

과학탐구 상황에서 고등학생들의 반회의주의적 태도 측정도구 개발 및 적용

아리프 라흐마툴라흐 · 하민수*

노스캐롤라이나 주립 대학교 · 강원대학교

Developing and Applying the Questionnaire to Measure High School Students' Unskeptical Attitude in Science Inquiry

Arif Rachmatullah · Minsu Ha*

North Carolina State University · Kangwon National University

Abstract : The purpose of the study is to develop a questionnaire that examines unskeptical attitudes in scientific inquiry context. The questionnaire items were developed through literature research, expert review, and statistical analyses for validity and the differences in scores were identified by gender and tracks. A total of 363 high school students participated in the study. To explore the validity evidence of items, the Rasch analysis and the reliability of internal consistency were performed, and the two-way ANOVA was performed to compare the scores of the unskeptical attitudes between gender and academic track. Self-reporting and Likert-scaling 23 items were developed to measure unskeptical attitudes in scientific inquiry context. The items were developed in the sub-domain of scientific inquiry: 'questioning and hypothesis generating,' 'experiment designing,' and 'explaining and interpreting.' Second, the validity and reliability of the unskeptical were identified in a rigorous method. The validity of items were identified by multi-dimensional partial score model analysis through the Rasch model, and all 23 items were found to be fit to model. Various reliability evidences were also found to be appropriate. It was found that there were no significant differences of unskeptical attitude score between the gender and academic track except one comparison. The developed questionnaire could be used to check an unskeptical attitude in the course of scientific inquiry and to compare the effects of scientific inquiry classes.

keywords : scientific skepticism, scientific inquiry, questionnaire development, science education

I. 서론

과학교육의 중심 목표는 학생들이 자연현상을 탐구하여 새로운 지식과 합리적 의사결정을 할 수 있는 역량을 향상시키는 것이다(MOE, 2015). 과학은

자연이 어떤 방식으로 작동하는지, 그 과정과 현상에 대한 원인을 탐구하는 학문이다(Carey *et al.*, 1989; Driver *et al.*, 1994). 과학적 활동에서는 감각기관을 통해 수집된 다양한 관찰 사실들을 종합하여 관찰하지 않은 새로운 지식을 추론하는 것도 포함된다. 이 과정에서 자신의 추론에는 합당한 증

* : (msha@kangwon.ac.kr)

** 2016

(NRF -2016S1A5A8017660)

***2018 09 27 , 2018 11 29

, 2018 11 29

<http://dx.doi.org/10.21796/jse.2018.42.3.308>

거들이 논리적으로 연계되어야 하며, 이 과정에서 비판적 사고가 반드시 필요하다. Lipman(1987)은 비판적 사고를 정의하면서, 주장들이 어떤 이유를 근거로 정당화 되는지를 판단하고 문제를 해결하는 사고라 하였다. 그렇기 때문에 비판적 사고는 어떤 이론이나 주장을 무조건 수용하는 것이 아니라 이론적 근거에 대해 의심하고 근거가 적절한지 판단하는 등의 합리적인 사고를 하는 것이다. 비판적 사고를 통한 창의적 인재 양성을 주된 교육 목표로 설정한 이후부터 새로운 지식 생성이라는 사회적 요구에 부응할 수 있었다(Abrami *et al.*, 2015). Beyer(1995)은 비판적 사고를 통해 지식을 형성하는 과정을 오래 수행하면 지식의 옳고 그름에 대한 반복적인 판단을 수행할 수 있어 궁극적으로 합리적 의사 결정 능력도 향상된다고 한다.

비판적 사고에 대해서 Glaser(1941)는 3가지 요소로 정리하였다. 첫째, 자신의 경험 범위 내에서 발생하는 문제에 대해서 신중히 고려하는 태도이며, 둘째 문제를 논리적으로 조사하고 추론하는 방법적 지식이다. 마지막으로 실제로 방법을 적용할 수 있는 기술이 요구된다. 이 중에서 신중히 자신의 문제를 고려하는 태도는 회의주의와 밀접하다. 자신의 경험, 생각, 지식들에 대해 대부분의 사람들은 비판적이지 않고 옳다고 믿지만 실제로는 틀렸을 수 있다. 잘못된 지식을 활용하여 수행한 의사결정은 큰 문제를 양산하기도 한다(McPeck, 1981). 따라서 경험 내에서 일어난 일을 그대로 받아들이지 않고 근거에 대해 의심하는 태도를 가지는 회의주의가 필요한데 과학과 관련된 것을 과학적 회의주의라 한다. 회의주의적 관점은 인간의 이성의 불완전성을 인정하고, 완벽하지 않다고 믿음으로서 인간의 이성을 더 편안하게 해준다. 또한 자신이 경험 지식들의 옳고 그름에 대한 추가적인 조사를 수행할 수 있으므로 더 좋은 의사 결정을 수행할 수 있는 선택지를 제공한다. 회의주의자는 관찰 사실에 대하여 신중하고 진지하게 탐구하는 사람을 의미한다(Popkin, 1967). 과학적 회의주의자는 경험적 지식의 진위에 대해서 한 번 더 고민하는 방법을 통해 더 확실한 지식을 생성할 수 있게 해준다. 또한 지식이 넘쳐나는 현대 사회에서

올바른 지식의 선택과 선별에서 회의주의는 중요한 기능을 수행할 수 있다.

현대 사회의 높은 수준의 과학적 발전으로 인하여 많은 지식들이 생성되고 있다. 특히 과학 지식이 국가적 발전이나 개인의 이득과도 밀접하게 관련되어 있기 때문에 이해중립성에 어긋나는 과학적 지식도 끊임없이 양산되고 있다(Macfarlane & Cheng, 2008). 따라서 경험적 지식을 의심하고 한 번 더 생각하도록 해주는 회의주의는 그런 점에서 중요한 역량이라 할 수 있다. 미래 사회의 비판적, 합리적, 창의적 인재를 교육시키기 위해서는 과학을 회의주의에 대한 교육이 반드시 필요할 것이다. 특히 회의주의적 태도는 인지적 발달과 함께 성장하기 보다는 사회적인 경험에 의해 생성되기 때문에 회의주의 교육은 필요하다(Heyman *et al.*, 2007).

