

## ICILS 2013 컴퓨터·정보 소양 검사의 국가, 성별에 따른 차별기능문항 분석

이 문 수\*

한국교육과정평가원

본 연구의 목적은 ICILS 2013 검사에서 비슷한 성취 수준을 보이는 대한민국, 노르웨이, 폴란드 3국의 국가, 성별에 따른 차별기능문항을 탐색하고 원인을 분석하는 것이다. 모듈 A의 19개 문항을 대상으로 Wald-test를 이용하여 분석한 결과 남학생의 경우 9개 문항, 여학생의 경우에는 7개 문항이 노르웨이와 폴란드 두 유럽국가 학생에게 유리하였으며, 대한민국 남학생과 여학생에게는 1개 문항과 3개 문항이 유리하였다. 이 중 1번 문항은 국가 간 사용 언어에 따라 차별적으로 기능할 가능성이 높은 문항으로 차기 검사에서는 수정이 필요해 보인다. 성별에 따른 차별기능문항이 존재하는지 살펴본 결과, 대한민국은 15번 문항에서 여학생들의 성취가 높게 나타났으며, 언어 능력이 우수한 여학생들에게 유리하게 작용하였을 것으로 판단된다. ICILS 2013에서는 전체적으로 구성형 문항에서 차별기능문항이 많이 발견되었다.

주제어 : ICILS, ICT 리터러시, 다집단 차별기능문항, Wald-test

---

\* 교신저자 : 이문수, 한국교육과정평가원, [mslee9@kice.re.kr](mailto:mslee9@kice.re.kr)

## I. 서 론

4차 산업혁명을 주제로 다룬 ‘세계경제포럼 2016’에서는 ‘21세기 기술’이라는 16가지 미래 핵심 기술을 제안하였다. 비판적 사고, 창의성, 의사소통 등 복잡한 도전사항들에 대하여 어떻게 대처하는지 보는 핵심 역량(Competencies), 도전정신, 리더쉽, 적응력 등 변화하는 환경에 어떻게 적응하는지를 보는 인성 자질(Character Qualities)과 함께 문해력(literacy), 수해력(numeracy), 문화 및 시민문해(cultural and civic literacy) 등 일상생활에서 핵심 기술을 어떻게 적용시키는지 보는 기초문해(Foundational Literacies)가 있다(World Economic Forum, 2016). 기초문해 중 하나인 정보통신기술(Information and Communication Technology: ICT) 문해력은 컴퓨터, 스마트폰, 인터넷 등을 이용하여 문제해결에 필요한 정보를 탐색, 분석, 평가, 조직, 창출, 활용, 관리, 공유할 수 있는 능력을 의미한다. 4차 산업혁명 시대에는 혁신형 고부가가치 산업사회의 도래로 인하여 교수자에 의해 일방적으로 전달되는 지식이 아닌 학습자 스스로 또는 동료들과의 협업을 통해 정보를 탐색하여 가치를 창출할 수 있는 능력이 중요해진다. 따라서 학교는 물론 일상생활에서 ICT 기기를 이용하여 컴퓨터·정보 소양(Computer and Information Literacy; CIL) 또는 ICT 리터러시 능력을 함양하는 것이 비판적 사고력, 문제해결 능력을 갖춘 창의적 인재를 기르는 교육의 주요 과제가 되고 있다.

전 세계적으로 컴퓨팅 사고력을 기반으로 문제해결력을 기르고 창의적인 인재를 양성하기 위해 소프트웨어 교육을 필수 정규 교육과정으로 적극 도입하고 있으며, 우리나라도 2015 개정 교육과정에서 초등학교와 중학교에 소프트웨어 교육을 전면 실시하고 있다(교육부, 2015a, 교육부, 2015b). 구체적으로 초등 실과에 도입되는 소프트웨어 교육은 놀이 중심의 알고리즘 체험과 교육용 도구를 활용한 프로그래밍 체험을 통해 ICT 기기 및 각종 프로그램을 쉽고 재미있게 배울 수 있도록 하며, 2019년부터 17시간 이상 적용될 예정이다. 중학교에서는 실생활의 문제들을 컴퓨터 과학의 원리를 활용하여 해결하는 능력을 함양하도록 구성하여 2018년부터 34시간 이상 적용되고 있다(교육부, 2015a, 교육부, 2015b).

컴퓨터·정보 소양의 중요성이 점차 강조됨에 따라 국제적으로 학생들의 CIL 수준을 파악하기 위해 2010년 EC(European Commission)의 주도 아래 중학교 2학년(8학년)을 대상으로 하는 국제 컴퓨터·정보 소양 연구(International Computer and Information Literacy Study: ICILS)가 시작되었다(김수진 외, 2015). 수학, 과학 성취도 추이변화 국제비교연구(TIMSS)로 잘 알려진 국제 교육성취도 평가 협회(International Association for the Evaluation of Educational Achievement; IEA)에서 주관하는 ICILS는 중학생들의 컴퓨터·정보 소양 및 관련 교육맥락을 분석하여 국제적으로 비교하려는 목적에서 시작된 5년 주기 연구로 2013년에 첫 평가가 이루어졌다. 우리나라도 이러한 세계적인 관심과 ICT 관련 정부 정책의 변화에 따라 2011년부터 연구에

참여하였으며, ICILS 2013에서 유럽의 노르웨이, 폴란드 등과 비슷한 높은 성취를 보였다. 그러나 처음 실시되었던 ICILS 2013에서는 비슷한 수준의 능력을 가진 학생들의 정답률이 전혀 다르게 산출되기도 하였다. 이같은 차이가 피험자의 순수한 능력의 차이인지 아니면 검사에 차별기능문항(Differential Item functioning; 이하 DIF)이 존재하였는지에 대한 추가 분석이 필요한 상황이다. 이에 본 연구에서는 ICILS 2013에서 국가 및 성별에 따른 DIF가 존재하는지 탐색하고 그 원인을 분석하고자 하였다. ICILS의 대부분의 참가국이 유럽이었다는 점과 IT 첨단 기술의 강소국이 많은 유럽과 우리나라의 ICT 관련 교육의 비교를 위하여 대한민국, 노르웨이, 폴란드 3개국을 연구 대상 국가로 선정하였다. 구체적인 연구문제는 다음과 같다.

첫째, ICILS 2013 검사에서 대한민국, 노르웨이, 폴란드 3개국의 국가에 따른 차별기능문항은 어떠한가?

둘째, ICILS 2013 검사에서 대한민국, 노르웨이, 폴란드 3개국의 성별에 따른 차별기능문항은 어떠한가?

## II. 이론적 배경

### 1. ICT 리터러시 및 ICILS

전통적으로 리터러시는 읽기, 쓰기 등 문자 언어를 활용한 사회적 의사소통 능력을 의미하였으나 사회 환경의 변화 및 사회적 언어 도구(문자, 영상, 디지털 등)의 다양화로 인하여 그 의미가 점차 확대되고 있다(Gardiner, 1998). 리터러시와 관련하여 초기에는 단순한 문자 해독 그 자체에 집중하였으나, 여러 사회적 의미와 결합되면서 정치, 경제, 문화생활을 향유하기 위해 필요한 능력을 구비할 수 있는 생활 기능(life-skill)으로의 기능적 문자 해독의 의미가 점점 강조되고 있다(김영애, 서정희, 2011; Gardiner, 1998). 최근에는 정보통신기술이 발달함에 따라 정보를 수집, 생산, 가공, 전달, 활용하여 의사소통 할 수 있는 ICT 리터러시가 더욱 중요시되고 있다(김혜숙, 진성희, 2006).

국내 ICT 리터러시 연구는 2000년대 중반부터 한국교육학술정보원(KERIS)의 국가수준 초·중학생 ICT 리터러시 수준 측정 연구를 중심으로 꾸준히 수행되었다. 초기에는 ICT 리터러시 개념 정립 및 검사도구 개발 연구(김경성 외, 2009; 2010; 백순근 외, 2008; 서순식 외, 2008; 이원규 외, 2007)가 수행되었다. 2010년 이후에는 초등학생, 중학생을 대상으로 전국 단위 ICT 리터러시 수준 측정 연구(김경성 외, 2010; 김종민 외, 2017; 김현철 외, 2011;

안성훈 외, 2014; 2015)가 진행되어 컴퓨터·정보 소양 관련 다양한 연구 결과를 산출하고 있다.

