

2015 개정 과학과 교육과정에 기초한 과학과 핵심역량 조사 문항의 개발 및 적용

하민수¹, 박현주², 김용진^{3*}, 강남화⁴, 오피석⁵, 김미점³, 민재식³, 이윤형³, 한효정³, 김무경³, 고성우⁶, 손미현⁷
¹강원대학교, ²조선대학교, ³경상대학교, ⁴한국교원대학교, ⁵경인교육대학교, ⁶아라초등학교, ⁷서울대학교

Developing and Applying the Questionnaire to Measure Science Core Competencies Based on the 2015 Revised National Science Curriculum

Minsu Ha¹, HyunJu Park², Yong-Jin Kim^{3*}, Nam-Hwa Kang⁴, Phil Seok Oh⁵, Mi-Jum Kim³, Jae-Sik Min³,
Yoonhyeong Lee³, Hyo-Jeong Han³, Moogyong Kim³, Sung-Woo Ko⁶, Mi-Hyun Son⁷

¹Kangwon National University, ²Chosun University, ³Gyeongsang National University, ⁴Korea National University of Education,
⁵Gyeongin National University of Education, ⁶Ara Elementary School, ⁷Seoul National University

ARTICLE INFO

Article history:

Received 21 May 2018

Received in revised form

30 June 2018

9 August 2018

Accepted 18 August 2018

Keywords:

2015 revised national
science curriculum,
science core competency,
questionnaire of core
competency

ABSTRACT

This study was conducted to develop items to measure scientific core competency based on statements of scientific core competencies presented in the 2015 revised national science curriculum and to identify the validity and reliability of the newly developed items. Based on the explanations of scientific reasoning, scientific inquiry ability, scientific problem-solving ability, scientific communication ability, participation/lifelong learning in science presented in the 2015 revised national science curriculum, 25 items were developed by five science education experts. To explore the validity and reliability of the developed items, data were collected from 11,348 students in elementary, middle, and high schools nationwide. The content validity, substantive validity, the internal structure validity, and generalization validity proposed by Messick (1995) were examined by various statistical tests. The results of the MNSQ analysis showed that there were no nonconformity in the 25 items. The confirmatory factor analysis using the structural equation modeling revealed that the five-factor model was a suitable model. The differential item functioning analyses by gender and school level revealed that the nonconformity DIF value was found in only two out of 175 cases. The results of the multivariate analysis of variance by gender and school level showed significant differences of test scores between schools and genders, and the interaction effect was also significant. The assessment items of science core competency based on the 2015 revised national science curriculum are valid from a psychometric point of view and can be used in the science education field.

1. 서론

최근 세계 교육 추세는 ‘미래 사회를 위한 준비’라는 보편성을 내세우며 학교 교육의 방향을 근본적으로 재조정하려는 노력을 기울이고 있다. 기존 교육체계의 한계를 인식하고 학습자가 미래 사회 변화에 대해 유연하게 적응하고, 더 나아가 변화를 선도할 능력을 함양할 수 있도록 교육적 변화를 도모하고 있는 것이다. 이러한 움직임의 핵심 키워드로 등장한 것이 핵심역량(core competencies)이다(Kwak, 2013). 이것은 학교교육에서 지향할 것은 정형화된 지식이 아니라 학습자가 실제로 정보와 지식을 활용할 수 있는 능력의 함양이라는 인식에서 근거한 것이다(Jin, Sohn & Chu, 2011; Kim, 2014; Kim, et al., 2009; Lee et al., 2010; OECD, 2003; 2006; So et al., 2010; Yoon et al., 2007).

세계 각국은 교육의 질적 제고를 위해 이미 역량중심의 교육과정 체제로 개편하였으며(OECD, 2003), 우리나라는 2015 개정 교육과정에서 핵심역량 중심의 교육과정을 제시하였다(MOE, 2015). 이는 미래 사회에서 요구되는 인재 육성에 대한 개인적·사회적 요구가 반영

되어 지식 중심의 교육과정에서 핵심역량 중심의 교육과정 패러다임으로 변화한 것이다. 즉 학교교육을 통해 전인적 측면에서 학생들의 다면적 역량을 키우고 궁극적으로 국가 차원의 경쟁력을 제고하기 위한 것이다. 2015 개정 교육과정의 변화에 따라 각 교과에서도 역량 중심의 교육과정으로 변화하였다. 이에 따라 과학과 교육과정도 핵심역량을 설정하고 그에 맞게 개정되었다. 과학과 핵심역량은 과학적 사고력, 과학적 탐구능력, 과학적 문제해결력, 과학적 의사소통 능력 및 과학적 참여와 평생학습 능력이다. 새로운 교육과정에 바탕을 둔 학교 과학교육은 위의 다섯 가지 과학과 핵심역량을 향상시키는 것을 주된 목표로 하고 있다.

핵심역량 기반 교육이 강조되면서 학생들의 역량 및 변화 추이를 분석하기 위한 핵심역량 조사도구 개발과 관련된 다양한 연구들이 진행되어 왔다(Noh & Lee, 2013a; 2013b). 한국교육개발원의 학생역량 조사 도구 개발 및 지수 분석 연구(Kang et al., 2014; Kim et al., 2013), 한국교육과정평가원의 핵심역량 관련 교육과정 연구(Lee et al., 2008; Lee et al., 2009; Lee et al., 2013; Yoon et al., 2007), 한국청소년정책연구원의 청소년들의 핵심역량 진단을 위한 조사 도

* 교신저자 : 김용진 (gnuedu@nate.com)

<http://dx.doi.org/10.14697/jkase.2018.38.4.495>

구 개발 및 지수 분석 연구(Chang *et al.*, 2015; Kim *et al.*, 2008; Kim *et al.*, 2009; Seong, Baek & Jin, 2014) 등이 추진된 바 있다. 대학생을 대상으로 한 역량 진단 도구도 개발되어 왔으며, 한국직업능력개발원의 K-CESA가 대표적이다. 대학 자체적인 역량을 설정하고 그에 따른 진단 도구를 개발한 사례는 동덕여대의 역량 진단 도구, 숙명여자대학교의 SM-CLA, 성균관대학교의 SCCA, 건양대학교 핵심역량 진단 평가, 동국대학교의 동국 역량진단과 D-CESA의 핵심역량 진단 도구 등이 있다(Son, Song & Lee, 2016). 핵심역량을 측정하기 위한 검사도구 개발과 더불어 이를 활용한 실증연구(Baek, 2013) 및 그 결과를 교과·비교과 프로그램과 연계하는 연구(Kim & Lee, 2012; 2013; Lee & Lee, 2012)들도 일부 진행되었다.

그러나 지금까지 핵심역량을 측정하기 위해 수행된 연구들은 주로 역량을 범주화하기 위한 다양한 기준을 적용하고, 일반적인 핵심역량을 진단하는데 초점이 맞추어져 있었다. 또한 학생역량 지표와 국가 교육과정과의 연계성이 부족하고, 학생역량만을 위주로 조사하였기 때문에 학생역량 현황 외에 추가적으로 얻을 수 있는 정보가 부족한 아쉬움이 있다.

과학과 핵심역량은 2015 개정 과학과 교육과정의 주요 교육 목표로 설정(MOE, 2015)되어 있으며, 교육과정 문서를 통해 과학과 핵심역량의 특성이 제시되어 있다. 새로운 과학과 교육과정의 효과를 측정하거나 역량 신장을 위한 목표로 개발된 많은 교수·학습 프로그램들의 효용성을 확인하기 위해서는 학생들의 과학과 핵심역량을 조사하고, 그 변화를 측정 할 수 있는 효율적인 검사도구의 개발이 필수적이다. 따라서 본 연구는 2015 개정 과학과 교육과정에서 기술하고 있는 과학과 핵심역량의 영역별 내용에 기초하여 학생들의 과학과 핵심역량을 조사할 수 있는 문항들을 개발하였다. 구체적인 연구내용은 다음과 같다.

1. 개발된 과학과 핵심역량 문항의 타당도와 신뢰도는 어떠한가?
2. 성별, 학교급별 과학과 핵심역량의 점수는 유의미하게 차이가 나타나는가?
3. 성별, 학교급별 과학과 핵심역량의 요인별 상관관계에 의미 있는 차이가 있는가?

