인할 수 있다.

내적일관성신뢰도의 기준에 대한 의견은 문헌에 따라 약간의 차이가 있다. 또한 이 분석에서 사용하는 것과 같은 개념평가에서 부분점수모형일 때와 응답의 척도가 분명한 Likert 척도문항의 경우 내적일관성신뢰도 alpha값의 기준은 다르게 적용되어야 할 것이다. 정신측정학에서 일반적으로 통용되는 alpha값의 기준점은 0.7이다(George & Mallery, 2003). 일부 문헌에서는 0.6의 경우도 분석가능한 수준의 일관성이 있다고 보고하고 있다(Hair et al., 2006). 이와 같은 상황들을 근거로 판단하였을 때 0.664의 인간문항과 0.727의 동물문항, 0.761의 식물문항의 신뢰도는 의미 있는 수준이라고 판단된다. 또한 이 결과는 창조론에서 자연선택까지 5가지의 보기를 제시하여 하나를 선택하게 하는 방법에서 학생들이 다양한 상황(문항)이 있음에도 자신의 생각을 일관성 있게 제시했음을 확인할 수 있다.

고전검사이론에 근거한 Cronbach's Alpha보다 문항반응이론에 근거한 라쉬분석에서 보여주는 문항의 MNSQ의 값은 문항의 적합도를 판단하는데 더욱 적합하다(Neumann et al., 2011). Table 3에서는 18개 문항의 MNSQ의 값을 보여준다. 보다 엄격한 수치인 Unweighted MNSQ를 확인하면 전체 18개 문항 모두 평정척도를 사용한 문항을 평가할 때의 기준인 6-1.4 기준에 만족하고 있다. 또한 이 연구에서 사용한 방법이 부분점수모형이지만 더 엄격한 선택형 모형의 MNSQ 기준인 7-1.3의 기준에도 모두 만족한다. 이 결과를 통하여 5가지 과학적으로 구분된 응답을 활용한 부분점수모형방식의 접근은 상당히 높은 수준의 타당도가 있음을 확인할 수 있다. 인간문항, 동물문항, 식물문항의 MNSQ값을 통하여 비교해보면, 인간문항이 세 가지 항목

중에서 가장 낮은 수준의 적합도를 보이는 것을 확인할 수 있다. 이 결과와 앞서 제시한 내적일관성 신뢰도 두 가지를 토대로 판단하였을 때 학생들의 반응이 인간진화문항에서 일관성이 떨어지고 문항반응이 상대적으로 부적합함을 확인할 수 있다.

Figure 1은 3차원 부분점수모형 라쉬분석에서 생성된 Wright Map을 보여주고 있다. 오른쪽에는 문항들의 난이도에서 상위 단계로 넘어가는 임계치(threshold)에 대한 척도(item measure)값이, 왼쪽에는 학생들의 능력치에 관한 척도(person measure)값이 제시되어 있다. 문항의 난이도와 학생들의 능력치 모두 인간, 동물, 식물문항으로 구분되어 제시되어 있다. 단계별 임계치를 살펴보면, 창조론적 사고에서 목적론적 사고로 넘어가는 수준이 평균 -1.7정도이며, 목적론적 사고에서 의도적 사고는 -0.5, 용불용설로 넘어가는 수준은 약 0.0, 마지막으로 용불용설에서 자연선택으로 넘어가는 수준은 약 2.0 수준이다. 학생들의 능력치 분포를 살펴보면 인간문항에 대한 로짓값이 0.0 수준이 49.09%, 0.5수준이 31.15%로 대부분이었으며, 동물문항에 대한 로짓값은 0.0수준이 31.85%, 0.5수준이 32.96%, 1.0이 19.52%으로 인간문항에 비하여 보다 높은 수준이었다. 식물문항에 대한 로짓값은 -0.5수준이 11.75%, 0.0수준이 35.18%, 0.5수준이 23.50%, 1.0수준이 15.72%로 범위의 분포가 다른 문항에 비하여 넓은 특징이 있었다. 전체적으로 보았을 때 용불용설에서 자연선택적 사고로 넘어가는 단계가 가장 어려운 것으로 나타났으며, 그 다음은 목적론에서 의도적 사고로 넘어가는 것이다. 의도적 사고에서 용불용설로 넘어가는 수준은 상대적으로 크지 않았다.