과학적 회의주의 교육을 위한 과정으로 본 연구에서는 과학적 반회의주의적 태도 검사도구를 개발하고자 한다. 교육의 시작과 끝은 진단평가와 형성평가이다. 그러므로 과학적 회의주의 교육을 위해 반드시 필요한 것은 측정 도구이다. 우리의 일상적인 반회의주의적 태도의 변화를 지속적으로 살펴야 한다. 또한 검사도구 개발과정에서 과학적 회의주의에 대한 문헌 고찰을 통하여 구인을 구체화함으로써 과학적 회의주의의 교육목표도 확인할 수 있다. 반회의주의 태도 문항을 바탕으로 어떻게 하면 회의주의적 태도를 향상시킬 수 있을지에 대한 교수 전략도 함께 논의할 수 있다. 따라서 이 연구에서는 그 동안 개발된 적이 없는 과학탐구 상황에서 반회의주의 문항을 다양한 문헌 고찰과 통계적 방법을 활용하여 개발하는 것을 1차적 목표로 한다. 또한 개발된 도구를 통하여 고등학생들을 대상으로 조사하고 성별과 계열별로 그 차이를 확인하는 것이 목표이다. 자세한 연구 목적은 다음과 같다.

1. 과학탐구상황에서 반회의주의 태도의 세부 구인과 구인을 측정할 문항을 개발한다.
2. 과학탐구상황에서 반회의주의 태도 측정 문항의 타당도와 신뢰도를 통계적 방법으로 확인한다.
3. 고등학생들을 대상으로 성별과 계열별에 따라 과학탐구상황에서 반회의주의 태도에 유의미한 차이가 있는지 확인한다.

II. 연구 방법

1. 문항 개발 및 타당화

이 연구에서 문항의 개발은 문헌 연구를 시작으로 전문가 검토, 1차 문항 선별 및 통계 자료 수집, 비적합 문항 제거, 재분석의 과정을 거치면서 진행되었다. 문헌 연구는 과학탐구의 본성, 과학적 회의주의, 과학 지식의 생성의 키워드로 광범위한 조사를 수행하였다. 문헌 연구를 통해 과학탐구 상황을 크게 ‘문제 인식과 가설 생성’, ‘실험 설계’, ‘관찰’, ‘설명과 해석’으로 단계를 구분하고 각 단계에 우리가 일상적으로 하고 있는 반회의주의적 태도를 정리하였다. 1차적으로 정리된 자료들을 바탕으로 3인의 과학교육 전문가와 2인의 과학 교사가 추출된 문구를 바탕으로 문항화 하였다.

개발된 문항의 타당도는 Messick(1995)의 6가지 구인 타당도의 기준 중에서 내용 타당도, 구조 타당도, 실제에 기초한 타당도, 일반화 타당도를 중심으로 확인하였다. 과학적 회의주의 문항에 대한 첫 개발이므로 이전에 개발되어 발표된 유사한 문항이 없기 때문에 외적 준거에 기초한 타당도는 확인하지 않았다. 또한 자기 보고형 태도 검사도구이기 때문에 공정성 등의 결과에 기초한 타당도는 해당되지 않는 것으로 판단되어 이 연구에서 조사하지 않았다. 내용 타당도는 문헌 분석과 전문가 검토를 통해 확인하였으며, 구조 타당도는 차원 분석을 통해 확인하였다. 실제에 기초한 타당도는 문항 적합도 등의 라쉬 분석 결과를 활용하였으며, 일반화 타당도는 차별적 문항기능 분석을 통해 확인하였다. 자세한 방법은 연구 결과에 기술되어 있다.

2. 참여자

이 연구를 위하여 3개 학교의 363명의 고등학생들이 참여하였다. 참여자 중 1학년은 34.4%, 2학년은 65.6%이었다. 남학생은 56.5%, 여학생은 43.5%이었다. 문과계열은 46.2%, 이과계열은

53.8%로 나타났다. 2017년 10월 경 자료가 수집되어 1학년의 경우 계열 선택이 이루어지고 있는 시점이었다. 따라서 1학년 학생은 희망하는 계열을 묻는 것으로 조사하였다.

3. 분석 방법

과학탐구 상황에서 반회의주의 태도에 관한 측정 문항의 통계적 타당도와 신뢰도를 확인하기 위하여 다양한 방법을 활용하였다. 먼저 4차원 부분점수 모형 라쉬 분석(four dimensional-partial credit model-Rasch analysis)를 사용하였다. 이 분석을 위하여 라쉬 분석 도구인 ConQuest 프로그램을 사용하였다. 먼저 개발된 문항이 1차원으로 분석하는 것과 4차원으로 세부 요인별로 분석하는 것 중 어떤 것이 더 높은 설명력과 타당함을 보이는지 비교하였다. Final Deviance와 Akaike Information Criterion(AIC)의 값을 비교하였다(Neumann *et al.*, 2011). 학생들의 응답 분포가 문항의 타당도를 확인하는데 적합한 수준으로 분포하는지 확인하기 위하여 Item Separation Reliability를 확인하였고, 각 요인별 세부 문항들이 학생들의 반회의주의적 태도를 구분할 수 있는지 확인하기 위해 EAP/PV Reliability를 확인하였다.

라쉬 분석을 통하여 확인된 문항 적합도(MNSQ, mean-square)를 통해 학생들의 반응의 적합도를 확인하였다. MNSQ가 1을 기준으로 낮거나 높아질 경우 문항적합도가 낮다고 해석할 수 있다. 본 문항은 평정척도이면서 저부담 평가이므로 Wright & Linacre(1994)가 제시한 0.6-1.4가 활용될 수 있으며, 이 보다 더 관대한 기준인 0.5-1.5 역시 기준이 될 수 있다(Boone *et al.*, 2014). 자료의 내적 일관성 신뢰도를 확인하기 위하여 Cronbach alpha를 확인하였다. 남학생과 여학생, 계열별 응답 반응의 차이를 확인하여 문항이 집단별로 동등하게 작동했는지 확인하기 위한 차별적 문항반응도 분석하였다. 요인별 상관관계는 Pearson correlation 분석으로, 집단간 반회의주의 태도의 점수 비교는 이원분산분석을 통해 확인하였다. 이원분산분석의

효과크기의 확인은 partial eta squared(PES)를 사용하였다. 라쉬 분석을 제외한 기타 통계 분석은 모두 SPSS 22버전을 사용하였다.