II. 연구 방법

1. 문항 개발

2015년도에 개정된 교육과정은 각 교과별로 역량을 설정하여 제시하였다. 과학과 핵심역량은 과학적 사고력, 과학적 탐구능력, 과학적 문제해결력, 과학적 의사소통능력, 과학적 참여와 평생 학습 능력 등의 5가지로 정하고, 각 핵심역량별 설명이 교육과정에 제시되어 있다. 따라서 본 연구에서 개발한 과학과 핵심역량 능력에 대한 인식 조사

문항은 2015 개정 과학과 교육과정에서 기술하고 있는 과학과 핵심역량에 대한 설명을 바탕으로 구성하였다. 각 역량별 5문항씩 총 25개 문항을 과학교육전문가 5인과 교육심리전문가 1인이 개발하고 초·중·고 현장 교사 6인과 함께 협의와 검토 과정을 거쳤다. 이후 현장 교사의 소속 학교에서 한 학급씩의 학생들을 대상으로 예비조사를 하여 문항의 이해 정도를 파악하고 수정 보완하였다. 학교 현장에서 교사들은 과학과 교육과정의 설명을 바탕으로 핵심역량 향상을 위한 수업 목표, 내용, 활동, 평가 등을 설계할 것으로 기대된다. 그러므로 교육과정에 제시된 핵심역량의 설명에 기초하여 핵심역량을 측정하는 문항을 구성하는 것은 교육과정과 평가 문항의 일관성을 유지하는 가장 효율적인 방법이 될 것이다.

2. 자료 수집

2015 개정 과학과 교육과정에 근거하여 개발한 과학과 핵심역량 문항의 타당도와 신뢰도, 그리고 집단별 점수 차이 등을 확인하기 위하여 초·중·고등학교 학생 11,348명을 대상으로 자기 평가 형태의 설문 조사를 하였다. 조사에 참여한 학생들이 25개 문항에 응답하는 시간은 약 10분 내외가 소요되었다.

본 연구의 조사에 응답한 학생들은 전국 17개 시·도에서 학생 참여형 과학수업 선도학교를 운영하고 있는 초·중·고 60개교에 재학 중인 학생들이다. 선도학교는 서울, 부산, 인천, 대전, 울산, 대구, 광주, 세종시 등의 8개 도시와 경기, 강원, 충북, 충남, 경북, 경남, 전북, 전남, 제주도 등의 9개도에서 초등학교, 중학교, 고등학교별로 각각 1개교 이상씩 지정되어 있다. 따라서 본 연구의 조사 대상은 전국적인 표본에 근접한다고 볼 수 있다. 조사에 응답한 학생들의 분포는 Table 1과 같다.

3. 문항 타당도 분석 및 통계 검정 방법

과학과 핵심역량 문항의 타당도와 신뢰도를 확인하기 위하여 Messick(1995)의 타당도 모델에 근거하여 다양한 통계분석이 실시되었다. 첫째, 문항에 대한 학생들의 반응 등 문항 적합도와 여러 신뢰도를 확인하기 위하여 라쉬 모델 분석을 활용하여 Infit MNSQ, Outfit MNSQ, Person reliability, Item reliability를 확인하였다. 둘째, 내적 일관성 신뢰도를 확인하기 위하여 Cronbach alpha도 확인하였다. 셋째, 내적 구조에 관한 타당도와 관련된 개발된 문항의 내적 구조의 적합성을 확인하기 위하여 구조방정식을 활용한 확인적 요인분석을 실시하였다. 넷째, 적합도는 RMR(root mean square residual), GFI(goodness of fit index), NFI(normed fit index), IFI(incremental fit index), TLI(Tucker Lewis index), CFI(comparative fit index), RMSEA(root mean square error of approximation)을 통해 확인하였

Table 1. Number of participants by gender and school level.

성별	학교급				특성화고	과학고	전체
	초등학생	중학생	고등학생 (문과희망)	고등학생 (이과희망)			
남학생	1,769	2,594	552	801	283	98	6,097
여학생	1,718	1,786	842	858	0	47	5,251
전체	3,487	4,380	1,394	1,659	283	145	11,348

다. 다섯째, 차원 분석은 ConQuest 4.5.0을 활용하여 Akaike Information Criterion(AIC), Final Deviance, Chi-square의 값을 비교하는 방법을 활용하였다. 마지막으로, 성별과 학교급별 점수 비교를 위하여 이원분산분석을 활용하였으며, 각 요인별 상관관계 확인을 위하여 Pearson 상관관계를 확인하였다. 본 연구에서 라쉬모델 분석은 WINSTEPS 3.92.1을 사용하였으며, 기타 통계 분석은 SPSS v.22를 사용하였다.

III. 연구 결과 및 논의

1. 조사 문항의 타당도 및 신뢰도

Messick(1995)이 제안한 구인 타당도 중에서 첫 번째는 내용타당도이다. 본 연구에서 개발하고자 하는 과학적 핵심역량에 관한 조사 문항의 내용은 전문가 또는 이론적으로 정의된 과학 역량의 내용과 일치해야 된다. 과학적 역량 문항의 내용타당도를 확보하기 위하여 이 연구에서는 2015 개정 과학과 교육과정에서 제시된 과학적 역량에 대한 정의를 최대한 활용하였다(Table 2). 예를 들어 Table 2에 제시된 과학적 사고력에 대한 2015 개정 과학과 교육과정의 설명과 개발된 문항을 살펴볼 수 있다.

2015 개정 과학과 교육과정에서는 과학적 사고력의 구인을 정의할

때 과학적 주장과 증거의 관계, 과학의 지식과 방법, 논리적 추론, 독창적 아이디어 산출 등을 과학적 사고력의 중요한 구성요소로 제시하였다. 이 요소들을 토대로 문항을 구성하였으며, 학생들의 수준에서 이해하기 어려운 내용에 해당하는 일부 요소들은 문항에 포함시키지 않았다. 예를 들면, 과학적 사고력에서 과학적 세계관과 자연관에 대하여 교육과정에 제시되었으나 과학적 세계관과 자연관에 대한 학생들의 이해가 낮기 때문에 이 항목들은 조사 문항에 포함되지 않았다. 이 점은 문항 개발의 제한점으로 이해될 수 있다. 해당 구인의 중요한 내용 요소가 빠져서 문항의 대표성이 낮아지는 문제점이 발생할 수 있으나 학생들이 이해할 수 없는 문항을 포함시키는 것 역시 다른 문제를 야기한다. 또한 이 연구에서 개발한 문항들이 과학과 핵심역량의 다양한 요소들을 학생들이 신장시켰는지 진단하는 기능 보다는 각 구인별로 학생들의 수준을 가늠하는 기능의 수준에서는 일부 내용요소가 빠진다고 하여 문제되지 않는다고 판단하였다.

이와 같은 다양한 판단 과정을 거쳐 문항에 대한 1차 초안을 만들었고, 과학교육 전문가 5인을 포함한 연구진 전체에 의하여 지속적인 협의를 거쳐 최종 25개 문항을 개발하였다. 2015 개정 과학과 교육과정의 서술문이 교육과정 개발에 참여한 과학교육 전문가들의 협의체에 의해 개발되었고, 핵심역량 조사 문항의 개발에 참여한 과학교육 전문가들의 전문성을 고려할 때 개발된 문항의 내용 타당도는 충분하다고 판단하였다.

Table 2. Explanation of science competency in 2015 revised science curriculum and items developed in this study.