대표적인 국외 ICT 리터러시 검사는 ETS 검사(2005), ACER 국가수준 검사(2010), PISA에 포함된 ICT 리터러시 평가(OECD, 2003), 그리고 IEA의 ICILS(김수진 외, 2014; 2015) 등이 있다. 이 중 ICILS는 전 세계 20여개 국가가 참여하고 있는 5년 주기 검사로, 지난 2013년에 이어 2018년 2주기 검사 시행을 앞두고 있다. ICILS 2013은 컴퓨터·정보 소양과 관련하여 크게 ‘정보 수집 및 관리’와 ‘정보 생산 및 교환’을 평가의 주요 요소로 보고 있으며, 4가지 주제의 모듈로 구성되어 있다(김수진 외, 2014; Fraillon, Schulz, & Ainley, 2013). ICILS 2013에서 대한민국은 536점으로 ICILS 평균 점수인 500점에 비해 높은 성취를 보였으며, 전체 14개 참여국<sup>1)</sup> 중 폴란드, 노르웨이 등에 이어 5위<sup>2)</sup>를 기록하였다(Fraillon et al., 2014).

대한민국 학생들의 CIL 점수를 남학생과 여학생으로 나누어 살펴보면, 여학생의 평균점수는 556점으로 체코(559점)에 이어 두 번째로 높았던 반면, 남학생은 517점으로 1위 국가인 체코의 남학생(548점)보다 31점이 낮았다. 이같은 수치는 우리나라와 유사한 성취를 보였던

〈표 1〉 ICILS 2013 컴퓨터·정보 소양 국제비교 결과

국가	CIL 평균 점수	ICILS 평균 기준	대한민국 평균 기준	점수 분포
체코	553	▲	▲	
호주	542	▲		
폴란드	537	▲		
노르웨이(9학년)	537	▲		
대한민국	536	▲		
독일	523	▲	▼	
슬로바키아	517	▲	▼	
러시아연방	516	▲	▼	
크로아티아	512	▲	▼	
슬로베니아	511	▲	▼	
리투아니아	494		▼	
칠레	487	▼	▼	
태국	373	▼	▼	
터키	361	▼	▼	

▲ : 기준보다 유의하게 높음    ▼ : 기준보다 유의하게 낮음

\* 출처 : 김수진 외, 2014, p.25; 김수진 외, 2015, p.9; Fraillon et al., 2014, pp.96

호주, 노르웨이, 폴란드와 같은 상위권 국가들보다 상대적으로 낮은 것이다. [그림 1]은 ICILS 2013 검사 참여율을 만족한 14개국의 여학생 CIL 점수에서 남학생의 점수를 차감한 수치를 보여준다. 모든 국가에서 여학생의 성취가 남학생보다 높게 나타났고 여학생과 남학생 점수 차이의 평균은 18점이었다. 대한민국은 여학생과 남학생의 CIL 평균점수 차이가 38점으로 유사한 성취를 보였던 호주(24점), 폴란드(13점), 노르웨이(23점)에 비해 차이가 가장 크게 나타났다. 이처럼 대부분의 국가에서 여학생의 성취가 높게 나타나는 현상에 대하여 단순히 여학생들의 ICT 관련 능력이 뛰어난 것인지 아니면 차별기능향 등의 다른 요인이 작용하는 것인지에 대하여 살펴봐야 할 필요성이 제기된다.



\* 출처: 김수진 외, 2014, p.33

(그림 1) 성별 CIL 평균점수 차이(여-남)

## 2. 대한민국, 노르웨이, 폴란드 ICT 교육

현재 전 세계 많은 국가에서 ICT 리터러시 교육에 많은 노력을 기울이고 있다. 우리나라에서는 2015 개정 정보과 교육과정에 단순히 컴퓨터를 사용하는 ICT 실습이 아니라, 문제 해결을 위한 아이디어와 알고리즘을 도출하고, 협력적 프로그래밍 활동을 통해 해법을 구현하며, 상호작용적 피드백 반응을 통한 산출물의 정교화 내용을 포함하였다(교육부, 2015c).

노르웨이의 교육 체제는 크게 유아, 초·중학교, 고등학교 및 고등 교육 4단계로 구성되어 있다. 국가 필수교육과정은 초등학교와 중학교에 해당하는 1~10학년까지, 고등학교 과정은 11~13학년까지이며 다른 인근의 유럽 국가와 마찬가지로 일반계 고등학교와 전문 고등학교로 구분하여 직업교육을 시행하고 있다(노르웨이 교육부, 2016). 노르웨이는 디지털 역량을 국가교육과정에 기본 역량으로 지정하고 있으며(김홍래, 이승진, 2013) 학생들의 실질적인 ICT 능력 향상을 위해 프로그래밍 언어에 대한 이해와 문제 해결에 초점을 두고 코딩, 게임 프로그래밍 및 3D 디자인 등의 내용으로 2016학년도부터 2년간 146개 학교를 대상으

로 ICT 교육을 실험적용하고 있다. 이와 더불어 교사 연수를 위한 다양한 교재 및 온라인 MOOC를 개발하여 프로그래밍에 대한 다양한 지원을 병행하고 있다(노르웨이 교육부, 2016).

폴란드의 교육 제도는 국가교육부와 과학 및 고등 교육부에 의해 중앙 관리되고 있으며, 현재 교육 체제는 1999년에 도입된 초등 6년, 중등 3년, 고등 3년, 대학의 구조이다. 초등 및 중등학교 운영에 대한 책임은 코핀에 위임되며, 교육 감독 책임은 16개도에서 지역 교육 당국의 장이 맡고 있다. 폴란드 교육과정에는 코딩 교육과 컴퓨터 프로그래밍이 통합되어 있으며, ICT와 관련하여 중학교에 “informatyka”라고 불리는 “컴퓨터 과학”이 주당 2시간씩 필수과목으로 배정되어 있다. 여기서는 다음의 6가지 주제를 다루고 있다. ① 인간 생활의 컴퓨터, ② 컴퓨터 활용, ③ 소프트웨어 사용(텍스트·그래픽 편집기, 스프레드시트, 데이터베이스), ④ 멀티미디어 활용, ⑤ 알고리즘, ⑥ 시뮬레이션 및 모델링(Gurbiel et al, 2005).

ICT 교육은 교육 프로그램의 구성이나 교사의 능력 등의 요소도 중요하지만 일반적으로 컴퓨터, 스마트폰 등의 매체와 인터넷 제반 환경이 일정 수준 이상 갖춰져야만 교육적 효과를 기대할 수 있다. 따라서 각국의 소득수준 및 ICT 인프라 구축 수준의 비교를 통하여 ICT 교육 여건을 확인할 수 있다. 대한민국, 노르웨이, 폴란드 3국의 ICT 인프라 및 경제적 특징을 살펴보면 <표 2>와 같다. 먼저, 주민 100명당 유선 초고속인터넷 신청 수를 살펴보면 대한민국이 37.2로 가장 높았고, 노르웨이가 우리나라와 유사하였으나 폴란드는 우리나라의 절반 이하 수준이었다. ICT에 대한 접근성, 이용도, 활용력 등을 종합평가하여 정보통신 발전 정도를 나타내는 ICT 개발 지수 점수는 우리나라가 8.57로 가장 높았으며, 노르웨이(8.13), 폴란드(6.31) 순이었다. 교육 분야에 대한 공공 지출은 노르웨이가 7.3%로 높았으며, 우리나라(5.0%)는 폴란드(5.1%)와 유사한 수준이었다(김수진 외, 2015). 이상의 지표들을 통해 우리나라의 ICT 인프라가 두 유럽 국가에 비해 결코 뒤떨어지지 않고 몇몇 분야에서는 앞서가고 있음을 알 수 있다.

〈표 2〉 ICT 인프라 및 경제적 특징

국가	주민 100명당 유선 초고속인터넷 신청 수※	ICT 개발 지수 점수※ (국가 순위)	1인당 GDP※※ (2005 ppp \$)	소득 지니계수※※	교육분야에 대한 공공지출※※ (% of GDP)
대한민국	37.2	8.57(1)	27,541	31.1	5.0
노르웨이	36.3	8.13(6)	46,982	25.8	7.3
폴란드	15.6	6.31(37)	18,087	34.1	5.1

\* 출처: 김수진 외, 2015, p.28; Fraillon et al., 2014, p.55, 재구성

※ ITU(International Telecommunications Union)에서 2012년에 조사한 자료(Fraillon et al., 2014, 재인용)

※※ Human Development Report 2013의 자료(Fraillon et al., 2014, 재인용)

다음으로, ICT와 관련한 학교 정책들을 살펴보면 폴란드가 ICT 장비, 소프트웨어 구입 및 취득, ICT를 위한 기술적 지원 등 많은 항목에서 학교 자율로 운영하고 있었고, 노르웨이는 학생 평가 부분에서만 자율적이었다. 우리나라는 학생 평가와 교사들의 ICT 활용 연수 참여 기회에서는 완전히 자율적이었지만, ICT 교육과정의 계획 및 실행은 자율적인 운영이 이루어지지 않고 있었다(김수진 외, 2015).