Ⅲ. 연구 결과 및 논의

1. 과학적 회의주의에 대한 문헌 연구 및 세부 요인 설정

회의주의는 지식, 이론 등에 대한 자기 반성적인 의문을 제시하고, 잘못되었을 수 있음에 대한 여지에 대한 다양한 가능성을 제시한다. 이러한 지속적인 의문 제기는 실제 행동에 영향을 줄 수 있다(Korsgaard, 1986). 과학탐구 상황에서의 회의주의도 마찬가지이다. 과학적 방법은 자연 현상에 대해 연구하고 새로운 지식을 구축해 나가는데 사용되는 방법이다. 경험과 측정에 근거한 증거를 과학적 주장을 위하여 사용하여 자연 현상의 원리를 밝혀내는 과정이다. 회의주의는 실증 가능성과 반증 가능성을 염두에 두어 주장을 평가하며, 직관적인 믿음과 일회적인 증거에 의해 제시되는 주장을 수용하지 않는다. 과학실험에서도 비판적으로 접근해야 되며, 탐구한 결과들이 정당성 있게 수용되기 위해서는 의심할 여지가 없는 합리적이고 타당한 증거가 필요하다(Parascandola, 2004).

학교에서 이루어지는 탐구 상황을 고려하였을 때, 학생들은 교사의 수업과 교과서의 내용을 그대로 믿지 않고 의심하는 태도를 지니며, 주장에 합리적인 근거를 찾아내는 능력을 갖추어야 한다(Hofer & Pintrich, 1997). 이 과정에서 무조건적인 의심이 강조되는 것이 아니다. 과학의 한계에 대해 이해하면서, 여러 조건 상황에서 실제로 탐구 조사가 가능한 것인지, 문헌을 통해 정당화를 해야 되는 것인지, 추가적인 실험이 가능한지 지속적으로 생각하는 태도가 교실에서의 과학적 회의주의이다(Chinn & Malhotra, 2002). 계속된 의문 제기 와 근거를 찾아내면서 학생들은 자신의 의견을 만

들어 낼 수 있을 것이고, 탐구 능력의 향상뿐만 아니라 다양한 개념 학습까지 효과적으로 수행할 수 있다(Aikenhead, 2006).

이와 같이 회의주의는 과학을 배우는데 있어서 굉장히 중요한 역할을 한다. 교실 내에서 일어나는 탐구 상황을 고려하면서 중요한 회의주의적 태도를 살펴보았다. 가장 먼저 연구의 시작인 문제 인식과 가설 생성이다. 우리는 자연현상을 보면서 다양한 의문을 제기하나 많은 문제들은 직관적이고 의미가 없을 수 있다. 또한 제시되는 의문에 성급하게 가설을 설정하는 것도 지양해야 된다. 따라서 두 번 생각하고 보다 신중하게 접근하는 태도, 다양한 문제와 가설에 대한 염두는 중요하다(Kahneman, 2013). 또한 가설은 인과적 설명이기 때문에 인과성에 대하여 면밀히 검토해야 된다. 두 사건이 동시에 일어난다고 하여 인과적 설명이라고 할 수 없다. 논리적으로 인과성을 갖추기 위해서는 다양한 검토가 필요하다(Sloman *et al.*, 2009). 정리하면 문제 인식과 가설 생성에서는 두 번 생각하고 보다 신중하게 접근하는 태도, 다양한 문제와 가설에 대한 염두, 인과성에 대한 면밀한 검토, 직관에 의한 빠른 결론 생성의 경계가 요구된다.

실험 설계에서도 다양한 요소들이 요구된다. 관찰 하려는 대상과 주변은 서로 복잡한 상호작용을 하고 있기 때문에 한 가지 방법만 사용하는 것이 아니라 여러 가지 방법들을 사용하여 관찰해야 한다(Chinn & Malhotra, 2002). 가능한 한 주변 환경에서 발견할 수 있는 모든 데이터를 수집하는 것이 좋으며, 그런 태도를 유지해야 된다. 연구 주제에 관련된 자료 수집은 기본이며, 그 이외에 생성되는 자료들에 대한 추가 수집도 염두를 해야 된다. 실험을 다시 수행하는 것은 어려운 일이기 때문에 사전에 충분한 논의를 거쳐서 수집 가능한 전체 자료를 확보해야 다양한 관점에서 설명을 생성할 수 있다. 정리를 하면 실험에 영향을 주는 보이지 않는 요인 등 많은 요인에 대한 고려, 복잡한 방법에 대한 탐색, 다양한 탐색에 대한 동기가 실험 설계에서 요구된다.

실험 과정에서 관찰을 하는 것에도 많은 회의주의가 요구된다. 특히 관찰은 과학의 본성에서도 관

찰의 한계에 대해 명시적으로 제시한다. 자연 세계는 서로 상호작용을 하고 있기 때문에 특정 요인을 중심으로 관찰하기 보다는 여러 통제 변수를 설정하는 것이 더 유의미한 설명을 생성할 수 있다(Chinn & Malhotra, 2002). 특히 관찰자는 자신의 이론에 의하여 제한된 관찰을 수행할 수 있다. 자신이 원하는 답에 맞는 증거를 수집할 가능성이 높으며, 관심 사항에 따라 관찰을 수행할 수 있다(Nickerson, 1998). 정리하면 다양한 관찰, 확장편향에 대한 방지, 관심에 대한 통제가 관찰 과정에서 주의를 기울여야 하는 요소이다.