Figure 2는 각 집단별 학생들의 능력치의 평균값과 표준오차를 보여준다. 통계적 검정을 위하여 인간문항, 동물문항, 식물문항의 라쉬점수를 토대로 MANOVA를 활용하여 분석하였다. 그 결과 집단별 유의미한 차이를 확인할 수 있었다($F[12,4482]=22.02$, $p=0.000$, $PES=0.049$). ANOVA의 결과 역시 세 항목 모두 유의미하였으며, 그 중에서도 동물문항에서 가장 큰 효과크기를 보였다(인간문항: $F[4,1696]=52.23$, $p=0.000$, $PES=0.110$; 동물문항: $F[4,1696]=56.66$, $p=0.000$, $PES=0.118$; 식물문항: $F[4,1696]=48.41$, $p=0.000$, $PES=0.102$).

인간문항의 경우 초등학생이 -0.231, 중학생이 -0.188, 고등학생이 -0.031, 비전공교사가 0.004, 전공교사가 1.217 수준이었다. 동물문항의 경우 초등학생이 -0.053, 중학생이 0.010, 고등학생이 0.220, 비전공교사가 0.261, 전공교사가 2.104으로 인간문항에 비하여 상대적으로 높았다. 식물문항은 초등학생이 -0.234, 중학생이 -0.150, 고등학생이 0.079, 비전공교사가 0.098, 전공교사가 2.135로 동물문항과 비슷한 수준이었다. Tukey 방법을 활용한 사후검정을 한 결과 유의미한