과학 핵심역량	2015 개정 과학과 교육과정의 설명	개발된 문항
과학적 사고력	과학적 사고력은 과학적 주장과 증거의 관계를 탐색하는 과정에서 필요한 사고이다. 과학적 세계관 및 자연관, 과학의 지식과 방법, 과학적인 증거와 이론을 토대로 합리적이고 논리적으로 추론하는 능력, 추리 과정과 논증에 대해 비판적으로 고찰하는 능력, 다양하고 독창적인 아이디어를 산출하는 능력 등을 포함한다.	1. 나는 의견을 주장 할 때, 논리적 근거를 들어서 한다. 2. 나는 어떤 일을 판단 할 때, 과학적 증거가 있는지 잘 따져본다. 3. 나는 어떤 일을 판단 할 때, 과학적 지식을 잘 사용한다. 4. 나는 다른 사람의 말이나 토론에서 옳고 그름을 논리적으로 잘 따진다. 5. 나는 남들이 생각해 내지 못하는 아이디어를 자주 만들어낸다.
과학적 탐구 능력	과학적 탐구 능력은 과학적 문제 해결을 위해 실험, 조사, 토론 등 다양한 방법으로 증거를 수집, 해석, 평가하여 새로운 과학 지식을 얻거나 의미를 구성해 가는 능력을 말한다. 과학적 탐구를 위해서는 과학 탐구 기능과 지식을 통합하여 적용하고 활용하는 능력이 필요하며 과학적 사고력이 이 과정에 기초가 된다.	1. 나는 관찰이나 실험, 제작 등의 활동을 할 때, 필요한 도구를 잘 다룬다. 2. 나는 관찰이나 실험을 통해 증거를 찾고 해석하는 것을 잘 한다. 3. 나는 어떤 일이 생기면 직관적으로 판단하기보다는 여러 상황을 자세히 조사한다. 4. 나는 과학 탐구를 할 때, 관련된 과학 지식과 잘 연계하여 생각한다. 5. 나는 과학 탐구의 결과를 통해 과학적 지식이나 의미를 잘 구성해 낸다.
과학적 문제 해결력	과학적 문제 해결력은 과학적 지식과 과학적 사고를 활용하여 개인적 혹은 공적 문제를 해결하는 능력이다. 일상생활의 문제를 해결하기 위해 문제와 관련 있는 과학적 사실, 원리, 개념 등의 지식을 생각해 내고 활용하며 다양한 정보와 자료를 수집, 분석, 평가, 선택, 조직하여 가능한 해결 방안을 제시하고 실행하는 능력이 필요하다. 문제 해결력은 문제 해결 과정에 대한 반성적 사고 능력과 문제 해결 과정에서의 합리적 의사 결정 능력도 포함한다.	1. 나는 과학 관련 문제를 해결할 때, 과학적 지식을 잘 사용한다. 2. 나는 조사나 실험 등으로 증거 자료를 잘 찾고 분석하여 과학적 문제를 해결한다. 3. 나는 일상생활에서의 현상들에 대해 어떤 과학적 지식과 관련이 있는지 곰곰이 생각한다. 4. 나는 과학 문제를 해결하기 위해 정보매체 (PC, 스마트폰이나 신문 등)로 관련 자료를 잘 찾아낸다. 5. 나는 일상생활에서 문제가 생기면 합리적인 해결 방안을 찾으려고 깊이 생각한다.
과학적 의사소통 능력	과학적 의사소통 능력은 과학적 문제 해결 과정과 결과를 공동체 내에서 공유하고 발전시키기 위해 자신의 생각을 주장하고 타인의 생각을 이해하며 조정하는 능력을 말한다. 말, 글, 그림, 기호 등 다양한 양식의 의사소통 방법과 컴퓨터, 시청각 기기 등 다양한 매체를 통하여 제시되는 과학기술 정보를 이해하고 표현하는 능력, 증거에 근거하여 논증 활동을 하는 능력 등을 포함한다.	1. 나는 어떤 문제를 해결하는 과정을 친구에게 효율적으로 설명한다. 2. 나는 과학 지식이나 학습 결과물을 정보매체 (PC, 스마트폰 등)의 프로그램을 사용하여 잘 표현한다. 3. 나는 과학 지식이나 학습 결과물을 글이나 그림, 기호 등으로 종이나 칠판에 잘 나타낸다. 4. 나는 과학 문제와 관련하여 다른 사람과 질의 응답하는 토의나 토론을 잘한다. 5. 나는 다른 사람의 말을 끝까지 잘 듣고 판단한 후 내 의견을 조리 있게 말한다.
과학적 참여와 평생 학습 능력	과학적 참여와 평생 학습 능력은 사회에서 공동체의 일원으로 합리적이고 책임 있게 행동하기 위해 과학기술의 사회적 문제에 대한 관심을 가지고 의사 결정 과정에 참여하며 새로운 과학기술 환경에 적응하기 위해 스스로 지속적으로 학습해 나가는 능력을 가리킨다.	1. 나는 과학기술로 생기는 사회 문제와 그 해결 방안을 찾는 것에 관심이 많다. 2. 나는 협동적으로 해야 하는 과학 활동의 경우, 적극 참여하여 내가 맡은 책임을 다한다. 3. 나는 친구들의 학습 활동이나 결과물에 대해 평가할 경우, 공정하게 평가한다. 4. 나는 새로운 과학기술 제품이 나오면 그 사용법을 쉽게 이해하고 익힌다. 5. 나는 나이가 들어도 새로운 과학 지식이나 정보를 찾아 배울 것이라고 생각한다.

두 번째로 확인한 구인 타당도는 실제에 기초한 타당도 (Gadermann, Guhn & Zumbo, 2011)이다. 실제에 기초한 타당도는 피평가자와의 면담을 통해 확인하거나 본 연구에서와 같이 라쉬모델 분석을 통해 양적방법으로 확인할 수 있다. 이 방법은 학생들이 특정 문항에 대하여 비정상적 반응을 하는 경우가 어느 정도인지를 이론적으로 예측된 확률과 실제 자료를 비교하는 방법으로 확인한다. 문항 적합도는 라쉬모델 분석을 통해 획득한 문항적합도인 Infit MNSQ와 Outfit MNSQ 값을 통해 확인한다. Boone, Staver & Yale(2014)은 MNSQ가 0.5~1.5에 있으면 적합한 수준이라고 하였다. 과학적 핵심역량의 5개요인 25개 문항 모두가 이 기준에 만족하는 것으로 나타났다(Table 3). Person reliability 값은 각 문항들이 학생들의 과학적 핵심역량을 구분하는데 적합한지 확인할 수 있는 수치로서, 0.7이나 또는 그 이상일 경우 적합하다고 할 수 있다(Boone, Staver & Yale, 2014). 이 연구에서는 가장 낮은 Person reliability 값이 0.81로 5개 요인의 문항들은 모두 적합한 Person reliability를 보여주었다. Item reliability 값은 연구에 참여한 학생들의 역량 수준이 문항의 적합도를 판단하는데 적절한지 확인하는 통계값이다. Item reliability 값의 경우 0.9이상 이면 적합하다고 할 수 있는데, 5개 항목 모두 1.0으로 적합한 것으로 확인되었다(Boone, Staver & Yale, 2014). 마지막으로 내적 일관성 신뢰도인 Cronbach alpha(CA)이다. 모든 요인에서 0.8 이상으로 적합한 수준의 내적 일관성 신뢰도를 보여주고 있다(George &

Mallery, 2003). 일반적인 문항 개발 과정에서는 혹시라도 부적합 문항이 있을지 모르기 때문에 예상 문항수보다 더 많은 수의 문항을 개발하고, 문항 적합도를 통해 부적합 문항을 제거하는 방법으로 최종 문항을 선별한다. 이 연구에서는 25개 문항을 개발하였고 학생들에게 투입한 결과 25개 모든 문항의 적합도와 신뢰도가 적합한 것으로 나타나 문항 제거 과정은 없었다.

다음으로 내적 구조에 기초한 타당도이다. 이 연구에서는 과학적 핵심역량을 2015 개정 과학과 교육과정에 근거하여 5개 요인으로 구분하였다. 이 구분이 학생들의 응답을 바탕으로 통계적으로 적합한 지에 대한 분석을 실시하였다. 분석은 구조방정식을 통한 확인적 요인분석을 실시하였다(Figure 1). 또한 대안 모델로서 과학적 사고력, 과학적 탐구능력, 과학적 문제해결력을 하나의 요인으로 하고, 과학적 의사소통과 과학적 참여와 평생학습을 두 번째 요인으로 한 대안 모델도 설정하였다. 대안 모델의 이론적 근거는 과학적 문제 해결의 인지적 요인(사고력, 탐구능력, 문제해결력)과 사회적 요인(의사소통, 참여)로 묶일 수 있음을 근거로 하였다. 그 결과 5개 요인으로 구분한 모델이 더 적합한 모델임이 확인되었다. 5개 요인으로 구분된 모델의 모델 적합도는 RMR이 0.027, GFI이 0.915, NFI이 0.945, IFI이 0.946, TLI이 0.939, CFI이 0.946, RMSEA이 0.060(90% CI: 0.059 - 0.061)로 구조방정식 모델 적합도의 일반적인 기준을 모두 만족하고 있었다. 2개 요인으로 구분된 모델은 대부분 만족스러웠으나 GFI가 0.9 이하

Table 3. Person and Item reliability and MNSQs of Rasch analyses.