〈표 3〉 학교 정책에 대한 학교의 자율성

국가	학교 자치 (예. 학교 운영 위원회 등)	교사 채용	학생 평가	ICT 장비와 소프트웨어 구입/취득	교사의 ICT 활용 연수 참여규정	ICT 교육 과정의 계획 및 실행	ICT 기술적 지원
대한민국	●	●	●	●	●	○	●
노르웨이	●	●	●	●	●	●	●
폴란드	●	●	●	●	●	●	●

● 완전히 자율적임, ● 부분적으로 자율적임, ○ 자율적이지 않음

\* 출처: 김수진 외, 2015, p.28; Fraillon et al., 2014, p.53, 재구성

학년별 ICT 관련 교과와 필수 또는 선택 여부를 살펴보면 <표 4>와 같다. 먼저, 폴란드는 초등과 중등에서 ICT 관련 교과가 모두 필수 교과인 반면 노르웨이는 중등 후기에만 선택 교과로 지정되어 있었다. 우리나라는 중등 전기와 후기 선택 과목으로 지정되어 있었으나, 2015 개정 교육과정이 적용되는 2018년부터는 중학교에서 정보 교과가 필수 과목으로

〈표 4〉 학년별 ICT 관련 교과 및 ICT 평가 정책

국가	ICT 관련 교과				ICT 능력 평가 및 모니터링과 관련된 학교 수준의 요구	ICT 학생 평가(교육부)				
	중 초 등	중 등 전 기	중 등 후 기	교과명 (중등전기)		진 단 평 가	형 성 평 가	총 괄 평 가	국가(주) 질 관리 프로그램	디지털 학습 산출물(예, e-portfolio)
대한민국	▲	▲		정보	○	○	●	●	○	○
노르웨이			▲		○	●	●	●	●	●
폴란드	◆	◆	◆	컴퓨터 과학	●	○	○	○	●	●

◆ 필수 교과, ▲ 선택 교과, ● ICT의 활용 정책 / 계획에 관한 내용 있음, ○ ICT의 활용 정책 / 계획에 관한 내용 없음

\* 출처: 김수진 외, 2015, p.29; Fraillon et al., 2014, p.62 재구성

전환되었다(교육부, 미래창조과학부, 2015.7.21).

마지막으로 대한민국, 노르웨이, 폴란드 교사의 ICT 활용 능력 개발 지원 정책 현황은 <표 5>와 같다. 예비교사를 대상으로 ICT 활용에 관한 기술적 능력의 개발 및 교수학습에 ICT 활용하는 것에 대한 정책적 지원은 3개국 모두에서 이뤄지고 있었으며, 폴란드의 경우 현직교사에 대한 지원도 지속되고 있었다. ICT를 활용한 협력 또는 의사소통 능력의 개발과 학생 평가 시 ICT의 활용에 있어 노르웨이는 예비교사 교육 단계, 대한민국과 폴란드는 현직교사 교육, 연수에서 지원하는 것으로 나타났다(김수진 외, 2015).

<표 5> 교사의 ICT 활용 능력 개발을 위한 지원 정책

국가	ICT 활용에 관한 기술적 능력		교수학습에 ICT의 활용		ICT를 활용한 협력 또는 의사소통		학생 평가에서 ICT의 활용	
대한민국	●	■	●	-	-	■	-	■
노르웨이	●	-	●	-	●	-	●	-
폴란드	●	■	●	■	-	■	-	■

● 예비교사 교육에 대한 지원, ■ 현직교사 교육 또는 연수에 대한 지원

\* 출처: 김수진 외, 2015, p.30; Fraillon et al., 2014, p.65 재구성

### 3. 차별기능문항 개념 및 분석방법

차별기능문항은 동일한 수준의 잠재적 능력을 가진 피험자들이 집단 특성 때문에 정답 확률이 달라지는 것을 말한다(Clauser & Mazor, 1998; Holland & Wainer, 2012). 특정집단에 유리 혹은 불리하게 작용하는 검사 문항의 편파성, 공정성과 관련한 DIF는 국외에서는 주로 성별, 인종, 사회·경제적 배경(SES)과 관련하여 연구가 수행되어 왔고, 국내에서는 주로 성별에 따른 DIF 연구(김종민 외, 2016; 성태제, 1994; 송미영, 2001; 이대용, 김석우, 길임주, 2014; 이정, 이순목, 2009)가 진행되었다.

DIF 연구에서 주로 다뤄진 분석 방법은 Mantel-Haenszel(MH), SIBTEST, MIMIC 모형, 로지스틱 회귀 및 IRT Likelihood Ratio(IRT-LR), Wald-test 방법 등으로 지금까지 집단이 3개 이상인 상황을 탐색한 연구는 많지 않았다. Penfield(2001)는 1개의 참조(reference) 집단과 4개의 연구(focal) 집단을 대상으로 GMH(Generalized MH) 및 MH 방법의 검정력을 모의실험 연구를 통해 살펴보고, 연구 결과 GMH가 대부분의 조건에서 작은 type I 오차를 산출하였으나 모든 집단에 DIF가 존재할 때는 상대적으로 낮은 검정력을 보였다. Magis 등(2011)은 프랑스어 사



용 지역에 거주하는 캐나다 학생들의 영어 시험에서 DIF가 존재하는지를 평가하기 위해 4개 집단을 대상으로 분석을 진행하였고 GLR(Generalized LR) 방법과 특정 문항에 대해서는 GMH와 Lord's chi-square 테스트를 함께 적용하였다. 이 연구에서는 GLR이 다른 방법과는 달리 참조집단을 선택할 필요가 없는 장점이 있다고 보고하였다. Asparouhov와 Muthén(2014)은 측정 불변성(집단간 요인 부하량의 동등성)과 스칼라 불변성(집단간 요인 절편의 동등성)을 이용하여 alignment procedure 방법의 검정력을 조사하였다. 모의실험을 진행한 결과, 매개 변수 추정 편차가 낮고 표본이 큰 경우 안정적인 결과를 산출하였으며, 표본 수가 작은 경우에도 편차의 정도가 크지 않으면 편향이 상대적으로 작게 나타난다고 보고하였다.

본 연구에서는 Wald-test 방법을 이용하여 CIL 점수가 비슷한 대한민국, 노르웨이, 폴란드 3국을 대상으로 검사에 나타나는 차별기능항(국가 DIF 및 성별 DIF)을 탐색하고자 한다. 이 연구에서 사용한 Wald(1943)  $\chi^2$ -test는 두 집단 간의 성취 차이를 검증하기 위해 Wald 통계량을 이용한다. 즉, DIF 분석을 위해 참조집단과 연구집단의 문항 모수들을 각 집단의 표준오차 또는 분산-공분산 행렬로 나눈 통계량을 비교하는 방법이다(김종민 외, 2016).

$$\chi_i^2 = \mathbf{v}_i^T \boldsymbol{\Sigma}_i^{-1} \mathbf{v}_i \quad (1)$$

$$\mathbf{v}_i^T = [\hat{a}_{Fi} - \hat{a}_{Ri}, \hat{b}_{Fi} - \hat{b}_{Ri}],$$

$\hat{a}_{Fi}, \hat{a}_{Ri}, \hat{b}_{Fi}, \hat{b}_{Ri}$  : 연구집단, 참조집단의 문항 변별도/난이도 추정치