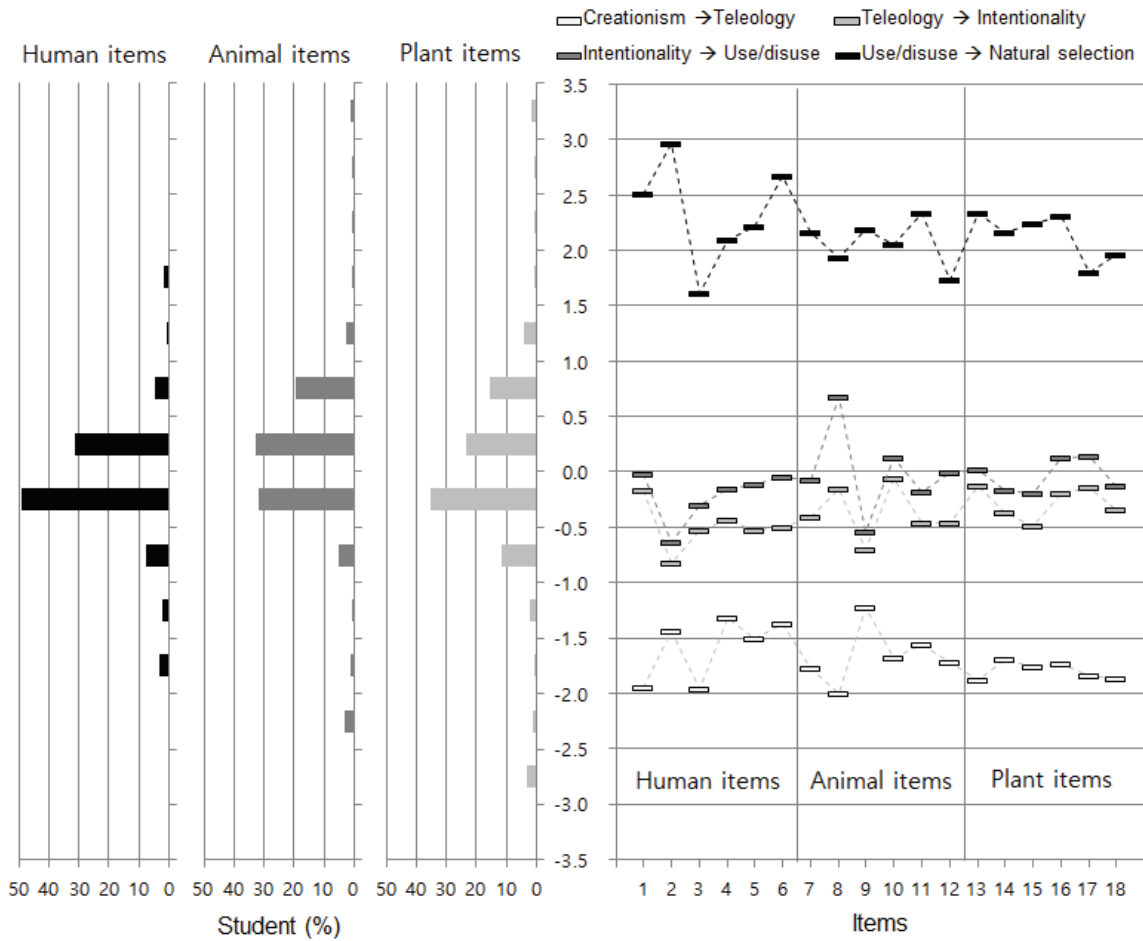


Figure 1. Wright map of 3-dimensional partial credit model Rasch analysis

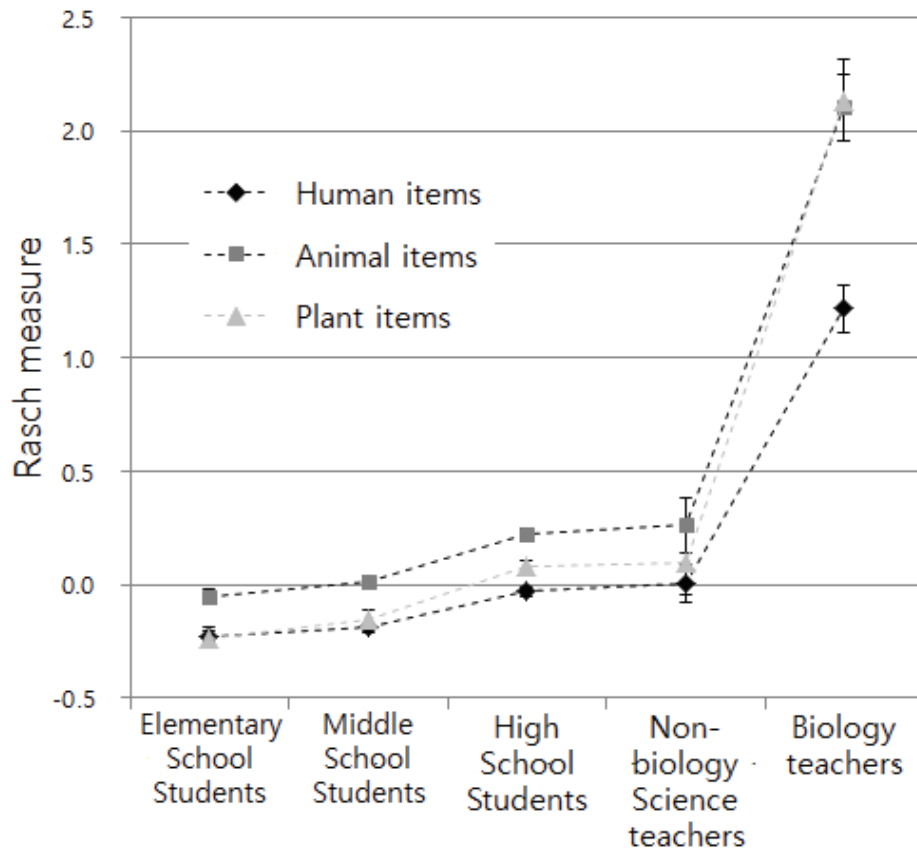


Figure 2. The change of Rasch measures across groups

수준의 향상은 중학교와 고등학교에서 나타났으며, 비전공교사 이후에 전공교사에서 마지막 단계로 도약하는 것을 확인할 수 있었다.

마지막으로 대응표본 t-test를 통하여 인간, 동물, 식물의 상황에서 학습의 발달이 어떤 차이점을 보이는지 확인하였다. 동물의 진화의 경우에는 지학년일 때 단계가 의도 또는 용불용설 이상의 수준에서 출발하는데 비해, 식물과 인간의 진화는 그보다 낮은 목적론 수준이었다. 가장 높은 개념발달 수준을 보이는 전공교사의 경우에도 인간 진화의 경우는 동물과 식물의 진화에 비하여 상당히 낮은 수준의 개념발달을 보이고 있다. 그리고 그 수준은 Cohen's d가 0.71, 0.68로 큰 효과크기를 보이고 있다.