과학 핵심역량	문항 번호	MEASURE	Infit MNSQ	Outfit MNSQ	PTMA	Person Reliability	Item Reliability	문항 삭제시 CA	CA
과학적 사고력	R_1	-0.42	0.918	0.907	0.813	0.83	1.00	0.835	0.867
	R_2	0.30	0.861	0.856	0.837			0.828	
	R_3	0.38	0.856	0.851	0.839			0.827	
	R_4	-0.34	0.927	0.920	0.811			0.835	
	R_5	0.08	1.400	1.403	0.728			0.870	
과학적 탐구능력	I_1	-0.79	1.277	1.290	0.779	0.86	1.00	0.893	0.897
	I_2	0.22	0.812	0.792	0.873			0.863	
	I_3	0.11	1.165	1.141	0.812			0.883	
	I_4	0.18	0.844	0.827	0.867			0.865	
	I_5	0.28	0.851	0.835	0.866			0.865	
과학적 문제해결력	P_1	0.09	0.898	0.884	0.838	0.84	1.00	0.845	0.879
	P_2	0.24	0.812	0.799	0.856			0.839	
	P_3	0.28	0.960	0.953	0.828			0.850	
	P_4	-0.20	1.182	1.199	0.779			0.869	
	P_5	-0.40	1.105	1.090	0.790			0.862	
과학적 의사소통 능력	C_1	-0.06	0.933	0.919	0.839	0.85	1.00	0.861	0.889
	C_2	-0.18	1.131	1.123	0.804			0.874	
	C_3	0.18	0.993	0.982	0.833			0.864	
	C_4	0.48	0.883	0.877	0.853			0.857	
	C_5	-0.43	1.016	1.004	0.819			0.867	
과학 참여와 평생학습 능력	J_1	0.71	1.039	1.049	0.793	0.81	1.00	0.827	0.854
	J_2	-0.27	0.923	0.912	0.799			0.818	
	J_3	-0.82	1.055	1.063	0.759			0.832	
	J_4	-0.04	0.961	0.953	0.797			0.820	
	J_5	0.42	0.982	0.982	0.799			0.823	

이며, RMSEA가 0.6이상으로 일부 항목에서 만족스럽지 못하였다 (RMR = 0.032, GFI = 0.885, NFI = 0.921, IFI = 0.922, TLI = 0.914, CFI = 0.922, RMSEA = 0.071).

5개 요인 모델과 2개 요인 모델의 적합도 차이를 보다 명확하게 확인하기 위하여 차원분석도 실시하였다. ConQuest 4.5.0을 활용하여 5개 요인으로 분석하였을 때와 2개 요인으로 분석하였을 때의 Akaike Information Criterion(AIC), Final Deviance, Chi-square의 값을 비교하는 방법을 사용하였다(Neumann, Neumann & Nehm, 2011). AIC값의 5개 요인 모델이 152606, 2개 요인 모델이 155065으로 그 차이가 -2460으로 5개 요인 모델이 더 적합한 것을 확인할 수 있다. Final Deviance 역시 5개 요인 모델이 152376, 2개 요인 모델이 154859으로 그 차이가 -2484로 나타나 5개 요인 모델이 더 적합하였다. 마지막으로 Chi-square 5개 요인 모델이 3790, 2개 요인 모델이 1312로 그 차이가 2478로서 5개 요인 모델이 더 적합하였다. 구조방정식을 활용한 확인적 요인분석과 차원 분석의 두 방법을 통해서 2015 개정 과학과 교육과정에서 제시된 5개 과학과 핵심역량의 각 요인들의 구분이 명확히 이루어지는 것을 확인할 수 있다.

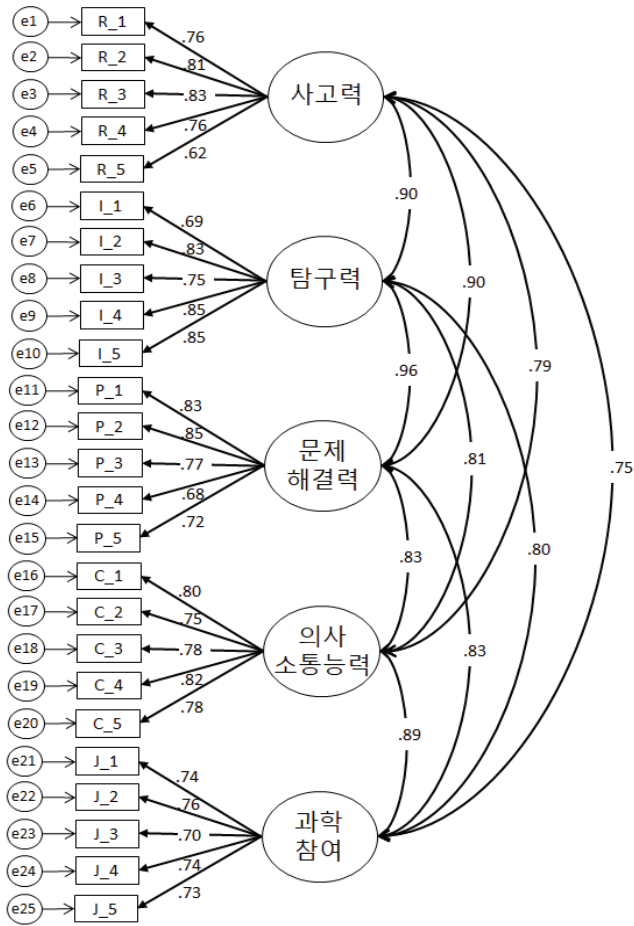


Figure 1. Confirmatory factor analysis using structural equation modeling

세 번째로 확인한 타당도는 일반화 타당도이다. 일반화 타당도는 검사도구의 기능이 응답자 그룹의 특성에 따라 변화하지 않고 일관성이 있는지를 확인하는 타당도로 차별적 문항반응을 통해 확인한다. 과학적 핵심역량에 관한 검사도구가 학교급 간, 성별 간 비교 연구에 활용될 수 있음을 가정하고, 차별적 문항반응은 학교급과 성별에 따라

수행하였다. 라쉬 모델 분석에서 DIF값의 차이가 0.64 이상일 경우 차별적 문항 기능이 나타난다고 할 수 있다(Boone, Staver & Yale, 2014).

성별로 분석한 경우 25개 문항 모두 유의미한 수준의 DIF 차이가 나타나지 않았다. 학교급 간에 비교한 경우 초등학교에서 이과 고등학교 학생까지의 4개 집단으로 짝지어진 총 6개의 집단 간 구분과 25개 문항으로 구성된 총 150개의 DIF 차이값 중 0.64이상은 과학적 탐구능력 첫 번째 문항의 초등학생과 문과 고등학생 사이에서, 초등학생과 이과 고등학생의 사이에서만 0.64 이상의 차이를 보여주었다 (Table 4). 차별적 문항기능 분석을 근거로 개발된 과학적 핵심역량 검사 도구는 학교급 간 및 성별 간 비교 분석 연구에서 활용하기에 적합한 도구임을 확인할 수 있었다.

이 연구에서는 2015 개정 과학과 핵심역량에 관한 설명을 근거로 개발된 과학과 핵심역량 문항의 내용 타당도, 실제에 기초한 타당도, 내적 구조에 기초한 타당도, 일반화 타당도의 4가지 타당도 수준을 확인하였다. 내용 타당도의 경우 수치화된 방법으로 확인하지 않았는데, 그 이유는 국내 과학교육 전문가들의 합의에 의하여 도출된 과학과 교육과정의 설명을 근거로 하였으며, 개발 과정에서 5명의 과학교육 전문가들 및 6명의 현장 과학교사들이 여러 번에 걸친 검토와 합의를 거쳤기 때문에 충분하다고 판단된다.

통계적 방법으로 검증한 실제에 기초한 타당도, 내적 구조에 기초한 타당도, 일반화 타당도 역시 만족스러운 것으로 확인되었다. 학생들의 응답과 라쉬 모델에 근거하여 추론된 이론적 값과의 차이점을 근거로 확인하는 문항 적합도(MNSQ)는 실제에 기초한 타당도를 확인하는 통계적 방법 중 하나이다. 문항 적합도 MNSQ에 대한 기준은 라쉬 모델 분석의 권위자인 Linacre 박사가 매우 많은 자료를 통해 시뮬레이션 한 뒤 확인하고, 많은 연구에서 일반적으로 통용되고 있는 기준인 MNSQ 0.5~1.5를 주로 활용한다(Boone, Staver & Yale, 2014). 또한 라쉬 모델 중 평정척도 모델 등을 개발한 Wright 박사와 Linacre 박사가 함께 제안한 기준인 0.6~1.4 역시 기준점으로 활용되고 있다(Wright & Linacre, 1994). 보다 엄격한 0.6~1.4 기준을 적용한다고 하더라도 과학적 사고력의 마지막 문항의 Outfit MNSQ 값인 1.403을 제외하고 모든 문항이 만족스러운 결과를 보이고 있다. 또한 1.403 역시 기준점과 의미 있는 차이가 없는 수준으로 학생들의 문항에 대한 반응은 적합한 수준에서 이루어진다고 할 수 있다.