마지막으로 설명과 해석 과정에서도 회의주의적 태도가 필요하다. 자신의 결론을 지지하는 데이터를 얻었다라도 한 번에 만족해서는 안 되고 많은 자료를 얻기 위해서 노력해야 한다. 자료의 신뢰도와 타당도를 확보하기 위해서는 많은 조사가 필요하다(Trout, 2002). 조사를 마친 후에는 연구 문제를 충분히 설명할 수 있는지 여러 번 검토해야 한다. 또한 수집한 데이터를 해석 할 때 회의적이어야 한다. 원하는 내용과 반대되는 내용에 대한 것 까지 조사하여 확인해야 한다(Gilovich, 2008). 데이터를 한 가지 측면으로만 해석 하는 것이 아니라 여러 방면으로 해석하여 상황에 맞게 사용해야 한다. 특히 원하는 결과가 나타났다고 과정이 옳았을 것으로 해석하면 안 된다(Gilovich, 2008). 또한 기존의 이론이나 교사의 설명, 교과서의 설명에 끼워 맞추는 식의 해석 등 권위적인 지식에 결과 해석을 맞추는 것도 주의해야 될 태도이다(Gilovich, 2008). 정리하면 다양한 해석의 가능성, 성급한 일반화에 대한 경계, 동기 기반 추론 경계, 결과 편향 경계, 권위에 대한 호소 경계가 실험과 해석 과정에서 요구되는 회의주의적 태도이다.

Table 1은 이상의 문헌 조사를 통해 확인된 회의주의적 태도와 반대되는 반회의주의 태도에 대한 문항이다. 이 연구에서 학생들의 과학적 회의주의에 대한 인식이 아닌 반회의주의적 태도를 측정하는 도구를 개발하고자 하는 것은 회의주의는 하나의 지향점이며, 반회의주의는 우리가 현재 가지고 있는 심리적, 행동적 요소들이다. 추상적인 지향점에 대한 인식이 아닌 현재 우리가 하고 있는 행동

이 보다 구체적이기 때문에 측정이 가능하다. 인지 편향과 인간의 비합리성에 관한 다양한 연구는 자연스러운 우리의 모습은 반회의주의적 태도를 가진 인간임을 강조한다(Kahneman, 2013). 따라서 이 연구에서는 합리적인 과학탐구를 수행하는데 필요한 회의주의적 태도와 상반되는 반회의주의적 태도에 대한 측정 도구를 개발하였다.

2. 개발된 문항의 통계적 타당도 및 신뢰도

개발된 과학탐구 상황에서 반회의주의적 태도 측정 문항은 과학탐구 상황을 ‘문제 인식과 가설 생성’, ‘실험 설계’, ‘관찰’, ‘설명과 해석’으로 구분하였다. 개발된 23개 문항을 하나의 요인으로 설정하고 분석하는 것과 4개 요인으로 구분하여 분석하는 것 중에서 어느 것이 더 타당한지 확인하기 위하여 각 차원으로 분리하여 적합도를 확인하였다. 또한 리커트 척도를 활용한 연구이기 때문에 각 척도의 점수가 문항마다 다를 수 있다는 것을 전제로 한 Partial Credit Model과 모든 문항이 동일한 척도 수준을 가진다는 Rating Scale Model에서 어느 것이 더 적합한지도 같이 확인하였다. Table 2에는 1차원과 4차원, Partial Credit Model과 Rating Scale Model로 구분된 총 4가지 경우의 Final Deviance와 Akaike Information Criterion 점수를 비교하였다. 두 분석 모두 가장 낮은 Final Deviance와 Akaike Information Criterion 값을 보이는 것이 좋은 모델로 판단한다(Neumann *et al.*, 2011). 그 결과 4개 요인으로 구분하여 분석하고, Partial Credit Model을 사용하는 것이 가장 적합한 것으로 확인하였다.

두 번째로 실제에 기초한 타당도를 확인하였다. 라쉬 분석은 라쉬 모델을 바탕으로 예측된 점수와 실제 자료의 차이점을 활용하여 문항의 적합도를 확인한다. 그 값은 mean-square(MNSQ)로 나타내어진다. 적합한 값의 기준은 Wright & Linacre(1994)가 제시한 0.6-1.4를 활용하거나, Boone *et al.* (2014)가 제시된 0.5~1.5 사이의 값을 기준으로 활용한다. 최종 23문항을 분석한 결과 Infit MNSQ의 경우 0.84~1.14, Outfit MNSQ의

Table 1. Developed items

가	1			.
	2			.
	3	가		가
	4			.
	5			.
가	1			.
	2			.
	3			.
	4			가
	5			.
	6			/ ,
가	1			.
	2	가		.
	3	가		.
	4			.
가	1			.
	2			.
	3			.
	4			가
	5	가		.
	6			가
	7			.
	8			.

Table 2. Analysis of appropriate dimensions and model for Rasch analysis

	Partial Credit Model		Rating Scale Model	
	Final Deviance	AIC	Final Deviance	AIC
1	18111	18297	18293	18347
4	17893	18098	18130	18203

경우에도 0.84~1.14로 적합한 것으로 확인되었다 (Bond & Fox, 2001). EAP/PV Reliability는 문항이 학생들의 반회의주의적 태도를 효율적으로 구분해 내는가 확인하는 신뢰도이며, EAP/PV Reliability의 경우에는 0.8이상일 경우 아주 좋은 신뢰도, 0.7~0.8사이의 경우는 바람직한 결과 0.60~0.69사이의 경우는 받아들여 질 수 있는 정도로 구분된다(DeVellis, 2003). 23문항에 대한 결과는 문제 인식과 가설 생성의 값은 0.765, 실험 설계의 값은 0.835, 관찰의 값은 0.659 설명과 해석의 값은 0.765로 가장 낮은 신뢰도를 보인 관찰의 경우에도 수용될 수 있는 수준의 결과를 확인할 수 있었다. 참여자들의 수준이 문항의 신뢰도를 확인하는데 적합한지를 확인해주는 Separation Reliability 신뢰도의 경우에도 0.97로 1에 매우 근접한 유의미한 결과를 확인 했다. 전체적으로 라쉬

분석 결과는 타당한 것으로 확인되었다.

수집한 자료의 내적 일관성 신뢰도를 확인하기 위하여 Cronbach α 계수를 산출했다. 23개의 문항의 전체 문항 신뢰도와 각 요인별 신뢰도, 문항 삭제 후 신뢰도는 Table 3에 제시되어 있다. 전체 문항의 Cronbach α 신뢰도는 0.890로 높은 수준으로 확인되었다. 분류별 신뢰도에서 문제 인식과 가설 설정의 2번 문항을 제거하였을 때만 0.731에서 0.733으로 소폭 증가하는 것을 확인했다. 전체적으로 0.7이상의 적합한 수준의 내적 일관성 신뢰도를 보여주고 있다(Fauth *et al.*, 2014).