$\boldsymbol{\Sigma}_i$  : 두 집단 문항모수 차이의 분산-공분산 행렬

DIF 연구 초기에는 분산-공분산 행렬 계산상의 어려움 때문에 Wald-test 방법이 분석에 많이 활용되지 못했으나 근래 EM 알고리즘의 효율적인 개선(Cai, 2008) 및 가교 문항을 활용한 Wald-1 test 방법(Cai, Thissen, & du Toit., 2011), 가교 문항을 이용하지 않는 Wald-2 test 방법(Langer, 2008), IRTPRO(Cai, Thissen, & du Toit., 2011), flexMIRT(Cai, 2013) 등 새로운 IRT 프로그램의 개발로 말미암아 Wald-test를 이용한 DIF 연구가 가능해졌다(김종민 외, 2016). 먼저, Wald-2 test에서는 가교 문항 없이 두 집단의 문항 모수를 같게 설정하고 연구집단의 평균과 표준편차를 추정한다. 이후 식 (1)에 따라 Wald-test 값을 산출한 후,  $\chi^2$  분포에 의해 DIF 문항을 판별한다(김종민 외, 2016). 다음으로 Wald-1 test에 가교문항을 설정한 후, 연구집단의 평균, 표준편차는 문항모수들과 함께 추정하고, 참조집단의 평균, 표준편차는 0, 1로 고정시킨다. 이때, 두 집단의 가교 문항 모수들은 동일하게 제약하고, 다른 문항들의 모수를 추정한다. 다집단 DIF 분석에서 모형의 적합도 문제로 긴 분석 시간이 요구되는 IRT-LR 방법에 비

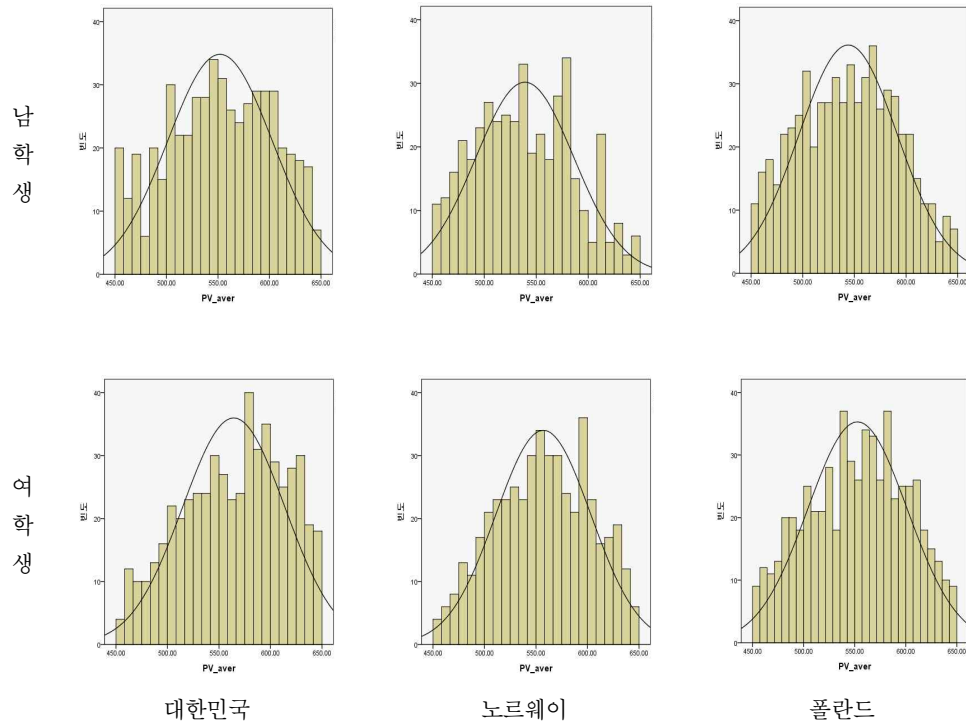
해 Wald-1, Wald-2 test의 연계 알고리즘을 이용하면 효율적인 분석이 가능하다(Woods, Cai, & Wang, 2012). 또한 Wald-test에서 가교문항 설정의 정제(purification)과정을 거치면 제1종 오류를 감소시킬 수 있는 장점이 있다(Kim & Cohen, 1995; Lord, 1980). 본 연구에서는 이러한 분석상의 장점을 이용하고자 Wald-test 방법을 이용하여 다집단 DIF 분석을 진행하였다.

### III. 연구방법

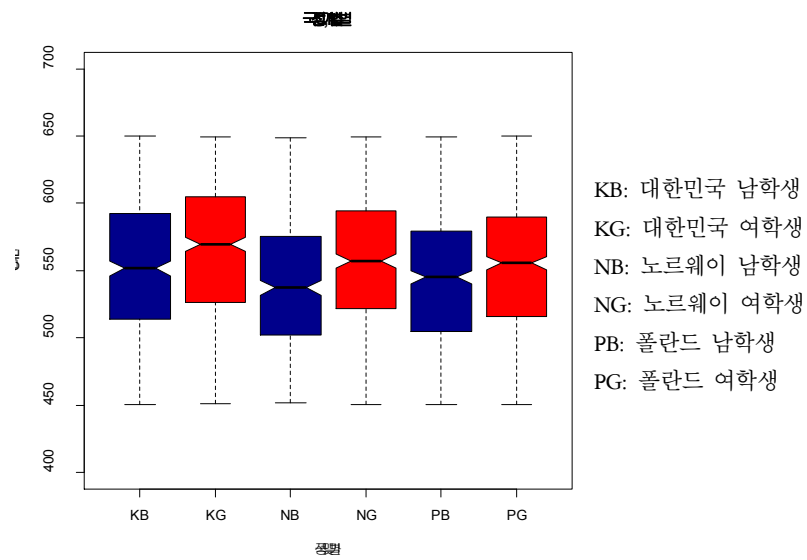
#### 1. 연구 대상

본 연구에서는 ICILS 2013에서 수집된 대한민국, 노르웨이, 폴란드 중학생의 성취 자료를 분석하였다. 분석에 사용된 모듈 A에 응답한 학생들은 대한민국 남학생 743명(52.5%), 여학생 673명(47.5%), 노르웨이 남학생 506명(49.3%), 여학생 521명(50.7%), 폴란드 남학생 734명(50.9%), 여학생 707명(49.1%)이었다. 모듈 A의 성적 분포를 평균, 표준편차 순으로 살펴보면 대한민국 남학생(520.41, 92.30), 여학생(560.71, 76.92), 노르웨이 남학생(522.82, 68.53), 여학생(551.71, 61.13), 폴란드 남학생(551.71, 61.13), 여학생(539.27, 71.85)으로 우리나라 남학생과 여학생의 점수 차이는 40여점으로 나타나 노르웨이(29점), 폴란드(12점)보다 격차가 컸다.

본 연구에서는 양 극단에 위치한 학생들로 인한 오차를 줄이기 위해 능력 추정치(plausible value) 평균 450점 이상 650점 이하의 학생들만 분석에 포함하였다. 이같은 조치를 취한 것은 ICILS가 고부담 검사(high-stake tests)가 아닌 관계로 불성실하게 응답한 것으로 보이는 응답자의 비율이 높았기 때문에(특히 대한민국 남학생의 최저점 비율이 가장 높음) 결과의 왜곡을 줄이고자 하는 방편이었다. 극단의 이상치(outlier)를 처리하는 일반적인 방법으로 이상치 제외 방법(trimming)과 이상치 값을 대체하거나 가중치를 조정하는 윈저화 방법(winsorization), 그리고 M-estimation 등의 로버스트 기법을 적용하여 이상치의 값을 추정하는 방법(Chambers et al, 2000, Kokic & Bell, 1994) 등이 있다. 본 연구에서는 표본수가 충분하다는 점에서 대체(imputation) 등의 방법을 사용하지는 않았다. 이에 최종적으로 분석에 사용된 인원은 대한민국 남학생 532명(49.8%), 여학생 537명(50.2%), 노르웨이 남학생 429명(47.6%), 여학생 472명(52.4%), 폴란드 남학생 605명(50.1%), 여학생 602명(49.9%)으로 최초 표본의 75~87% 수준이었다. 이들의 평균 및 표준편차는 각각 대한민국 남학생(551.95, 50.80), 여학생(564.31, 49.65), 노르웨이 남학생(539.04, 47.27), 여학생(557.13, 46.15), 폴란드 남학생(543.95, 47.71), 여학생(552.70, 48.36)이었으며, 각 집단별 학생들의 분포는 [그림 2] 및 [그림 3]과 같다.