이상의 결과로부터 논의하면 다음과 같다. 순위선다방식(ordered multiple choice)을 활용하여 진화 개념에 대한 과학사적 발달 과정(창조론적 사고, 목적론적 사고, 의도적 사고, 용불용설적 사고, 자연선택적 사고)에 따라 부분점수를 주는 방법은 통계적으로 타당하였다. 내적일관성신뢰도, 문항반응 적합도 등에 관한 다양한 기준을 적용하였을 때 과학사적 발달순서에 따라 순위점수를 주는 방법은 타당하고 신뢰할 수 있다. 학생들의 응답은 18개의 문항에서 일관성이 있었으며, 문항의 반응 역시 적합하였다. 그러므로 Ha *et al.*(2010), Ha *et al.*(2012)의 연구에서 다윈의 자연선택만 정답으로 간주하여 채점하는 방법보다 창조에서 자연선택으로 이어지는 사고의 단계를 부분점수화 한다면 많은 정보를 평가결과로부터 취할 수 있을 것이다. 예를 들어, 학생들의 사고 수준이 다섯 단계의 사고 발달과정에서 어느 수준인지 확인할 수 있다. 또한 Ha *et al.*(2006)의 연구에서는 창조론에서 자연선택까지 다양한 개념을 빈도로 비교하였는데, 빈도분석보다 비율적도인 라쉬점수를 통하여 분석하면 더 다양한 통계방법이 활용될 수 있다. 목적론, 의도, 용불용설을 오답지로만 활용한 Conceptual Inventory of Natural Selection(Anderson *et al.*, 2002), Measure of Understanding of Macroevolution(Nadelson, Southerland, 2009)에서도 학생들이 선호하는 응답에서 나타나는 오개념이 어떤 것인지 확인하고 부분점수를 제공하는 방법이 제안될 수 있다. Kampourakis, Zogza (2007)이 지적한 바와 같이, 용불용설은 목적론이나 의도에 비하여 상당히 과학적인 생각이었으며 결과적 지식이

아닌 과정적 지식 역시 중요한 과학적 지식으로 인정하는 과학교육에서는 사고의 발달을 확인하는 방법은 필요하다. 그런 면에서 과학사에 근거하여 창조론에서 자연선택까지 부분점수모형으로 평가하는 방법은 과학교육의 여러 목표를 달성하는데 많은 장점들이 있는 것으로 판단된다.

두 번째로 인간, 동물, 식물 문항의 반응의 3차원 라쉬분석을 통한 문항의 맥락에 따른 개념발달은 차별적이었다. 동물문항의 경우 지학년도 용불용설 단계에서 출발하는 반면, 식물문항과 인간문항은 의도 수준에서 개념발달이 시작하는 것을 확인할 수 있다. 라마르크의 용불용설은 'Philosophie Zoologique'의 동물학 관련 책에서 제안하였다. 라마르크뿐만 아니라 당대 진화론에 대한 많은 논쟁들은 실제로 동물의 진화와 관련이 많다. 당시에 동물의 경우 의지가 있고, 특정형질을 의도를 활용해서 많이 사용할 수 있는 능력이 있기 때문이다(Kampourakis, Zogza (2007). 에라스무스 다윈이 의도가 형질의 변화를 초래한다고 한 책도 동물관련 서적(Zoonomia)이다. 창조에서 자연선택까지 5가지 설명에서 의도와 용불용설은 동물의 형질과 보다 어울리며, 반대로 식물의 경우 그 두 가지 사고가 적용되기 어려울 수 있다. 하지만 특이한 점은 현대적 관점에서의 용불용설인 후성유전학(epigenetics)의 경우 식물학에서 더욱 더 활발하게 연구되고 있다(Grant-Downton & Dickinson, 2005). 실제 생물학에서는 식물에서 획득형질의 유전이 더 많이 관찰됨에도 학생들은 동물소재를 중심으로 용불용설 개념을 형성하여 용불용설적 설명을 식물에게 쉽게 적용할 수 없는 것이다. 인간의 경우 돌연변이와 자연선택적 설명으로 개념이 발달이 다른 항목에 비하여 쉽게 이루어지지 않는 것 역시 진화에서 인간중심적 사고가 있음을 반영하고 있다(Seoh *et al.*, 2016). 맥락에 따라 개념의 수준이 다르다는 것은 기존의 연구에서도 여러 차례 확인된 바 있다(Ha *et al.*, 2006; Nehm & Ha, 2011; Opfer *et al.*, 2012). 선행연구와 달리 이 연구의 결과는 횡단 자료의 추가로 맥락에 따라 개념 수준의 발달이 다르다는 논의 역시 가능하다. 맥락에 따라 학생들의 진화에 대한 설명의 수준이 차등적이라는 것은 평가와 교수방법에 관한 다양한 함의를 제공한다. 예를 들어 학생들이 돌연변이와 자연선택의 진화개념을 기린의 목뿐만 아니라 식물의 잎