다음으로 일반화에 기초한 타당도이다(Table 4). 일반화에 기초한 타당도는 검사 도구가 응답자의 특성에 영향을 받지 않고 동등하게 작용하고 있는지 보여주는 타당도이다. 특정 문항이 특정 집단에 다르게 이해되거나 다른 반응을 유도할 경우 전체

Table 3. Rasch analysis results of developed unskeptical attitude items

	MEASURE	Infit MNSQ	Outfit MNSQ	EAP/PV Reliability		
가	1	-0.11	1.12	1.12	0.765	0.733
	2	0.48	0.98	0.98		0.712
	3	-0.20	0.85	0.85		0.629
	4	-0.17	0.84	0.84		0.637
	5	0.01	0.93	0.93		0.700
	1	0.40	1.05	1.04	0.835	0.773
	2	-0.90	1.14	1.14		0.786
	3	0.15	0.95	0.95		0.764
	4	-0.07	1.04	1.04		0.781
	5	0.16	0.90	0.90		0.749
	6	0.26	0.92	0.93		0.752
	1	-0.16	1.06	1.03	0.659	0.671
	2	-0.22	1.03	1.02		0.644
	3	0.29	0.99	0.98		0.645
	4	0.09	1.07	1.06		0.718
	1	-0.31	0.96	0.96	0.765	0.752
	2	-0.16	0.93	0.93		0.744
	3	0.31	0.98	0.98		0.750
	4	0.62	1.08	1.08		0.773
	5	0.07	0.90	0.90		0.746
	6	0.18	0.96	0.95		0.753
	7	-0.47	0.99	0.99		0.753
	8	-0.24	0.96	0.96		0.746

점수의 결과 해석이 왜곡될 수 있다. 본 연구에서는 성별과 계열에 따라 개발된 문항이 동등한 효과를 보이는지 라쉬 분석의 차별적 문항 분석 기능인 DIF 차이값을 사용하여 확인했다. 집단간 DIF값의 차이가 0.64 보다 큰 경우 그룹별로 문항에 대한 반응이 다르게 나타나는 것으로 볼 수 있다(Boone *et al.*, 2014). 23개의 문항을 분석한 결과 성별의

경우 문제 인식과 가설 설정의 2번 문항의 DIF값의 차이가 0.84, 계열별에서는 문제 인식과 가설 설정의 5번 문항이 0.66으로 0.64 기준을 벗어났다. 그 외에 모든 문항은 이 기준을 만족하는 것으로 확인되었다.

각 요인별로 생성된 점수의 상관관계는 Table 5에 제시되어 있다. 문제 인식과 가설 생성, 실험

Table 4. Results of differential item functioning analysis

	1	0.03	-0.43	0.46	-0.33	0.04	-0.37
	2	0.42	1.27	-0.84	0.95	0.65	0.30
가	3	-0.21	-0.47	0.27	-0.23	-0.43	0.20
	4	-0.48	-0.01	-0.47	-0.04	-0.48	0.44
	5	0.24	-0.37	0.61	-0.40	0.26	-0.66
	1	0.47	0.47	0.00	0.80	0.24	0.56
	2	-0.89	-1.52	0.63	-1.42	-0.96	-0.47
	3	0.27	0.38	-0.12	0.37	0.29	0.08
	4	-0.12	-0.32	0.20	-0.56	0.07	-0.63
	5	0.07	0.50	-0.43	0.39	0.14	0.26
	6	0.18	0.49	-0.32	0.43	0.21	0.23
	1	-0.16	-0.13	-0.03	-0.32	-0.05	-0.26
	2	-0.19	-0.45	0.26	-0.30	-0.30	0.00
	3	0.30	0.51	-0.21	0.47	0.36	0.11
	4	0.07	0.07	0.00	0.14	0.00	0.15
	1	-0.36	-0.70	0.34	-0.62	-0.42	-0.20
	2	-0.17	-0.46	0.29	-0.11	-0.39	0.28
	3	0.25	0.54	-0.29	0.31	0.42	-0.11
	4	0.78	1.35	-0.56	1.21	0.85	0.36
	5	0.06	-0.11	0.17	0.05	-0.10	0.15
	6	0.19	0.62	-0.42	0.52	0.34	0.18
	7	-0.47	-0.85	0.38	-0.86	-0.45	-0.40
	8	-0.28	-0.39	0.11	-0.51	-0.24	-0.27

Table 5. Correlations of scores from four factors

가					
가	1.000	-	-	-	-
	0.951 [‡]	1.000	-	-	-
	0.594 [‡]	0.641 [‡]	1.000	-	-
	0.734 [‡]	0.768 [‡]	0.873 [‡]	1.000	-

([‡] $p < 0.01$)

설계, 관찰, 설명과 해석의 4개요인 모두 높은 상관관계를 보이고 있다. 이와 같은 점은 내적 일관성 신뢰도 분석에서 전체 요인을 하나로 설정한 분석에서 높은 내적 일관성(0.89)을 통해서도 확인할 수 있다. 특히 관찰에 관한 반회의주의적 태도는 다른 요인과 상관관계가 상대적으로 낮게 나타난다. 실제 이 연구에서 이와 같은 현상에 대한 근본적인 이유를 추론해 낼 수는 없지만 관찰 영역의 문항이 다른 영역에 비하여 보다 행동적 관점으로 접근되어 있는 것으로 조심스럽게 추론한다. 문제 인식과 가설 생성, 실험 설계, 설명과 해석은 자신의 생각과 믿음에 관한 문항이 많다면 관찰은 실제 행동에 관한 질문으로 구성되어 있다. 이와 같은 차이점이 관찰 요인의 점수가 다른 요인과 상대적으로 상관관계가 낮은 것의 원인이 아닐까 추측한다.