(그림 2) 3개국 남·여학생의 점수분포 히스토그램



(그림 3) 3개국 남·여학생의 점수분포 상자그림

## 2. 측정 도구

ICILS 2013 평가들은 컴퓨터·정보 소양과 관련된 지식의 틀을 구성하는 개념적 범주인 주요영역(Strand)과 주요영역에 따른 특정 내용 범주인 하위요소(Aspect)로 구성되어 있다(Fraillon et al., 2013). 주요영역은 다시 ‘정보 수집 및 관리(주요영역 1)’와 ‘정보 생산 및 교환(주요영역 2)’ 두 가지로 나눌 수 있으며, 주요영역 1에서는 컴퓨터 사용과 관련한 학생들의 정보 수집 및 관리 능력을 측정하고, 주요영역 2는 사고, 창조, 의사소통 도구로서 컴퓨터 사용 능력을 측정하기 위한 것이다(Fraillon et al., 2013). 평가들에 따른 점수 비율에서는 주요영역 1, 2에 각각 33%, 67%가 할당되어 있다(Fraillon et al., 2014).

ICILS 2013 검사 도구는 실생활과 관련한 하나의 주제를 중심으로 몇 가지 과제들을 연쇄적으로 해결하는 모듈 형태로 구성되어 있다. ICILS 2013는 방과 후 운동(After-School Exercise), 밴드 경연대회(Band Competition), 호흡(Breathing), 현장학습(School Trip) 4개의 모듈로 구성되어 있으며, 각 모듈은 5~8개 일반과제 문항과 1개의 종합과제로 구성되어 있다. 문항의 형식은 선다형, 구성형 문항이 함께 있으며, 모듈의 마지막은 종합과제로 구성되어 있다(김수진 외, 2014). 이번 연구에서는 4가지 모듈 중 일반에게 문항 및 관련 정보가 공개된 ‘방과 후 운동’ (모듈 A)를 분석 대상으로 하였다. 모듈 A의 경우 선다형 14문항, 구성형 5문항 총 19문항으로 구성되었다. 1번 문항부터 10번 문항까지는 일반 개별 문항이고, 11번부터 19번은 종합과제 문항으로 학생들은 특정 목적에 맞는 산출물을 만들어야 한다. 종합과제는 9가지 평가 기준<sup>3)</sup>에 따라 채점하게 되며, 이번 연구에서는 9가지 채점 기준에 따른 채점 결과를 각각 하나의 문항으로 간주하여 분석을 하였다.

## 3. 분석방법

본 연구에서는 대한민국, 그리고 남학생을 참조집단으로 노르웨이, 폴란드, 여학생을 연구집단으로 선정하여 균일적(uniform) DIF 분석을 진행하였다. 먼저, ICILS 2013 모듈 A가 일차원성 가정을 충족시키는지 확인한 후 다집단 DIF 분석을 시작하였다. 본 연구에서는 국가에 따른 분석과 성별에 따른 분석을 각각 분리하여 주효과(main effects)만을 다루었다. 3개 국가에 대한 DIF 분석에서는 대한민국과 유럽 상위 2개국(노르웨이, 폴란드)의 평균비교 및 노르웨이와 폴란드의 비교를 위하여(-2 1 1, 0 -1 1) 형태의 복합대비(complex contrast)를 적용하여 국가 간 차이를 분석하였다.

앞에서 언급하였듯이 차별기능문항 검증을 위하여 Wald-test 방법을 이용하였다. 최근 Wald-test와 관련한 연구에서 Wald-2 test는 제 1종 오류를 증가시키는 경향이 있기 때문에

Wald-2 test를 차별기능문항 최종 판정 단계보다는 중간 과정인 가교문항 선정에 사용하는 것이 유용하다는 결과를 발표되었다(Woods, Cai, & Wang, 2012). 본 연구에서도 먼저 모듈 A 전체 19 문항에 대하여 Wald-2 test를 통하여 가교문항(차별기능문항이 없는 문항)을 선정하였고, 이들 문항의 문항모수를 고정한 후 나머지 문항들에 대하여 Wald-1 test를 진행하였다. 차별기능문항의 전체적인 분석은 flexMIRT(Cai, 2013) 프로그램을 사용하였다.

## IV. 연구 결과

### 1. 국가 DIF 분석 : Wald-2 test 결과

대한민국, 노르웨이, 폴란드 3국 간 ICILS 2013 모듈 A에 따른 차별기능문항 검증을 위한 Wald-2 test 결과는 <표 6>과 같다. Wald-2 test에서는 유의수준( $p$ -value)이 .01보다 작은 문항들을 잠재적인 DIF 문항으로 간주하였고, 그 이외의 문항들을 DIF가 없는 Wald-1 test의 가교문항으로 분류하였다. 3국 간의 분류에서는 2개의 대비(contrast)식 중 1개라도 유의한 차이가 발견되면 차별기능문항이 존재하는 것으로 보았다.

분석결과 남학생 집단에서는 1번, 3번, 5번, 6번, 7번, 10번, 12번~19번 문항에서 대한민국, 노르웨이, 폴란드 국가에 따른 차이가 있는 것으로 나타났다. 이 중 1번, 5번, 10번, 12번~14번, 16번, 18번, 19번 문항은 2개의 대비식에서 모두 DIF가 존재하는 것으로 분류되었다. 여학생 집단의 국가 간 비교에서도 남학생의 경우와 마찬가지로 1번, 3번, 5번, 6번, 7번, 10번, 12번~19번 문항이 국가 간 차이가 발견되었으며, 1번, 5번, 7번, 12번, 13번, 15번, 16번, 18번, 19번 문항은 2개의 대비식에서 모두 유의한 차이를 보였다. 산출된 결과를 통하여, 3개국 간 비교에서 차별기능문항으로 검출된 문항이 전체적으로 남학생, 여학생 두 집단에서 매우 유사함을 확인할 수 있다. 문항에 따른 세부적인 차별기능문항에 대한 분석은 Wald-1 test 결과를 이용하여 진행하였다.

### 2. 국가 DIF 분석 : Wald-1 test 결과

Wald-2 test를 통해 잠재적인 DIF 문항으로 선택된 문항을 제외한 나머지 문항들(가교문항)을 대상으로 Wald-1 test를 진행하였다. 이 과정을 통하여 최종적으로 3개국 간 DIF가 존재하는 것으로 판별된 문항들은 <표 7>에 제시하였다.

먼저, 남학생의 경우 1번, 3번, 5번, 6번, 10번, 12번, 15번, 16번, 17번, 19번 총 10개 문항

〈표 6〉 국가 DIF 분석 : Wald-2 test 결과(표음영처리 수정)

문항번호	대비	남학생		문항번호	대비	여학생	
		$\chi^2 (df)$	p-value			$\chi^2 (df)$	p-value
1	1	73.0 (2)	0.000***	1	1	100.1 (2)	0.000***
	2	32.5 (2)	0.000***		2	21.0 (2)	0.000***
2	1	7.7 (2)	0.022	2	1	4.2 (2)	0.122
	2	2.7 (2)	0.259		2	1.5 (2)	0.482
3	1	23.5 (2)	0.000***	3	1	27.1 (2)	0.000***
	2	0.3 (2)	0.862		2	6.5 (2)	0.039
4	1	1.7 (2)	0.433	4	1	0.1 (2)	0.970
	2	5.3 (2)	0.072		2	9.1 (2)	0.011
5	1	27.8 (2)	0.000***	5	1	9.7 (2)	0.008**
	2	28.8 (2)	0.000***		2	18.0 (2)	0.000***
6	1	49.8 (2)	0.000***	6	1	30.0 (2)	0.000***
	2	5.1 (2)	0.078		2	4.9 (2)	0.085
7	1	15.7 (2)	0.000***	7	1	16.3 (2)	0.000***
	2	2.6 (2)	0.270		2	11.3 (2)	0.004**
8	1	4.2 (2)	0.125	8	1	4.5 (2)	0.105
	2	2.9 (2)	0.240		2	3.7 (2)	0.160
9	1	0.3 (2)	0.866	9	1	1.6 (2)	0.447
	2	0.7 (2)	0.702		2	0.9 (2)	0.638
10	1	23.3 (2)	0.000***	10	1	4.0 (2)	0.134
	2	43.1 (2)	0.000***		2	47.3 (2)	0.000***
11	1	0.0 (2)	0.984	11	1	4.9 (2)	0.085
	2	8.6 (2)	0.014		2	2.4 (2)	0.309
12	1	20.1 (2)	0.000***	12	1	21.4 (2)	0.000***
	2	26.9 (2)	0.000***		2	30.7 (2)	0.000***
13	1	176.9 (2)	0.000***	13	1	115.8 (2)	0.000***
	2	44.6 (2)	0.000***		2	67.3 (2)	0.000***
14	1	18.1 (2)	0.000***	14	1	25.8 (2)	0.000***
	2	12.2 (2)	0.002**		2	6.8 (2)	0.033
15	1	122.5 (3)	0.000***	15	1	68.2 (3)	0.000***
	2	8.2 (3)	0.043		2	12.7 (3)	0.005**
16	1	98.6 (3)	0.000***	16	1	58.0 (3)	0.000***
	2	67.9 (3)	0.000***		2	43.3 (3)	0.000***
17	1	24.7 (3)	0.000***	17	1	59.0 (3)	0.000***
	2	2.7 (3)	0.442		2	2.8 (3)	0.425
18	1	28.0 (3)	0.000***	18	1	32.2 (3)	0.000***
	2	53.8 (3)	0.000***		2	60.2 (3)	0.000***
19	1	59.7 (3)	0.000***	19	1	72.6 (3)	0.000***
	2	31.1 (3)	0.000***		2	25.9 (3)	0.000***