내적 구조에 기초한 타당도는 개발된 검사도구의 타당도뿐만 아니라 2015 과학과 교육과정의 다섯 가지 역량 설정에 관한 타당도 역시 의미 있는 결과를 보여주었다. 과학적 사고, 과학적 탐구, 과학적 문제 해결은 많은 경우에서 독립되지 않은 역량으로 이해될 수 있다. 예를 들어 과학적 탐구를 수행하는 과정에서 과학적 사고력이 요구되며, 과학적 문제해결을 수행하는 과정에서 과학적 탐구 능력이 요구된다. 따라서 학생들은 이 세 구인에 대한 명확한 구분이 없을 수 있으므로 이론적으로는 세 구인이 구별될 수 있으나 학생들의 이해 수준에서는 구분되지 않을 수 있다. 하지만 이 연구에서 학생들의 응답에 기초하여 확인한 25개 문항의 내적 구조는 5가지 역량에 대한 구인들이 명확히 구분됨을 보여준다. 또한 이 결과는 2015 과학과 교육과정의 5가지 핵심역량의 구인 설정이 타당하고 적합하게 되었음을 보여주고 있다.

마지막으로 일반화 타당도이다. 과학 관련 정의적 태도 및 역량에 관한 많은 연구들에서 성별이나 학교급 같은 집단 간 비교는 빠지지

Table 4. The differential item functioning using Rasch model analysis.

과학 역량	문항 번호	학교급 간(1:초등학교, 2:중학교, 3:고등학교 문과, 4:고등학교 이과)					
		1-2	1-3	1-4	2-3	2-4	3-4
과학적 사고력	0.000	0.101	0.266	-0.043	0.165	-0.144	-0.309
	-0.170	-0.034	-0.250	0.240	-0.216	0.274	0.490
	-0.178	-0.024	-0.382	0.182	-0.357	0.207	0.564
	0.298	0.193	0.417	0.125	0.225	-0.068	-0.293
	0.083	-0.210	0.010	-0.482	0.220	-0.272	-0.492
과학적 탐구능력	0.171	-0.521	-0.735	-0.824	-0.214	-0.303	-0.089
	-0.153	0.175	0.109	0.149	-0.066	-0.026	0.040
	0.164	0.136	0.502	0.110	0.366	-0.026	-0.391
	-0.093	0.095	0.046	0.337	-0.050	0.242	0.291
	-0.086	0.118	0.092	0.230	-0.025	0.112	0.138
과학적 문제해결력	-0.229	-0.035	-0.412	-0.141	-0.378	-0.106	0.272
	0.000	-0.030	-0.253	-0.141	-0.223	-0.111	0.112
	-0.166	-0.056	-0.243	-0.056	-0.187	0.000	0.187
	0.222	0.070	0.375	0.202	0.305	0.131	-0.173
	0.173	0.030	0.451	0.084	0.420	0.054	-0.367
과학적 의사소통 능력	0.000	0.060	0.060	0.190	0.000	0.130	0.130
	-0.160	0.080	0.220	0.210	0.130	0.130	-0.010
	0.310	0.020	-0.090	0.030	-0.110	0.010	0.120
	-0.220	0.020	-0.290	-0.220	-0.310	-0.250	0.060
	0.090	-0.180	0.090	-0.210	0.270	-0.030	-0.300
과학 참여와 평생학습 능력	-0.340	0.120	0.000	0.000	-0.120	-0.120	0.000
	0.430	-0.080	0.060	0.010	0.150	0.100	-0.050
	0.510	-0.120	0.040	-0.200	0.160	-0.080	-0.240
	-0.300	0.000	-0.120	-0.080	-0.120	-0.090	0.030
	-0.210	0.060	0.010	0.230	-0.050	0.170	0.220

않는 대표적인 연구 문제이다. 집단 간 비교를 실시하기 위해서는 사용한 검사 도구에 대한 반응이 집단별로 동등하다는 것을 전제로 가능하다. 따라서 집단 간 비교 연구를 수행하기 위하여 확인되어야 하는 타당도가 일반화 타당도이다. 차별적 문항반응 분석은 자료수에 민감한 분석이지만(Lai, Teresi & Gershon, 2005; Scott et al., 2009), 이 연구에서는 10000명 이상의 참여로 다양한 집단 간 차별적 문항 반응 분석이 가능하였다. 성별 간 분석에서는 DIF의 일반적 기준인 0.64를 넘지 않으며, 학교급 간 비교는 초등학교와 고등학교의 비교에서 과학적 탐구 능력 문항 중 1번 문항만 기준에 위배된다(Boone, Staver & Yale, 2014). Logistic regression과 문항반응이론을 혼합한 Hybrid regression을 활용하여 DIF의 효과크기를 확인하면 0.7~0.8의 DIF차이는 작은 효과크기로 구분된다. 따라서 점수 비교에 영향을 줄 정도의 의미 있는 효과크기는 아니라고 판단된다. 문항 타당도에 관한 여러 문헌과 기준들을 근거로 판단하였을 때 개발된 25개 문항의 타당도는 정신측정학(Psychometrics) 관점에서 적합하다고 판단되며, 학교 현장에서 활용 가능성과 과학 교육 연구 등에서 활용가능성이 있다고 판단된다.

2. 성별과 학교급별 과학적 핵심역량의 점수 비교

이 연구에서 개발된 과학적 핵심역량을 측정하는 25개 문항의 타당도를 앞 절에서 확인하였다. 이 절에서는 성별과 학교급별로 과학

적 핵심역량을 비교 분석하여 개발된 검사도구가 집단 간 차이를 유의미하게 보여줄 수 있는지를 판단하고, 집단 간 과학적 핵심역량 비교 분석하고자 하였다.

Figure 2는 과학적 사고력, 과학적 탐구능력, 과학적 문제해결력, 과학적 의사소통 능력, 과학적 참여와 평생 학습 능력으로 구분하여 평균과 표준의 표준오차(오차막대)를 학교급과 성별로 제시한 것이다. Y축은 라쉬 모델 분석을 통해 생성된 Person measure 값이다.

전체적인 경향성을 보면 남학생이 여학생에 비하여 과학적 핵심역량이 유의미하게 높았으며, 그 차이는 과학적 의사소통과 과학적 참여와 평생학습 능력에서는 상대적으로 적게 나타났다. 모든 요인에서 문과 고등학생이 가장 낮은 수준이었으며, 이과 고등학생이 가장 높은 수준이었다.

보다 면밀하게 집단 간 차이를 확인하기 위하여 5개 요인을 모두 함께 분석한 다변량 이원분산분석과 각 요인별로 확인한 이원분산분석을 실시하였다. 먼저 다변량 이원분산분석에서 학교급 간, 성별 간은 모두 유의미하였다(학교급 간 $F[15, 30112]=21.409, p=0.000, PES=0.010$, 성별 $F[5, 10908]=24.984, p=0.000, PES=0.011$). 학교급 간 및 성별 간 교호작용효과(interaction effect)를 확인한 결과 유의미하였다(교호작용 효과: $F[15, 30112]=2.344, p=0.002, PES=0.001$). 각 요인별로 확인한 이원분산분석결과는 Table 5에 제시하였다. 5가지 과학적 핵심역량 요인에서 학교급 간 차이는 모두 유의미하였으며, 성별은 과학적 의사소통능력을 제외하고 모두 유의미하였다. 교호작용

용 효과는 과학적 탐구능력과 과학적 참여와 평생학습 능력을 제외하고 유의미한 것으로 나타났다.

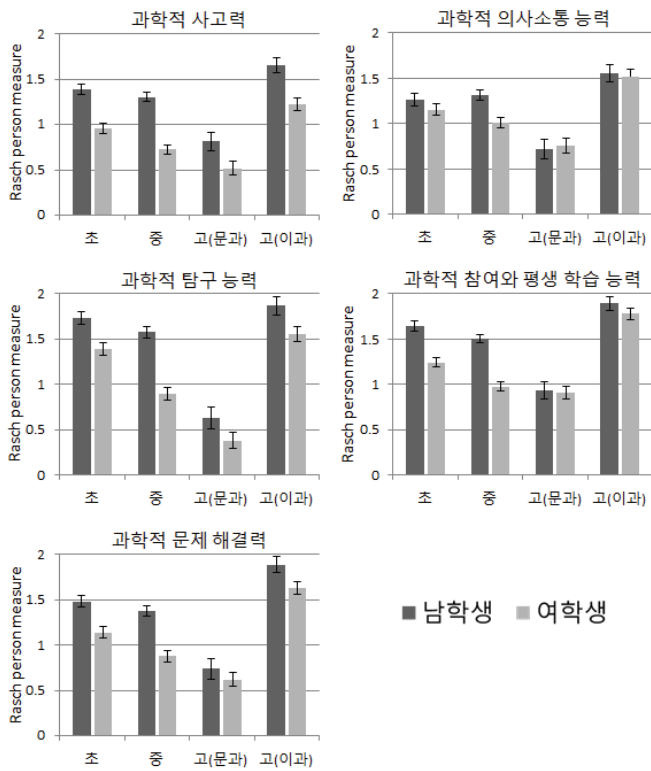


Figure 2. Comparison of scientific core competency score by gender and academic level

학교급 간 차이는 초등학교와 중학교의 차이는 거의 없으며, 문과와 이과 간 차이가 발생하는 것으로 보아 전공계열 선택이 학생 자신이 인식하는 과학 역량의 정도에 영향을 받을 수 있는 것으로 이해된다. 큰 의미 있는 차이가 나타나지 않는 학교급 간 차이에 비하여 매우 큰 차이를 나타내는 것이 성별 간 차이이다. 과학 역량에서 성별 간 차이점은 과학교육뿐만 아니라 많은 교육학 연구에서 관심을 가지고 있는 부분이다. 과학 관련 변인에서 남학생이 여학생에 비하여 높은 점수를 보인다는 것이 일반적이다.