Table 4. Paired sample t-test between assessment items

Subjects	Comparison	Mean difference	t	df	sig.	d
Elementary school student	Human - Animal	-0.178	-17.63	372	0.000	0.42
	Human - Plant	0.003	0.22	372	0.829	0.01
	Animal - Plant	0.181	21.46	372	0.000	0.34
Middle school student	Human - Animal	-0.198	-19.88	536	0.000	0.37
	Human - Plant	-0.038	-2.47	536	0.014	0.06
	Animal - Plant	0.160	19.56	536	0.000	0.23
High school student	Human - Animal	-0.251	-25.76	732	0.000	0.36
	Human - Plant	-0.110	-7.49	732	0.000	0.14
	Animal - Plant	0.141	20.14	732	0.000	0.16
Non-biology science teachers	Human - Animal	-0.258	-3.70	34	0.001	0.24
	Human - Plant	-0.094	-0.92	34	0.366	0.08
	Animal - Plant	0.163	4.02	34	0.000	0.12
Biology teacher	Human - Animal	-0.888	-9.21	22	0.000	0.71
	Human - Plant	-0.918	-6.89	22	0.000	0.68
	Animal - Plant	-0.030	-0.74	22	0.469	0.02

의 크기, 인간의 두뇌의 변화 등과 같은 다양한 상황에 적용하면 상황 의존적으로 개념이 형성되지 않을 수 있을 뿐만 아니라 다양한 상황에 지식을 전이시킬 수 있는 능력도 형성될 수 있다. 또한 평가에서는 다양한 설명소재가 활용되어야 보다 편향되지 않은 평가결과를 취할 수 있을 것이다(Nehm & Ha, 2011).

V. 결론

이 연구는 과학사에서 나타나는 진화개념의 발달과정과 학생들의 진화개념의 이해가 유사하다는 점을 근거로 개발된 개념평가도구를 사고의 발달이라는 관점에서 창조론, 목적, 의도, 용불용설, 자연선택의 단계별로 부분점수를 주는 순위선다형방법의 문항타당도와 개념의 발달과정을 확인하기 위해서 진행되었다. 구체적인 연구문제는 생물 형질에 대한 사고의 과학사적 발달 과정(창조론적 사고, 목적론적 사고, 의도적 사고, 용불용설적 사고, 자연선택적 사고)를 토대로 한 순위선다방식을 활용하여 부분점수모형을 사용하였을 때의 문항 반응 타당도의 확인과 인간, 동물, 식물 문항의 반응의 3차원 라쉬분석을 통한 문항의 맥락에 따른 개념 발달은 어떻게 진행되는지를 확인하였다.

문항의 내적일관성 신뢰도, 라쉬분석의 문항의 신뢰도, MNSQ값을 통한 문항반응적합도를 확인한 결과 과학사적 발달 과정을 토대로 순위선다방식을 적용한 분석 방법은 적합한 것으로 확인되었다. 또한 인간, 동물, 식물 문항으로 구분하여 실시한 3차원 라쉬분석은 문항의 맥락에 따라 개념발달이 차별적이라는 것도 보여주었다. 학생들의 진화개념발달은 동물문항이 다소 높았으며, 인간문항에서는 자연선택개념까지 진행되지 않는 것을 확인하였다. 과학사에 근거하여 창조에서 자연선택까지 단계별로 점수를 주는 부분점수모형으로 평가하는 방법은 학생들의 개념의 수준을 단계별로 구분할 수 있는데 유리할 것이다. 학생들의 개념 발달 수준의 단계를 확인한다면 학생들의 수준에 맞는 교수법 찾아서 제공할 수 있을 것이다. 맥락에 따라 차별적인 개념발달 역시 교육적 함의를 가진다. 평가에서 다양한 소재가 활용되어야 보다 편향되지 않은 평가결과를 취할 수 있고, 학생들 역시 다양한 소재를 활용하여 진화 개념을 학습한다면 개념이 상황 의존적으로 형성되는 것을 방지할 수 있을 것이다.