3. 고등학생들의 성별과 계열별에 따른 과학적 회의주의 점수 비교

이 연구를 통해 확인된 과학탐구 상황에서 고등학생들의 반회의주의적 태도 점수는 Figure 1에 제시되어 있다. 이원분산분석의 그래프를 막대그래프가 아닌 선형그래프로 제시한 이유는 두 선의 평행 정도를 통해 교호작용효과를 효과적으로 보기 위함이다. 문제 인식과 가설 설정 영역의 점수를 살펴보면, 문과계열의 경우 남학생의 평균 점수는 -0.269이었으며, 여학생의 평균 점수는 -0.176이었다. 이과계열의 경우 남학생의 평균 점수는 -0.163, 여학생의 평균 점수는 0.039이었다. 실험 설계의 경우 문과계열 남학생의 점수는 -0.261, 여학생은 -0.223이었으며, 이과계열의 경우 남학생이 -0.138, 여학생은 0.121의 평균 점수를 보였다. 관

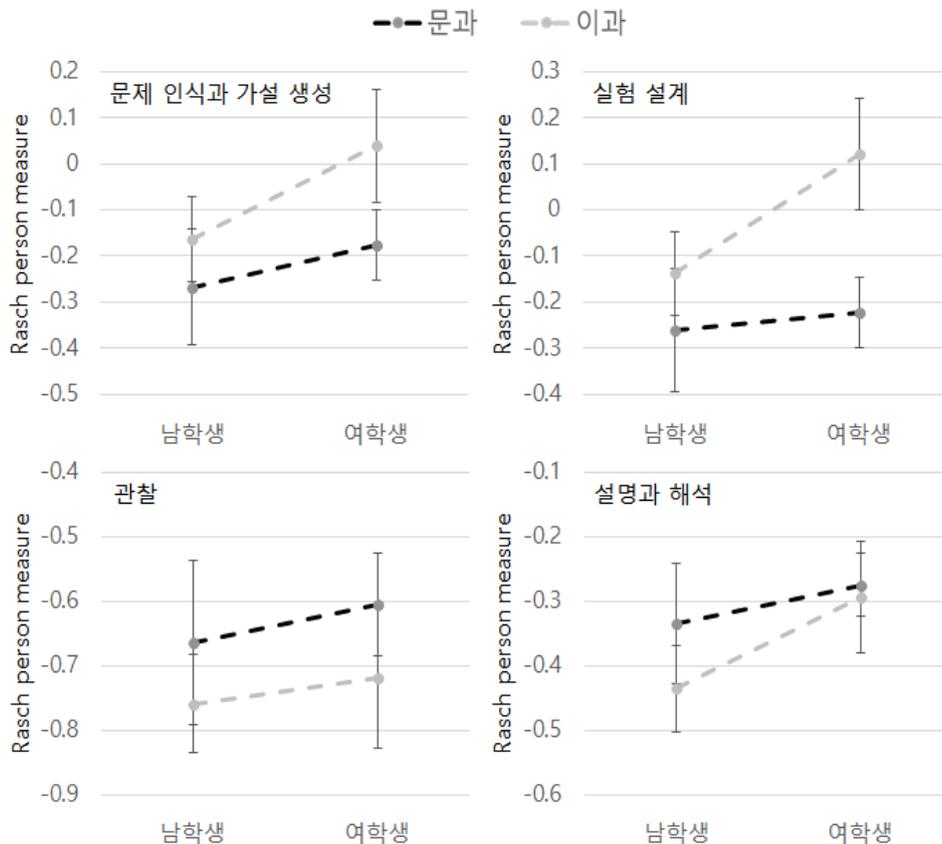


Figure 1. Unskeptical attitude scores of groups by academic track and gender

찰에서의 반회의주의적 태도를 보면 문과 계열의 남학생은 -0.664, 여학생은 -0.605의 점수를 보였다. 이과계열의 경우 남학생은 -0.759, 여학생은 -0.718의 평균 점수를 보였다. 마지막으로 설명의 경우 문과 계열의 남학생은 -0.335, 여학생은 -0.275의 점수를 보였다. 이과계열의 남학생은 -0.436, 여학생은 -0.294이었다. 이원분산분석을 나타내는 Figure 1을 보면 직관적으로 문제 인식과 가설 설정, 실험 설계에서는 이과계열의 학생의 평균값이 더 높으며, 관찰과 설명과 해석 영역에서는 문과계열의 학생들의 평균값이 더 높은 것을 확인할 수 있다. 전체적으로 여학생의 평균값이 더 높은 것을 확인할 수 있다.

보다 면밀한 통계 분석을 위하여 이원분산분석의 통계값들을 확인하였다(Table 6). 이원분산분석 결과 계열별, 성별 의미 있는 차이는 실험설계에서 계열별 차이를 제외하고 없는 것으로 확인되었다. 성별과 계열별 교호작용효과도 유의미하지 않았다. 과학탐구상황에서 반회의주의적 태도가 특정 그룹에서 형성되기 보다는 전체적으로 비슷한 수준인 것으로 이해할 수 있다. 특히 실험 설계에서 이과 계열의 학생들이 문과계열의 학생들에 비하여 0.05 수준에서 유의미한 차이를 보이는 것은 중요하게 논의할 가치가 있다. 이과계열의 학생들이 문과 학생들에 과학탐구 실험 수업을 보다 많이 경험하였을 것이며, 대부분의 실험들은 교과서에 제시된 단순한 실험이었을 가능성이 높다. 이와 같은 교과서

중심의 과학 실험은 확인 실험 중심의 무비판적인 활동을 반복할 가능성이 높을 것이다. 아마도 그런 영향이 일부 미치지 않았을까 조심스럽게 판단한다.

IV. 결론 및 제언

과학탐구 능력의 향상 및 비판적 사고, 합리적 의사결정 능력의 함양을 위하여 과학적 회의주의 교육의 필요성이 제기된다. 이 연구에서는 과학탐구 상황에서 반회의주의적 태도를 점검할 수 있는 검사도구를 개발하였다. 개발된 검사도구는 과학탐구 상황에서 회의주의적 태도를 견지하기 위하여 자기 점검용 도구로 활용될 수 있으며, 다양한 과학탐구 활동의 효과 검증용으로도 활용될 수 있다. 이 연구의 결론을 내리면 다음과 같다.