국내 연구를 살펴보면, Choi, Shin & Lee(2008)은 수학과 과학 영역에 관한 국제 성취도 비교 연구인 TIMSS와 PISA 결과를 활용하여 우리나라 남학생의 과학 성취도와 태도 부분이 우리나라 여학생에 비하여 통계적으로 높고, 우리나라 성취도 수준의 성별 간 차이는

다른 국가에 비하여 월등히 높다고 분석하였다. Park & Shin(2010)은 소수의 남학생과 여학생을 질적으로 관찰한 결과 남학생들이 과학 지식 및 탐구 문제를 해결하는 과정에서 창안하는 전략이 여학생들에 비하여 더 많은 것을 확인하였다. 우리나라 고등학생들의 과학 학습 동기를 분석한 Shin, Lee & Ha(2017)의 연구에서도 남학생이 여학생에 비하여 과학 학습 동기가 월등히 높은 것을 확인하였다. Lee et al.(2017)은 공대생들의 공학 창의성을 분석한 결과 남학생이 여학생에 비하여 공학 창의성이 더 높은 것을 확인하였다. 국내에서 이루어진 주요 문헌들의 경향성을 살펴보면, 남학생이 과학에 대한 성취도나 태도가 높은 것은 일반적이라 할 수 있다.

또한 이와 같은 경향성이 해외 연구에서도 비슷한 양상을 보인다는 것도 남학생의 높은 과학 역량을 보여준다. 예를 들어서 Cavallo, Potter & Rozman(2004)은 미국 서부 대학의 290명의 대학생 (남학생 103명, 여학생 187명)을 대상으로 남학생이 자기 효능감, 성과 목표 및 물리에 대한 이해가 더 높은 것을 확인하였다. 또한 남학생이 여학생에 비하여 유의미한 학습을 더 많이 하는 것으로 보고하였다. Benbow & Stanley(1983)는 1980년까지 4학년의 7학년 학생의 수학적 추리력을 비교한 결과, 남학생이 더 높은 것을 확인하였다. Lee & Burkam(1996)은 미국 중학교생 25,000 명의 8학년 학생을 대상으로 한 1988년 National Education Longitudinal study의 결과를 분석하여 과학 성취도의 성별 차이를 확인한 결과, 남학생이 여학생보다 유의미하게 더 높은 수준을 보인다고 제시하였다.

본 연구에서 개발한 도구를 활용하여 학교급 간, 성별 간 비교 연구의 결과는 문항의 타당도와 활용이라는 점에서 몇 가지 의미를 부여한다. 먼저 본 연구에서 개발한 도구에서 매우 높은 수준에서 나타나는 성차는 개발된 검사도구가 일반적으로 알려진 내용과 비슷한 양상을 보여준다는 점에서 Messick(1995)이 강조한 외적 준거에 기초한 타당도가 일부 성립되는 것으로 이해할 수 있다. 두 번째 본 연구에서 개발한 과학 역량 검사도구로 학생들의 진로에서 문·이과 계열 선택의 성향에 대한 예측도구로서 가치가 있을 것으로 판단된다. 문과와 이과 학생들의 과학 역량의 수준은 확실하게 구분되는 것으로 나타났으므로 5가지 항목의 점수를 활용하여 학생들의 문·이과 성향을 판단하고 진로 지도에 활용할 수 있을 것이다.

3. 집단별 과학적 핵심역량 요인 사이의 상관관계 분석

과학적 핵심역량 간 상관관계는 Figure 1의 확인적 요인분석에서도 확인할 수 있다. 이 절에서는 8개 집단별로 과학적 핵심역량 점수

Table 5. Two-way ANOVA of scientific core competency score by gender and academic level.

과학 역량	학교급 간			성별			교호작용효과		
	F	p	PES	F	p	PES	F	p	PES
과학적 사고력	28.009	0.000	0.008	70.480	0.000	0.006	1.383	0.246	0.000
과학적 탐구능력	55.683	0.000	0.015	40.425	0.000	0.004	3.467	0.015	0.001
과학적 문제 해결력	48.505	0.000	0.013	30.818	0.000	0.003	2.352	0.070	0.001
과학적 의사소통 능력	22.478	0.000	0.006	3.332	0.068	0.000	2.072	0.102	0.001
과학적 참여와 평생 학습 능력	49.836	0.000	0.014	31.485	0.000	0.003	6.631	0.000	0.002

Table 6. Pearson correlation coefficients between five factors of scientific core competency score by gender and academic level.

역량별 상관관계	초등학교		중학교		고등학교(문과)		고등학교(이과)	
	남	여	남	여	남	여	남	여
사고력-탐구능력	.825	.757	.808	.788	.797	.765	.796	.755
사고력-문제 해결력	.798	.773	.811	.790	.777	.774	.802	.758
사고력-의사소통 능력	.722	.670	.695	.627	.782	.767	.740	.762
사고력-과학 참여	.676	.609	.652	.593	.702	.692	.738	.685
탐구능력-문제 해결력	.839	.827	.860	.838	.865	.835	.851	.817
탐구능력-의사소통 능력	.712	.677	.725	.652	.829	.785	.807	.768
탐구능력-과학 참여	.700	.656	.711	.640	.796	.722	.787	.706
문제 해결력-의사소통 능력	.750	.708	.752	.669	.842	.800	.799	.793
문제 해결력-과학 참여	.712	.675	.733	.644	.828	.785	.840	.747
의사소통 능력-과학 참여	.788	.772	.800	.788	.809	.768	.787	.725

의 상관관계를 확인하여 과학적 핵심역량에서 집단 간 구조적 차이가 있는지 판단하기 위하여 Pearson 상관관계 분석을 실시하였다. 5개 과학적 핵심역량 간 상관관계의 쌍은 총 10개이며, 8개 집단으로 구분되어 분석되었으므로 총 80개의 상관관계 계수가 생성되었다. 이 중에서 가장 낮은 계수는 중학교 여학생의 사고력과 과학 참여의 상관관계 계수인 0.593이며, 가장 높은 계수는 문과 고등학생의 탐구능력 과 문제 해결력의 상관관계 계수인 0.865이었다. 모든 상관관계는 0.01수준에서 유의미하였다(Table 6). 상관관계는 두 변인 간 관계의 강함 정도를 나타낸다고 가정하면 특정 집단에서 과학적 핵심역량의 상관관계가 상대적으로 높은 경우 역량 간 상호 관련성이 높다고 할 수 있다. 전체적인 경향성을 볼 때 남학생이 여학생에 비하여 상관관계의 평균값이 높고, 문과와 이과 구분 없이 고등학생들이 초등학생 과 중학생에 비하여 더 높은 것을 확인할 수 있다.

역량 간 상관관계가 높다는 것은 모든 역량들이 동시에 발달하는 것을 의미할 수 있다. 각 역량들이 독립적으로 발달할 경우 역량별 점수는 독립적이고 따라서 상관관계가 낮을 수 있다. 이와 같은 점을 적용한다면 남학생은 여학생에 비하여, 고등학생들이 초등 및 중학생들에 비하여 과학 역량에서 동시에 발달하는 특성을 보이는 것으로 추론할 수 있다. 아마도 남학생이나 고등학생들은 과학 역량의 발달 기회가 여학생과 저학년에 비하여 상대적으로 많기 때문에 이와 같은 경향성이 나타났다고 이해할 수도 있을 것이다. 과학 역량 간 인과적 모델에 대한 명확한 이론이 부족하기 때문에 본 연구 자료로 구조방정식 등의 다양한 방법으로 확인할 수 없는 한계가 있으나, 과학 역량 간 상관관계에 관한 자료는 과학 역량의 구조적 측면에 대한 분석의 시작으로 의미가 있을 것으로 판단된다.