국문요약

학생들의 진화에 대한 대안개념(목적론, 용불용설 등)이 과학사에서 나타나는 설명의 발전 형태와 유사하다는 연구는 있었다. 하지만 과학사적으로 설명의 발달과정을 반영하여 부분점수를 주는 평가방식은 활용되지 않았다. 이 연구의 목적은 창조론에서 자연선택까지 과학사적 발달과정을 반영하여 부분점수를 주는 방법을 제안하고 이 방법이 타당한지에 대한 양적인 증거를 수집하는 것이다. 이 연구는 과학사에 근거하여 진화개념검사도구의 학생응답을 순위선다형점수로 변환하고 부분채점모형의 라쉬모델분석을 포함한 통계적 방법으로 새로운 평가방식이 타당한지 확인하였다. 또한 개념발달이 인간, 동물, 문항의 상황에 따라 다른지 확인하였다. Ha(2007)가 개발한 검사도구를 활용하여 1711명의 초, 중, 고등학생과 비전공, 전공생물 교사를 대상으로 생성한 자료를 통하여 분석하였다. 창조론, 목적,

의도, 용불용설, 자연선택에 0점에서 4점씩 부분점수로 제시한 평가 방법은 Cronbach alpha를 통한 내적일관성 신뢰도, 라쉬분석의 MNSQ값 등 통한 문항적합도를 확인한 결과 타당한 것으로 확인되었다. 초등학생과 중학생들의 개념수준은 의도에서 용불용설 단계에, 고등학생부터 용불용설 이후의 단계로 개념발달이 이루어지고 있었다. 진화설명 발달 과정은 인간, 동물, 식물에 따라 차이가 나타남을 확인할 수 있었다. 이 연구는 과학사와 학생들의 개념발달이 유사하다는 기존의 주장에 새로운 양적증거를 추가하고, 진화개념 평가를 위한 새로운 분석방법을 제안한다.

주제어 : 진화, 과학사, 순위선다형평가, 라쉬분석, 개념 발달

References

- Anderson, D. L., Fisher, K. M., & Norman, G. J. (2002). Development and evaluation of the conceptual inventory of natural selection. *Journal of Research in Science Teaching*, 39(10), 952-978.
- Bowler, P. J. (2009). *Evolution: The history of an idea*. Berkeley, CA: University of California Press.
- Churchill, F. B. (1968). August Weismann and a break from tradition. *Journal of the History of Biology*, 1(1), 91-112.
- Clement, J. (1983). A conceptual model discussed by Galileo and used intuitively by physics students. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 325-339). Hillsdale, NJ: Erlbaum.
- Fauth, B., Decristan, J., Rieser, S., Klieme, E., & Büttner, G. (2014). Student ratings of teaching quality in primary school: Dimensions and prediction of student outcomes. *Learning and Instruction*, 29, 1-9.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference*. 11.0 update (4th ed.). Boston: Allyn & Bacon.
- Grant-Downton, R., & Dickinson, H. G. (2005). Epigenetics and its implications for plant biology. 1. The epigenetic network in plants. *Annals of botany*, 96(7), 1143-1164.
- Greene, J. C. (1992). From Aristotle to Darwin: Reflections on Ernst Mayr's interpretation in the growth of biological thought. *Journal of the History of Biology*, 25(2), 257-284.
- Ha, M. (2007). Development of the instructional strategies of evolution based on the cross-sectional analysis of evolution conception. Unpublished master thesis. Cheong-ju: Korean National University of Education.
- Ha, M. Ch., Ha, Y., & Ku, S. (2012). A comparative study of Korean and United States college students' degree of religiosity, evolutionary interest, understanding and acceptance and their structures. *Journal of the Korean Association for Research in Science Education*, 32(10), 1537-1550.
- Ha, M., & Nehm, R. H. (2014). Darwin's difficulties and students' struggles with trait loss: cognitive-historical parallelisms in evolutionary explanation. *Science & Education*, 23(5), 1051-1074.
- Ha, M., Ku, S., & Cha, H. Y. (2010). Study of conceptions, interest and acceptance of evolution, and religiosity between biology majors and non-majors in colleges. *Biology Education*, 38(3), 467-475.
- Ha, M., Lee, J. K., & Cha, H. Y. (2006). A cross-sectional study of students' conceptions on evolution and characteristics of conception formation about it in terms of the subjects: Human, animals and plants. *Journal of Korean Association for Research in Science Education*, 26(7), 813-825.
- Hair, Jr., J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (6th ed.). Upper Saddle River, NJ: Pearson Prentice Hall.
- Johnson, R. L., & Peeples, E. E. (1987). The role of scientific understanding in college: student acceptance of evolution. *The American Biology Teacher*, 49(2), 93-98.
- Jungwirth, E. (1975). The problem of teleology in biology as a problem of biology-teacher education. *Journal of Biological Education*, 9(6), 243-246.
- Kampourakis, K., & Zogza, V. (2007). Students' preconceptions about evolution: How accurate is the characterization as "Lamarckian" when considering the history of evolutionary thought? *Science & Education*, 16(3-5), 393-422.
- Kelemen, D. (1999a). The scope of teleological thinking in preschool children. *Cognition*, 70(3), 241-272.
- Kelemen, D. (1999b). Function, goals and intention: children's teleological