먼저 과학탐구상황에서 반회의주의 태도를 측정할 수 있는 23개 리커트 형태의 자기보고형 검사도구를 개발하였다. 반회의주의적 태도 측정을 위하여 과학탐구 상황을 크게 ‘문제 인식과 가설 생성’, ‘실험 설계’, ‘관찰’, ‘설명과 해석’으로 구분하여 각 구인별로 측정할 수 있는 다수의 문항을 개발하였다. 개발된 문항은 학생들이 이해하기 쉬운 언어로 구성되었다.

둘째, 과학탐구상황에서 반회의주의 태도 측정 문항의 타당도와 신뢰도를 엄격한 방법을 통해 확인

Table 6. Results of two-way ANOVA

							*		
	<i>F</i>	<i>p</i>	<i>PES</i>	<i>F</i>	<i>p</i>	<i>PES</i>	<i>F</i>	<i>p</i>	<i>PES</i>
가	1.962	0.162	0.006	2.299	0.130	0.007	0.269	0.604	0.001
	1.937	0.165	0.006	4.798	0.029	0.014	1.089	0.297	0.003
	0.263	0.609	0.001	1.134	0.288	0.003	0.009	0.925	0.000
	1.783	0.183	0.005	0.640	0.424	0.002	0.291	0.590	0.001

하였다. 라쉬 분석을 통해 다차원 부분점수모델의 분석의 적합성을 확인하였으며, 라쉬 분석을 통해 문항 적합도를 확인한 결과 23개 문항 모두 적합으로 나타났다. EAP/PV Reliability, Separation Reliability 등 라쉬 분석에서 생성되는 신뢰도 지수도 적합하였으며, 자료의 내적 일관성 신뢰도인 Cronbach α 도 기준에 부합하는 것으로 확인되었다. 라쉬 분석의 차별적 문항 분석 기능인 DIF 차이값을 확인한 결과 성별과 계열별로 문제 인식과 가설 설정에서 1문항씩 기준에 약간 벗어나는 수준으로 개발된 문항이 집단간 응답반응에 의미 있는 수준의 영향을 주지 않는 것으로 확인되었다. 문제 인식과 가설 생성, 실험 설계, 관찰, 설명과 해석의 4개 요인의 점수 모두 높은 상관관계를 보이고 있었다.

남학생과 여학생, 문과계열과 이과계열별로 구분하여 반회의주의적 태도를 비교한 결과 유의미한 차이를 나타낸 비교는 실험설계에서 계열별 차이를 제외하고 없는 것으로 확인되었다. 성별과 계열별 교호작용효과도 유의미하지 않았다. 과학탐구상황에서 반회의주의적 태도가 특정 그룹에서 형성되기 보다는 전체적으로 비슷한 수준인 것으로 이해할 수 있다. 특히 실험 설계에서 이과계열의 학생들이 문과계열의 학생들에 비하여 0.05 수준에서 유의미한 차이를 보이는 것은 교과서 중심의 과학 실험은 확인 실험 중심의 무비판적인 활동을 반복할 가능성이 영향을 미치지 않았을까 조심스럽게 추측해 본다.

개발된 문항들을 활용하여 과학탐구 과정에서 반회의주의적 태도를 점검하고, 과학탐구 수업의 효과를 비교하는데 쉽게 활용될 수 있을 것이다. 23개 문항을 응답하는 시간은 5분이 되지 않을 정도이기 때문에 교실 현장에서 충분히 활용 가능성이 높다. 이상의 결론을 제시하면서 이 연구는 최소한 서론에서 제시한 연구 문제를 충분히 확인할 수 있었다.

그럼에도 불구하고 이 연구가 해결하지 못한 다

양한 제한점을 밝히고자 한다. 먼저 자기 보고형의 검사지의 한계점이다(Baars *et al.*, 2014; Kostons, Van Gog & Paas, 2012; Panadero *et al.*, 2012). 자기 보고형의 검사도구는 실용적인 측면에서 다양한 이점이 있기 때문에 널리 활용되며, 대안이 없기 때문에 근본적인 이유를 바탕으로 비판하기 어렵다. 그럼에도 불구하고 자기 보고형 검사도구를 해석할 때 주의해서 되는 것은 명백하다. 예를 들어 남자의 경우 여자에 비하여 능력을 과대평가하거나(Pajares & Graham, 1999; Pallier, 2003; Stankov & Lee, 2008), 어설픈 전문가들이 전혀 전문성이 없는 비전문가에 비해서도 자기 능력을 높게 평가하는 경향이 있다(Angner, 2006; Menkhoff *et al.*, 2013; Quirk, 2010; Chi, 2006). 자기 보고형 검사도구는 동료평가나 교사평가의 용도로도 활용될 수 있다는 점에서 일부 단점을 극복할 수 있다. 개발된 반회의주의적 태도 검사도구를 자기 평가로 할 수 있지만 모둠 활동 후 동료들이 평가할 수도 있다. 물론 개인의 생각이나 믿음과 같이 행동영역으로 드러나지 않는 부분은 한계가 있을 수 있다.

마지막으로 과학적 회의주의와 과학탐구 상황에서의 회의주의 등 과학 철학적으로 중요한 주제에 대하여 쉽게 접근하여 문항을 구성하였는지에 대한 비판이 제기될 수 있다. 학자마다 견해가 다를 수 있기 때문에 내용타당도에 대한 이의 제기가 있을 수 있다. 적어도 이 연구는 연구자들이 접근할 수 있는 문헌과 전문성을 근거로 문항을 개발하였다고 판단한다. 합리적 인재 양성과 지식 생성의 역량 기반 교육으로의 패러다임의 변화에 따라 과학적 회의주의의 중요성이 강조되고 있기 때문에 측정 문항은 반드시 개발되어야 한다. 모든 학자들이 일치된 의견이 제시될 때까지 평가 문항 개발을 하지 않고 기다리는 것은 바람직한 것은 아니다. 아마도 이 연구에서 개발된 문항들을 시작으로 비판하면서 과학적 회의주의에 관한 다양한 검사도구들이 개발될 수 있을 것으로 기대한다.