IV. 결론 및 제언

이 연구는 2015 개정 과학과 교육과정 문서에 제시하고 있는 과학 과 핵심역량에 대한 설명을 근거로 과학과 핵심역량 검사 도구를 개발한 후 전국 초·중·고등학교 11,348명 이상의 학생들에게 적용하여 그 타당도와 신뢰도를 확인하였다. 이 연구의 결과와 논의들을 중심으로 결론을 내리면 다음과 같다.

첫째, 2015 개정 과학과 교육과정에 제시된 과학과 핵심역량 설명에 기초하여 25개 문항이 개발되었고, 5명의 과학교육연구자와 6명의 현장 과학 교사들에 의하여 내용 타당도가 확인되었다. 모든 문항은

과학과 교육과정에서 제시하는 과학과 역량을 충실히 반영하고, 학생들이 이해하고 응답하기 쉬운 문항의 형태로 개발되었다.

둘째, 개발된 문항의 실제에 기초한 타당도, 내적 구조에 기초한 타당도, 일반화 타당도는 정신측정학 관점에서 타당한 수준을 보이고 있다. 문항 적합도, 확인적 요인 분석, 차별적 문항 반응의 통계 수치는 기준에 모두 충분한 수준이었다. 또한 다양한 신뢰도 수치 역시 다양한 문헌에서 제시한 기준을 넘었다. 개발된 검사도구가 2015 과학과 교육과정의 핵심역량에 대한 설명을 기초로 개발되었다는 점에서 개발된 검사도구의 타당도는 2015 과학과 교육과정의 타당도라고 할 수 있다. 2015 과학과 교육과정에서 설명된 과학과 핵심역량에 대한 학생의 응답 반응과 내적 구조 등이 타당하다는 점은 해당 교육 과정을 바탕으로 교실 현장에서 목표 설정을 할 수 있음을 의미한다. 이런 점에서 개발된 문항의 타당도에 관한 결과들은 시사점이 크다.

셋째, 개발된 과학과 핵심역량 문항의 점수를 성별 간 및 학교급 간 비교한 결과 대부분의 영역에서 유의미한 차이를 나타내고 있다. 특히 성별에 따른 차이는 대부분의 핵심역량 영역과 학교급에서 나타나는 것으로 확인되었다. 이와 같은 결과는 과학 영역에서 남학생이 여학생에 비하여 더 높은 수준을 보이고 있다는 다양한 선행연구들과도 일치하는 결과이다.

넷째, 과학과 핵심역량의 영역 간 점수의 상관관계를 통하여 집단 별 과학과 핵심역량의 구조적 차이를 확인하였다. 남학생이 여학생에 비하여, 고등학생이 초등 및 중학생에 비하여 영역 간 점수의 상관관계가 높다는 것은 과학과 핵심역량의 구조가 더욱 견고하며, 영역 간 발달이 동시에 이루어졌다는 추론을 가능하게 한다. 과학과 핵심역량 간 인과적 모델에 관한 이론적 모델이 성립된다면, 이 부분에 대한 추가적 연구가 요구된다.

개발된 문항들은 교실 현장에서 학생들의 과학과 핵심역량 수준을 점검하고, 다양한 과학교육 연구에서 활용될 수 있을 것이라고 기대한다. 25개 문항을 응답하는 시간은 10분이 되지 않을 정도이기 때문에 교실 현장에서 충분히 활용 가능성이 높다.

이 검사도구를 활용하는데 있어서 주의사항을 강조하고자 한다. 행동 과학 영역에서 널리 활용되는 자기 보고 검사지(self-report questionnaire)의 한계점에 대한 다양한 논의들이 있어 왔다(Baars *et al.*, 2014; Kostons, Van Gog & Paas, 2012; Panadero, Tapia & Huertas, 2012). 물론 한계점을 인정하면서도 실용적인 측면에서 대안이 없기 때문에 자기 보고 검사지는 널리 활용되고 있다. 이 연구에서

개발한 검사지 역시 자기 보고형이다. 특히 자기 보고 검사지를 활용한 성차와 학교급 간 비교 연구는 일부 주의해서 해석해야 된다. 예를 들어 남학생이 여학생보다 자기 능력을 과대평가한다고 알려졌다 (Pajares & Graham, 1999; Pallier, 2003; Stankov & Lee, 2008). 또한 전문가들은 비전문가에 비하여 자기 능력을 과대평가한다는 결과들 역시 많은 연구에서 확인되었다(Angner, 2006; Menkhoff, Schmeling & Schmidt, 2013; Quirk, 2010; Chi, 2006). 이와 같은 자기 보고 검사지가 가지는 문제점은 모둠기반 수업에서 동료평가와 교사 평가의 결과들을 혼용하여 사용하고, 검사 결과를 종합하여 판단한다면 해결될 수 있을 것이다. 특히 동료평가의 경우 2015 교육과정에서도 강조하는 평가 방법이므로 이 연구에서 개발된 검사 도구를 바탕으로 동료평가를 실시한다면 의미 있는 평가 결과를 얻을 수 있을 것이다. 또한 학생들의 자기 보고형 평가와 동료평가 뿐만 아니라 학생들의 수업 활동에서 교사의 관찰에 의한 과정중심 평가에서 평가의 영역과 기준으로 활용한다면 학생들의 핵심역량에 대한 평가를 보완하는데 유용할 것임을 제안한다.

마지막으로 이 연구의 제한점을 기술한다. 평가 문항의 타당도에서 가장 중요한 것은 내용 타당도이며, 본 연구에서 제시한 다양한 통계치는 문항이 적절한지에 관한 판단을 위한 보조 자료이다. 따라서 본 연구에서 개발한 과학과 핵심역량 문항들이 과학과 핵심역량을 측정하는데 내용상 적합한지 판단하는데 있어 중요한 요소는 학자들의 의견이다. 예를 들어 과학적 사고력과 과학적 탐구능력, 과학적 문제해결력을 구성하는 요소들이 중복되는 경우가 있다. Table 2에 제시된 2015 과학과 교육과정의 설명문을 보더라도 과학적 탐구 능력이 과학적 사고력에 기초한다고 되어 있다. 장기적으로 과학과 핵심역량에 대한 명확한 정의, 각 역량별 구성 요소에 대한 이론적 고찰 등의 다양한 연구들이 지속적으로 이루어져야 할 것이다. 하지만 과학적 핵심역량에 대한 연구가 이루어지기까지 평가 문항 개발을 하지 않고 기다리는 것은 바람직한 것은 아니다. 2015 과학과 교육과정에서 많은 학자들이 논의 끝에 과학과 핵심역량 5가지를 정의했다. 이 연구는 그 정의를 바탕으로 평가 문항을 만들어서 많은 수의 다양한 학생들에게 적용하였다. 그 결과 다양한 통계값들이 문항의 타당도를 평가하는데 활용되는 기준에 근거하여 적합한 것으로 확인되었다. 단순히 통계 분석 결과가 적합하다고 이 연구에서 제시한 문항이 타당하다고 단언하는 것은 한계가 있다. 이 연구에서 제시된 문항들에 대한 비판적 고찰을 시작으로 과학과 핵심역량에 관한 문항 개발이 시작될 것으로 기대한다.

국문요약

이 연구는 2015 과학과 교육과정에 제시된 과학과 핵심역량에 관한 서술문을 기초로 학생들의 과학과 핵심역량을 측정할 수 있는 문항을 개발하고, 그 타당도와 신뢰도를 확인하기 위하여 진행되었다. 2015 과학과 교육과정에 제시된 과학적 사고력, 과학적 탐구 능력, 과학적 문제해결력, 과학적 의사소통 능력, 과학적 참여와 평생 학습 능력에 대한 설명을 근거로 각 역량별로 5개 문항씩, 총 25개 문항이 개발되었다. 개발된 문항은 전국 17개 시도에 고루 분포한 학생 참여형 과학수업 선도학교의 초등학교, 중학교, 고등학교 학생 11,348명을 대상으로 과학과 핵심역량을 조사하여 문항의 타당도와 신뢰도를 분

석하였다. 타당도는 Messick(1995)이 제안한 6가지 타당도 중 내용 타당도, 실제 타당도, 내적 구조 타당도, 일반화 타당도 등 4가지를 확인하였다. 라쉬 모델의 문항 적합도(MNSQ)를 분석한 결과, 25개 문항 중 부적합 문항은 없었으며, 구조방정식을 활용한 확인적 요인 분석 결과 5개 요인 모델은 적합한 모델로 확인되었다. 성별과 학교급 별에 따른 차별적 문항 반응을 분석한 결과, 부적합 값은 175개 중 2개에서 나타났다. 성별과 학교급별 다변량 이원분산분석을 실시한 결과, 학교급 간 및 성별 간에서 통계적으로 유의미한 차이가 나타났으며, 교호작용 효과 역시 유의미하였다. 결론적으로 2015 과학과 교육과정에 제시된 과학과 핵심역량의 설명에 기초하여 개발한 과학과 핵심역량 평가문항은 정성측정학적 관점에서 타당하며, 따라서 교육 현장에서 그 활용 가치가 있을 것으로 판단된다.