- reasoning about objects. *Trends in Cognitive Sciences*, 3(12), 461-468.
- Kelemen, D., Rottman, J., & Seston, R. (2013). Professional physical scientists display tenacious teleological tendencies: Purpose-based reasoning as a cognitive default. *Journal of Experimental Psychology: General*, 142(4), 1074.
- Lee, M. S., & Lee, K. J. (2006). Analysis of student conceptions in evolution based on science history. *Journal of Korean Association for Research in Science Education*, 26(1), 25-39.
- Nadelson, L. S., & Southerland, S. A. (2009). Development and preliminary evaluation of the measure of understanding of macroevolution: Introducing the MUM. *The Journal of Experimental Education*, 78(2), 151-190.
- Nehm, R. H., & Ha, M. (2011). Item feature effects in evolution assessment. *Journal of Research in Science Teaching*, 48(3), 237-256.
- Nersessian, N. (1989). Conceptual change in science and in science education. *Synthese*, 80, 163-183.
- Neumann, I., Neumann, K., & Nehm, R. (2011). Evaluating instrument quality in science education: Rasch-based analyses of a nature of science test. *International Journal of Science Education*, 33(10), 1373-1405.
- Opfer, J. E., Nehm, R. H., & Ha, M. (2012). Cognitive foundations for science assessment design: Knowing what students know about evolution. *Journal of Research in Science Teaching*, 49(6), 744-777.
- Osler, M. J. (2001). Whose ends? Teleology in early modern natural philosophy. *Osiris*, 16, 151-168.
- Seoh, K. H. R., Subramaniam, R., & Hoh, Y. K. (2016). How humans evolved according to grade 12 students in Singapore. *Journal of Research in Science Teaching*, 53(2), 291-323.
- Van Driel, J. H., De Vos, W., & Verloop, N. (1998). Relating students' reasoning to the history of science: The case of chemical equilibrium. *Research in Science Education*, 28(2), 187-198.
- Vosniadou, S. & Brewer, W. F. (1987). Theories of knowledge restructuring in development. *Review of Educational Research*, 51, 51-67.
- Wandersee, J. H. (1985). Can the history of science help science educators anticipate student's misconceptions. *Journal of Research in Science Teaching*, 23, 581-597.
- Willard, A. K., & Norenzayan, A. (2013). Cognitive biases explain religious belief, paranormal belief, and belief in life's purpose. *Cognition*, 129(2), 379-391.
- Wiser, M. & Carey, S. (1983). When heat and temperature were one. In D. Gentner & A. L. Stevens (Eds.), *Mental models* (pp. 267-297). Hillsdale, NJ: Erlbaum.
- Wright, B.D. & Linacre, J.M. (1994). Reasonable mean-square-fit values. *Rasch Measurement Transactions*, 8(3), 370.