\*\* $p < .01$ , \*\*\* $p < .001$

〈표 7〉 국가 DIF 분석 : Wald-1 test 결과

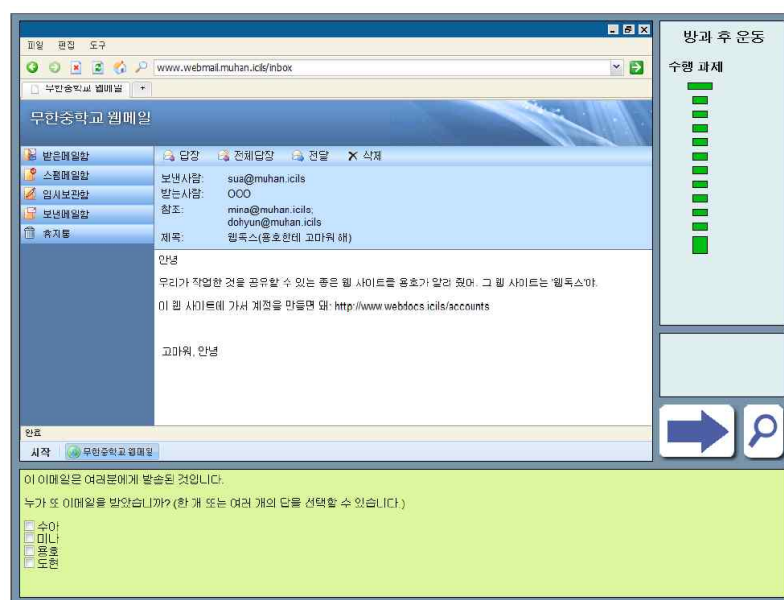
문항 번호	대비	남학생			문항 번호	대비	여학생		
		$\chi^2 (df)$	<i>p</i> - value	선호집단			$\chi^2 (df)$	<i>p</i> - value	선호집단
1	1	62.9 (2)	0.000***	N, P	1	1	63.1 (2)	0.000***	N, P
	2	29.4 (2)	0.000***	N		2	17.9 (2)	0.000***	N
3	1	17.7 (2)	0.000***	N, P	3	1	21.4 (2)	0.000***	K
	2	0.4 (2)	0.831			2	6.3 (2)	0.042	
5	1	21.5 (2)	0.000***	N, P	5	1	3.9 (2)	0.140	
	2	26.2 (2)	0.000***	N		2	17.7 (2)	0.000***	P
6	1	39.5 (2)	0.000***	N, P	6	1	20.5 (2)	0.000***	K
	2	4.8 (2)	0.091			2	5.5 (2)	0.065	
7	1	5.7 (2)	0.059		7	1	11.6 (2)	0.003**	N, P
	2	2.3 (2)	0.324			2	11.9 (2)	0.003**	P
10	1	16.7 (2)	0.000***	N, P	10	1	0.2 (2)	0.919	
	2	38.2 (2)	0.000***	N		2	41.2 (2)	0.000***	N
12	1	15.8 (2)	0.000***	N, P	12	1	7.9 (2)	0.020	
	2	30.5 (2)	0.000***	P		2	27.6 (2)	0.000***	P
13	1	56.4 (2)	0.000***	K	13	1	58.7 (2)	0.000***	K
	2	38.8 (2)	0.000***	N		2	57.9 (2)	0.000***	N
14	1	7.7 (2)	0.021		14	1	6.1 (2)	0.048	
	2	12.9 (2)	0.002**	P		2	8.0 (2)	0.018	
15	1	61.2 (3)	0.000***	N, P	15	1	26.6 (3)	0.000***	N, P
	2	7.9 (3)	0.049			2	11.4 (3)	0.010	
16	1	41.9 (3)	0.000***	N, P	16	1	27.5 (3)	0.000***	N, P
	2	68.4 (3)	0.000***	P		2	44.6 (3)	0.000***	P
17	1	15.0 (3)	0.002**	N, P	17	1	30.8 (3)	0.000***	N, P
	2	2.1 (3)	0.545			2	2.8 (3)	0.424	
18	1	10.0 (3)	0.018		18	1	12.0 (3)	0.008**	N, P
	2	51.8 (3)	0.000***	P		2	60.1 (3)	0.000***	P
19	1	18.2 (3)	0.000***	N, P	19	1	19.5 (3)	0.000***	N, P
	2	31.6 (3)	0.000***	P		2	28.1 (3)	0.000***	P

\*\* $p < .01$ , \*\*\* $p < .001$ 

K: 대한민국, N: 노르웨이, P: 폴란드

은 노르웨이와 폴란드 두 유럽국가에 유리한 것으로 나타났으며, 13번 문항은 대한민국 남학생에게 유리한 것으로 판명되어 총 11개의 DIF 문항이 발견되었다. 특히 1번 문항, 13번, 15번 문항 등은  $\chi^2(2)=62.9$  P-value=0.000,  $\chi^2(2)=56.4$  P-value=0.000,  $\chi^2(3)=61.2$ , P-value=0.000 으로 대한민국과 두 유럽 국가 간의 차이가 크게 나타났다. 여학생의 경우 1번, 7번, 15번~19번에서 노르웨이와 폴란드 두 유럽국가에 유리하였으며, 3번, 6번, 13번 문항은 대한민국 여학생에게 보다 유리하여 총 10개의 DIF 문항이 산출되었다. 남학생의 경우와 유사하게 1번, 13번 문항이 각각  $\chi^2(2)=63.1$  P-value=0.000,  $\chi^2(2)=58.7$  P-value =0.000의 통계량을 산출하여 대한민국과 유럽 두 국가와의 차이가 상대적으로 크게 나타났다.

두 유럽국가 학생들에게 유리하게 나타난 1번 문항은 발송된 이메일을 이용하여 이메일의 수신인을 찾는 선다형 문항이다. [그림 4]에 주어진 화면에서 학생들은 참조(cc)에 있는 이메일 주소를 통해 정답을 선택해야 한다(김수진 외, 2014). 이 문항은 ICILS 평가를 중 '정보 생산 및 교환' 영역의 '정보 공유'에 속하며, 평가 결과 문항난이도 474점에 해당하는 1 수준 문항(가장 낮은 수준)으로 분류되었다. ICILS 2013 참여국 평균 정답률은 66%였고, 대한민국 정답률은 57%, 노르웨이 85%, 폴란드 71%로 유럽 두 국가에서 높게 나타났다(김수진 외, 2014). 1번 문항의 경우 수신인의 이름(미나, 도현)이 이메일 주소에 제시되어 있어 쉽게 정답을 찾을 수 있을 것으로 보인다. 하지만 선택지에는 한글로 이름이 제시되어 있는



※ 출처: 김수진 외, 2014, P.35

(그림 4) 모듈 A 1번 문항



반면 이메일 주소에는 영어로 이름이 제시되어 있기 때문에 영어에 익숙한 유럽 국가 학생들에 비해 대한민국 학생들에게는 불리하게 작용하였을 것으로 보인다. 즉, 1번 문항의 경우 국가 간 사용 언어에 따라 차별적으로 기능할 가능성이 매우 높은 문항이다. 이메일의 사용 법이나 세부 기능 등에 대한 정보는 ICT 교육에서 제외할 수 없는 기본적인 사항이다. 일반적으로 사용하는 이메일 주소가 영어로 되어 있는 것도 당연한 현실이다. 따라서 차별기능 문항으로 분류된 1번 문항의 경우 국가 간 결과 비교에서는 별도의 항목으로 처리하는 방안이 적절해 보인다.