주제어 : 2015 개정 과학과 교육과정, 과학과 핵심역량, 핵심역량 조사 도구

References

- 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.
- Baek, P. (2013). A study on the relationships and characteristics between korean collegiate essential skills and undergraduates' individual variant factors. *Korean Journal of General Education*, 7(3), 349-387.
- Benbow, C. P., & Stanley, J. C. (1983). Sex differences in mathematical reasoning ability: More facts. *Science*, 222(4627), 1029-1031.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch Analysis in the Human Sciences*. New York, NY: Springer.
- Cavallo, A. M., Potter, W. H., & Rozman, M. (2004). Gender differences in learning constructs, shifts in learning constructs, and their relationship to course achievement in a structured inquiry, yearlong college physics course for life science majors. *School Science and Mathematics*, 104(6), 288-300.
- Chang, G., Seong, E., Choi, H., Jin, S., & Kim, G. (2015). *Measuring Youth Competency Index and International Comparative Study II(IEA ICCS 2016): General Report*. Sejong: National Youth Policy Institute.
- 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.
- Choi, K., Shin, D., & Lee, H. Y. (2008). Facts on gender difference in science education and solutional strategies. *Women's Studies Review*, 25, 117-158.
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2011). Investigating the substantive aspect of construct validity for the satisfaction with life scale adapted for children: A focus on cognitive processes. *Social Indicators Research*, 100(1), 37-60.
- George, D., & Mallery, M. (2003). *Using SPSS for Windows Step by Step: a Simple Guide and Reference*. Boston, MA: Allyn & Bacon
- Jin, M., Sohn, Y., & Chu, H. (2011). Study on development plan of K-CESA for college education assessment. *The Korean Journal of Educational Administration*, 29(4), 461-486.
- Kang, Y., Jung, D., Park, S., Lim, H., Hwang, J., & Seol, G. (2014). 2014 KEDI Student Competency Index Survey. KEDI Report RR 2014-26
- Kim, C., Eom, M., Kim, K., Kim, C., Park, J., Park, H., Lee, K., & Lee, S. (2013). A Study for Developing Student Competency Index. KEDI Report RR 2013-23
- Kim, D., Oh, H., Song, Y., Koh, E., Park, S., & Jung, E. (2009). Exploring students' core competencies in higher education from professors' perspectives: A case study of Seoul National University(SNU). *Asian Journal of Education*, 10(2), 195-214.
- Kim, E. (2014). *Readiness of university students in Korea for the core competencies in the 21st century*. Master Thesis. Ewha Womans University.
- Kim, H. & Lee, S. (2012). *Diagnosis and direction of the competence-based*

- educational model: Based on the course and effects. *Korean Journal of General Education*, 6(4), 11-40.
- Kim, H. & Lee, S. (2013). Application plans of competence assessment: Focusing on the connection between assessment and curriculum. *Korean Journal of General Education*, 7(4), 139-172.
- Kim, K., Kim, J., Chang, G., So, K., Kim, J., & Kang, Y. (2008). A Study on Development and Progress Plan of Lifetime Core Competency for Adolescent I : General Report. Seoul: National Youth Policy Institute.
- Kim, K., Maeng, Y., Chang, G., Ku, J., Kang, Y., & Cho, M. (2009). A Study on Development and Progress Plan of Lifetime Core Competency for Adolescent II: General Report. Seoul: National Youth Policy Institute.
- 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.
- Kwak, Y. (2013). Ways of restructuring key competences for a revision of science curriculum. *Journal of Korean Earth Science Society*. 34(4), 368-377.
- Lai, J. S., Teresi, J., & Gershon, R. (2005). Procedures for the analysis of differential item functioning (DIF) for small sample sizes. *Evaluation & the health professions*, 28(3), 283-294.
- Lee, G., Jeon, J., Huh, G., & Hong, W. (2009). A Study on the Vision of Elementary and Secondary School Curriculum for the Improvement of Core Competencies of Future Koreans: General Report RRC 2009-10. Seoul: Korea Institute for Curriculum and Evaluation.
- Lee, G., Kim, G., Kim, S., Kim, H., Lee, M., Lee, S., & Lee, I (2013). Exploring Curriculum for the Development of Core Competencies in the Future: Focusing on the Linkage of Curriculum, Teaching · Learning and Educational Evaluation. Research Report RRC 2013002. Seoul: Korea Institute for Curriculum and Evaluation.
- Lee, G., Kwon, H., Lee, J., & Lee, S. (2010). The study on proper way for engineering university education: Based on the perception of current competencies and expected competencies of engineering freshmen. *Journal of Engineering Education Research*, 13(6), 57-71.
- Lee, G., Min, Y., Jeon, J., & Kim, M. (2008). A Study on the Vision of Elementary and Secondary School Curriculum for the Improvement of Core Competencies of Future Koreans (II): Focusing on Sub-Element Setting by Core Competency Area. Seoul: Korea Institute for Curriculum and Evaluation.
- Lee, J. K., Shin, S., Rachmatullah, A., & Ha, M.(2017). Exploring the patterns of engineering college students' engineering-related creativity by gender, academic year, and engineering education accreditation program through latent class analysis. *Journal of Science Education*, 41(1), 16-35.
- Lee, S. & Lee, S. (2012). A new direction of undergraduate general education: Through a case study of competence-based education. *Korean Journal of General Education*, 6(2), 11-42.
- Lee, V. E., & Burkam, D. T. (1996). Gender differences in middle grade science achievement: Subject domain, ability level, and course emphasis. *Science Education*, 80(6), 613-650.
- 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.
- MOE (Ministry of Education) (2015). 2015 revised Science National Curriculum. Ministry of Education.
- 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.
- Noh, Y. & Lee, S. (2013a). The competency model development for Korean 4 year university students: A case study of D Women's University. *Journal of Corporate Education and Talent Research*, 15(1), 79-101.
- Noh, Y. & Lee, S. (2013b). A study on developing and validating the instrument for diagnosing the competency of university students: A case study of D Women's University. *The Korean Journal of Human Resource Development*, 15(3), 273-305.
- OECD (2003). Defining and selection of competencies: Theoretical and conceptual foundation. OECD Press.
- OECD (2006). Schooling for tomorrow: Thinking scenarios, rethinking education. OECD Press.
- 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.
- Park, B. T., Shin, D. (2010). Elementary school high achievers' gender characteristics of successful and unsuccessful process in problem solving. *Journal of Learner-Centered Curriculum and Instruction*, 10(2), 173-189.
- Quirk, P. J. (2010). The trouble with experts. *Critical Review*, 22(4), 449-465.
- Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., de Graeff, A., Groenvold, M., Gundy, C., Koller, M., Petersen, M. A., & Sprangers, M. A. (2009). A simulation study provided sample size guidance for differential item functioning (DIF) studies using short scales. *Journal of Clinical Epidemiology*, 62(3), 288-295.
- Seong, E., Baek, H., & Jin, S. (2014). Measuring Youth Competency Index and International Comparative Study I(IEA ICCS 2016): General Report. Sejong: National Youth Policy Institute.
- Shin, S., Lee, J. K., & Ha, M. (2017). Influence of career motivation on science learning in Korean high-school students. *EURASIA Journal of Mathematics Science and Technology Education*, 13(5), 1517-1538.
- So, K., Lee, S., Lee, J., & Heo, H. (2010). Review on curriculum reform in the New Zealand; Implementation of key competencies-based curriculum. *Comparative Education Society*, 20(2), 27-50.
- Son, Y., Song, C. & Lee, H. (2016). A Study on the Support for Core Competency of Korean Students in Collegiate Level. Sejong: Korea Research Institute for Vocational Education and Training.
- Stankov, L. & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology*, 100(4), 961.
- Wright, B.D. & Linacre, J.M. (1994). Reasonable mean-square-fit values. *Rasch Measurement Transactions*, 8(3), 370.
- Yoon, H., Kim, Y., Lee, G., & Jeon, J.(2007). A Study on the Vision of Elementary and Secondary School Curriculum for the Improvement of Core Competencies of Future Koreans (I): Focusing on Core Competency Criteria and Area Setting. Research Report RRC 2007-1. Seoul: Korea Institute for Curriculum and Evaluation.