참 고 문 헌

- Abrami, P. C., Bernard, R. M., Borokhovski, E., Waddington, D. I., Wade, C. A., & Persson, T. (2015). Strategies for teaching students to think critically: A meta-analysis. *Review of Educational Research, 85*(2), 275-314.
- Aikenhead, G. S. (2006). *Science Education for Everyday Life: Evidence-based Practice*. New York, NY: Teachers College Press.
- Angner, E. (2006). Economists as experts: Overconfidence in theory and practice. *Journal of Economic Methodology, 13*(1), 1-24.
- Baars, M., Vink, S., van Gog, T., de Bruin, A., & Paas, F. (2014). Effects of training self-assessment and using assessment standards on retrospective and prospective monitoring of problem solving. *Learning and Instruction, 33*, 92-107.
- Beyer, B. K. (1995). *Critical thinking. Fastback 385. Phi Delta Kappa*, 408 N. Union, PO Box 789, Bloomington, IN 47402-0789.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. New York, NY: Springer.
- Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). 'An experiment is when you try it and see if it works': a study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education, 11*(5), 514-529.
- Chi, M. T. (2006). Two approaches to the study of experts' characteristics. In K. A. Ericson, N. Charness, P. J. Feltovich & R. R. Hoffman (Eds.), *The Cambridge handbook of expertise and expert performance* (pp. 21-30). New York, NY: Cambridge University Press.
- Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education, 86*(2), 175-218.
- DeVellis, R. F. (2003). *Scale development: Theory and applications*. Thousand Oaks, CA: Sage Publications.
- Driver, R., Asoko, H., Leach, J., Scott, P., & Mortimer, E. (1994). Constructing scientific knowledge in the classroom. *Educational Researcher, 23*(7), 5-12.
- 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.
- Gilovich, T. (2008). *How we know what isn't so*. New York, NY: Simon and Schuster.
- Glaser, E. M. (1941). *An experiment in the development of critical thinking* (No. 843). New York, NY: Teachers College, Columbia University.
- Heyman, G. D., Fu, G., & Lee, K. (2007). Evaluating claims people make about themselves: The development of skepticism. *Child Development, 78*(2), 367-375.
- Hofer, B. K., & Pintrich, P. R. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67*(1), 88-140.
- Kahneman, D. (2013). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.
- Korsgaard, C. M. (1986). Skepticism about practical reason. *The Journal of Philosophy, 83*(1), 5-25.

- Kostons, D., Van Gog, T., & Paas, F. (2012). Training self-assessment and task-selection skills: A cognitive approach to improving self-regulated learning. *Learning and Instruction, 22*(2), 121-132.
- Lipman, M. (1987). Critical thinking: What can it be? *Analytic Teaching, 8*(1), 5-12.
- Macfarlane, B., & Cheng, M. (2008). Communism, universalism and disinterestedness: Re-examining contemporary support among academics for Merton's scientific norms. *Journal of Academic Ethics, 8*(1), 67-78.
- McPeck, J. E. (1981) *Critical Thinking and Education*. Oxford, UK: Martin Robertson.
- Menkhoff, L., Schmeling, M., & Schmidt, U. (2013). Overconfidence, experience, and professionalism: An experimental study. *Journal of Economic Behavior & Organization, 86*, 92-101.
- Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*(4), 5-8.
- Ministry of Education [MOE]. (2015). *2015 revised Science National Curriculum*. Sejong: Author.
- 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.
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220.
- Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology, 24*(2), 124-139.
- Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles, 48*(5), 265-276.
- Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences, 22*(6), 806-813.
- Parascandola, M. (2004). Skepticism, statistical methods, and the cigarette: A historical analysis of a methodological debate. *Perspectives in Biology and Medicine, 47*(2), 244-261.
- Popkin, R. H. (1967). "Skepticism". In P. Edwards (Ed.), *The Encyclopedia of Philosophy* (7, pp. 449-461) New York, NY: The Macmillan Company & The Free Press.
- Quirk, P. J. (2010). The trouble with experts. *Critical Review, 22*(4), 449-465.
- Slovan, S., Barbey, A. K., & Hotelling, J. M. (2009). A causal model theory of the meaning of cause, enable, and prevent. *Cognitive Science, 33*(1), 21-50.
- Stankov, L. & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology, 100*(4), 961.
- Trout, J. D. (2002). Scientific explanation and the sense of understanding. *Philosophy of Science, 69*(2), 212-233.
- Wright, B. D. & Linacre, J. M. (1994). Reasonable mean-square-fit values. *Rasch Measurement Transactions, 8*(3), 370.

국 문 요 약

이 연구는 과학탐구 상황에서 반회의주의적 태도를 조사하는 문항을 개발하는 것이 목적이다. 문헌 연구, 전문가 검토, 통계적 타당도 검증을 거쳐 문항을 개발하고, 성별, 계열별로 점수 차이를 확인하였다. 이 연구를 위하여 363명의 고등학생들이 참여하였다. 문항 타당도 확인을 위하여 라쉬 분석, 내적 일관성 신뢰도 확인 등을 수행하였고, 성별과 계열별 반회의주의적 태도의 점수 비교를 위하여 이원분산분석을 확인하였다. 연구 결과 과학탐구상황에서 반회의주의 태도를 측정할 수 있는 23개 리커트 형태의 자기보고형 검사도구를 개발하였다. 반회의주의적 태도 측정을 위하여 과학탐구 상황을 크게 '문제 인식과 가설 생성', '실험 설계', '관찰', '설명과 해석'으로 구분하여 각 구인별로 측정할 수 있는 다수의 문항을 개발하였다. 둘째, 과학탐구상황에서 반회의주의 태도 측정 문항의 타당도와 신뢰도를 엄격한 방법을 통해 확인하였다. 라쉬 분석을 통해 다차원 부분점수모델의 분석의 적합성을 확인하였으며, 라쉬 분석을 통해 문항 적합도를 확인한 결과 23개 문항 모두 적합으로 나타났다. 다양한 신뢰도 지수도 적합한 것으로 나타났다. 남학생과 여학생, 문과계열과 이과계열별로 구분하여 반회의주의적 태도를 비교한 결과 유의미한 차이를 나타낸 비교는 실험 설계에서 계열별 차이를 제외하고 없는 것으로 확인되었다. 개발된 문항들을 활용하여 과학탐구 과정에서 반회의주의적 태도를 점검하고, 과학탐구 수업의 효과를 비교하는데 활용될 수 있을 것이다.

주제어: 과학적 회의주의, 과학탐구, 문항 개발, 과학 교육