모듈 A의 종합과제는 제시된 소프트웨어를 이용하여 방과 후 운동 프로그램을 홍보하기 위한 포스터를 만드는 것으로 13번 문항은 종합과제 중 '설득력'의 평가와 관련한 부분이다. 포스터에는 [그림 5] 지시사항에 제시된 것처럼 제목, 프로그램 일정, 프로그램에 관한 정보, 참여를 위한 장비 및 복장 등의 정보가 포함되어야 한다. 13번 문항에 해당하는 평가 기준인 설득력은 '정보 변환'과 관련되어 있으며, 문항 난이도 643점 (3수준)에 해당하고 대한민국의 학생들의 정답률은 ICILS 평균(26%)보다도 두 배 이상 높은 60%로 나타나 참여국 중 가장 높았다.

'정보 변환'은 ICILS 2013 평가를 두 번째 주요영역인 '정보 생산 및 교환'의 하위요소로 컴퓨터 작업을 통해 청중과 목적에 맞게 정보를 명료하게 나타내는 능력을 의미한다. 정보



※ 출처: 김수진 외, 2014, P.40

[그림 5] 모듈 A 종합과제 세부 내용

를 통한 의사소통 효과를 높이기 위해 컴퓨터에서 사용 가능한 서식, 그래픽, 멀티미디어, 애니메이션 등을 사용하는 모든 것을 포함하고 있다(Fraillon et al., 2013). ICILS 2013의 학생을 대상으로 실시한 설문 결과에서는 대한민국 학생들의 상위 ICT 활용능력에 대한 자아 효능감(52점)이 전체 18개국 평균(50점)보다 높았으며, 노르웨이(49점), 폴란드(49점)는 상대적으로 하위권이었다. 상위 ICT 활용능력에 대한 자아 효능감은 ‘소프트웨어를 사용하여 바이러스를 찾고 제거하기’, ‘컴퓨터 네트워크 설정’ 등과 같이 컴퓨터를 활용한 응용 능력에 대한 자신감을 의미한다. 대한민국 학생들의 고수준 컴퓨터 활용 능력은 정보 변환을 통한 설득이 필요한 문제 해결에 유리하게 작용했을 가능성이 높다.

### 3. 성별 DIF 분석 : Wald-2 test 결과

ICILS 2013 모듈 A에 성별에 따른 차별기능문항이 존재하는지 여부를 조사한 Wald-2 test 결과는 <표 8>과 같다. 분석결과 대한민국에서는 6번, 15번 문항, 노르웨이는 5번, 10번 문항, 폴란드는 7번, 10번, 15번, 17번~19번 문항에서 남학생과 여학생에게 다르게 기능하는 모습이 발견되었다. 대한민국과 노르웨이에서는 총 19개 문항 중 각각 2개의 문항이 DIF가 존재하는 것으로 선정되었으나, 폴란드는 상대적으로 많은 6개 문항에서 DIF가 존재할 것으로 예측되었다. 국가 간 비교에서와 마찬가지로 문항별 세부 분석은 Wald-1 test 결과를 이용하여 진행하였다.

일반적으로 표본의 크기가 증가하면 집단 간의 차이가 존재하지 않아도 차이가 있다고 판별하는 제 1종 오류가 발생할 가능성이 있으며, 집단 간의 능력 차이가 존재하면 역시 통계적 오류가 증가하는 경향이 있다는 다수의 연구가 있다(e.g., DeMars, 2009; Li, Brooks, & Johanson, 2012). 따라서 DIF 연구에서 추가적인 정제 과정을 거친 후 DIF 문항을 최종적으로 선정해야 제 1종 오류를 줄일 수 있다(Holland & Wainer, 2012, Kim & Cohen, 1995, Lord, 1980). 이번 연구에서는 1단계에서  $\chi^2$  통계량을 구하여 잠재적인 차별기능문항들을 고정시킨 후 2단계에서 DIF 통계량을 재계산하는 방식의 2단계 정제 방법을 이용하였다.

### 4. 성별 DIF 분석 : Wald-1 test 결과

Wald-2 test에서 유의한 차이를 보인 문항들을 대상으로 Wald-1 test를 실시한 결과 대한민국은 15번 문항이 여학생에게 유리한 것으로 나타났다. 이 문항은 종합과제인 포스터 제작 과정에서 포스터의 제목이 적절한지 판단하는 능력을 평가하는 문항으로 세 개의 응답 범주(0, 1, 2)를 가지고 있다. 학생이 만든 포스터에서 제목의 위치와 역할이 명확하고, 내용이

〈표 8〉 성별 DIF 분석 : Wald-2 test 결과(표 음영처리 수정)

문항 번호	대한민국		노르웨이		폴란드	
	$\chi^2 (df)$	p-value	$\chi^2 (df)$	p-value	$\chi^2 (df)$	p-value
1	5.7 (2)	0.057	1.2 (2)	0.555	6.9 (2)	0.032
2	2.7 (2)	0.264	0.5 (2)	0.777	1.1 (2)	0.574
3	1.1 (2)	0.585	2.4 (2)	0.304	0.9 (2)	0.643
4	0.6 (2)	0.752	4.9 (2)	0.088	4.6 (2)	0.100
5	0.6 (2)	0.728	18.6 (2)	0.000***	8.8 (2)	0.012
6	9.9 (2)	0.007**	2.7 (2)	0.266	2.4 (2)	0.309
7	6.8 (2)	0.033	5.0 (2)	0.084	12.3 (2)	0.002**
8	0.3 (2)	0.869	0.5 (2)	0.795	0.7 (2)	0.706
9	0.7 (2)	0.690	0.8 (2)	0.655	0.4 (2)	0.825
10	0.7 (2)	0.720	12.2 (2)	0.002**	13.4 (2)	0.001**
11	7.7 (2)	0.022	5.9 (2)	0.051	0.1 (2)	0.931
12	1.2 (2)	0.553	3.9 (2)	0.145	4.3 (2)	0.116
13	5.3 (2)	0.069	2.4 (2)	0.307	5.7 (2)	0.059
14	5.1 (2)	0.078	4.9 (2)	0.087	1.7 (2)	0.424
15	28.8 (3)	0.000***	7.4 (3)	0.060	15.2 (3)	0.002**
16	3.4 (3)	0.339	7.8 (3)	0.051	0.6 (3)	0.896
17	3.3 (3)	0.344	1.1 (3)	0.781	22.5 (3)	0.000***
18	1.8 (3)	0.615	1.3 (3)	0.739	17.8 (3)	0.001**
19	4.4 (3)	0.221	1.6 (3)	0.662	12.8 (3)	0.005**

\*\* $p < .01$ , \*\*\* $p < .001$ 

적절한 경우 2점, 제목의 내용은 적절하지만 제목과 다른 글자의 구별이 어려운 경우는 1점, 포스터에서 제목의 구분이 어렵고 내용도 적절하지 않은 경우는 0점을 받는다(김수진 외, 2014). 15번 문항의 평가 기준인 제목 설정의 적절성은 총점이 2점으로 부분점수 1점은 평가들 중 '정보 생성'과 관련이 있고 CIL 문항난이도 492점(2수준)에 해당하는 것으로 나타났다. 이 기준에 대한 대한민국 학생들의 평균 정답률은 71%이었으며 ICILS 전체 평균은 67%였다. 2점 만점은 평가들의 '정보 변환'과 관련이 있으며 CIL 문항난이도 548점(2수준)에 해당한다. 이 기준에 대한 우리나라 학생들의 평균 정답률은 50%로 우리나라 학생들의 정답률

〈표 9〉 성별 DIF 분석 : Wald-1 test 결과

문항 번호	대한민국			문항 번호	노르웨이			문항 번호	폴란드		
	$\chi^2(df)$	p-value	선호 집단		$\chi^2(df)$	p-value	선호 집단		$\chi^2(df)$	p-value	선호 집단
6	8.9 (2)	0.012		5	16.9 (2)	0.000***	남학생	7	7.1 (2)	0.029	
15	23.3 (3)	0.000***	여학생	10	11.6 (2)	0.003***	남학생	10	8.2 (2)	0.016	
								15	17.3 (3)	0.001**	여학생
								17	23.6 (3)	0.000***	여학생
								18	17.7 (3)	0.001**	여학생
								19	14.9 (3)	0.002**	남학생

\*\* $p < .01$ , \*\*\* $p < .001$ 

은 ICILS 평균(48%)에 비해 약간 높은 수준이었다. 노르웨이(1점 75%, 2점 60%), 폴란드(1점 71%, 2점 61%)와 비교하면 1점은 비슷한 응답률을 보였으나 2점은 10%p 이상 낮게 나타났다.

15번 문항을 비롯하여 많은 문항들이 정보의 재구성, 각색 등 읽기 능력과 관련한 내용이 다수 포함되어 있으며, ICILS에서는 컴퓨터·정보 소양이 읽기 능력을 포함하는 종합적인 능력으로 보고 있다(Fraillon et al., 2014). 우리나라를 비롯하여 다수의 국가에서 여학생의 읽기 능력이 남학생보다 우수하다는 연구 결과가 다수 존재한다(김경희 외, 2010; 김기석, 김성열, 김성식, 2009; 최숙기, 박기범, 2012; OECD, 2010). 따라서 제목의 적절성을 평가하는 15번 문항이 전체 내용의 이해, 요약 등의 포괄적인 읽기 능력과 관련된다는 점을 고려할 때 이 문항은 읽기 능력이 우수한 여학생에게 유리하게 작용되었을 것으로 판단된다.

한편, 노르웨이에서는 5번과 10번 문항이 남학생에게 유리하게 작용하였다. 폴란드에서는 15번, 17번, 18번 문항은 여학생에게, 19번 문항은 남학생에게 유리한 것으로 나타났으며, 구성형 문항들에서 다수의 DIF 문항이 발견되었다.

## V. 논의 및 결론

세계경제포럼의 지구촌 의제 위원회(Global Agenda Council)가 복수의 최고 경영자 및 ICT 전문가들을 대상으로, 가까운 미래(2025년)에 일어날 것으로 예상되는 티핑 포인트(tipping points)에 대해 조사한 결과를 보면, “인류의 10%가 인터넷 연결 기능을 지닌 의류 착용”,

“3D 프린터로 제작한 간(liver)의 이식”, “블록체인(block-chain)을 통한 세금 징수” 등 지금과는 다른 미래 사회의 모습을 예상하였다(World Economic Forum, 2016). 이처럼 4차 산업혁명이 사회 전 분야에 영향을 주게 될 것으로 예상되는 가운데 교육 분야에 미치는 파급력은 더 클 것으로 보인다. 미래 사회에서는 표준화, 대중화된 산업사회와는 달리 기계적이고 틀에 박힌 방식이 아닌 새로운 방식으로 문제를 해결하는 능력(Novel and Adaptive Thinking), 다른 사람들과 직접적이고 깊게 교감·교류하는 능력(Social Intelligence)과 가상 팀의 멤버로 존재감을 드러내며 참여를 끌어내 생산성을 높이는 능력(Virtual Collaboration) 등이 더욱 중요해질 것이다(World Economic Forum, 2016). 이에 우리나라 교육 당국에서는 2015 개정 교육과정에서 초·중학교 정보 교과 관련 내용을 확대하는 방향으로 교육과정을 개정하였고 ICILS와 같은 ICT 국제 비교 평가에도 참여하는 등 ICT 리터러시, 컴퓨터·정보 소양 능력을 배양하기 위한 다양한 노력을 기울이고 있다.

이 연구에서는 ICILS 2013에서 비슷한 능력의 피험자 사이에 정답률 차이가 피험자의 순수한 능력의 차이인지 사회적, 문화적 또는 성별에 따른 집단 차이 때문이지를 알아보기 위해 DIF 분석을 진행하였다. 도출 A의 19개 문항을 대상으로 분석한 결과 남학생의 경우 10개 문항이 노르웨이와 폴란드 두 유럽국가 학생에게 유리하였으며, 13번 한 문항이 대한민국 남학생에게 유리한 것으로 판명되었다. 여학생의 경우에는 7개 문항이 노르웨이와 폴란드 두 유럽국가에 유리하였으며, 3개 문항이 대한민국 여학생에게 보다 유리하였다. 이 중 1번 문항의 경우 남, 여학생에게 공통적으로 국가 간 사용 언어(영어권 국가)에 따라 다르게 기능할 가능성이 매우 높은 문항으로 나타나 차기 검사에서는 다른 유형의 문항으로 교체하거나 국가 간 결과 비교에서 제외하여야 할 것으로 보인다. 또한, ICILS 2013에 전체적으로 두 유럽 국가에 유리한 문항이 많았다는 점을 볼 때 차후 전체 연구(National Research Coordinators, NRC) 미팅에서 이 점에 대하여 논의하고 개선점을 모색해야 한다.

국가별로 성별에 따른 차별기능문항이 존재하는지 살펴본 결과, 대한민국은 15번 문항이 여학생에게 유리한 것으로 나타났다. 이 문항은 포스터의 제목이 적절한지 판단하는 능력을 평가하는 종합과제 중 한 문항으로 정보의 재구성, 각색 등 읽기 능력과 관련한 내용이 포함되어 있다. ICILS 2013은 컴퓨터·정보 소양을 측정하는 검사이지만 읽기 능력을 비롯한 언어 능력이 검사에 많이 포함되어 있다(Fraillon et al., 2014). 이런 연유로 우리나라를 비롯하여 다수의 국가에서 여학생에게 다소 유리한 문항이 많았다고 판단된다. 물론, ICILS에서 읽기 능력 자체를 컴퓨터·정보 소양 검사에서 측정하고자 하는 중요한 구인으로 본다면 차별기능문항의 판단 기준을 다르게 볼 여지도 있다. 또한 ICILS 2013에서는 전반적으로 구성형 문항들에서 다수의 DIF 문항이 발견되어 이 문항들에 대한 보완책 마련이 요구된다. 예를 들어, 국가별로 채점 기준의 번역과정에서 해당 문항이 측정하고자 하는 바가 동일한

게 반영되고 있는지 여부를 재검토할 필요가 있다. 비록 차별기능문항이 문항의 편파성을 보여주는 것은 하지만 반드시 이 문항들이 검사에서 제거되어야 하는 것은 아니다. 검사 전체의 목적 및 방향성과 일치하고 타당성, 신뢰성을 가지고 있는 문항이라면 다소의 수정을 거쳐 검사에 잔존시켜야 할 것이다. 이와 함께 우리나라에서 ICILS 2013 결과 성차가 특히 크게 나타난 원인에 대해서는 좀 더 심층적인 연구가 필요하며, 이를 바탕으로 우리나라 학생들의 컴퓨터·정보 소양을 높일 수 있는 방안을 모색해야 한다.

이상의 연구과정 및 결과의 제한점과 함께 후속연구를 제안하면 다음과 같다. 첫째, 이번 연구는 ICILS 2018 문항 개발이 종료된 시점에 진행되었기 때문에 발표 시점상 연구 결과를 해당 검사에 직접 반영하기가 불가능하다. 그렇지만, 이번 연구를 통해 드러난 차별기능문항의 유형, 양상 등을 참고하면 향후 유사한 검사의 개발, 결과 분석에 시사점을 얻을 수 있으리라 생각한다. 둘째, ICILS의 국가 및 성별에 대한 차별기능문항을 보다 심도 있게 연구하기 위해서는 집단을 세부적으로 구분하여 문항의 유·불리가 변화되는 양상을 더욱 자세하게 분석할 필요가 있다. 예를 들어 성취 수준별로 집단을 구분하여 비교한다면 보다 의미 있는 결과를 얻을 수 있을 것이다. 셋째, 본 연구에서는 Wald-test 방법을 이용하여 차별기능문항을 검증하였다. 이 방법은 다집단 DIF 분석에 있어 정제작업의 간편함과 최신 프로그램의 개발로 집단 수가 증가하더라도 분석이 용이하다는 점 등 여러 장점을 가지고 있다. 그러나 어느 특정 방법이 항상 우수한 것은 아니며, 검증 방법에 따라 DIF 문항 추출 여부가 달라질 수 있기 때문에 다른 방법도 함께 적용해 볼 필요가 있다. 넷째, 본 연구에서 적용한 1수준 1차원 문항반응이론 모형에서 더 나아가 다수준, 다차원 문항반응이론 모형을 적용하여 모의실험 연구를 진행한다면 다양한 검사 상황에 맞는 차별기능문항 판별 방법을 제시할 수 있으리라 기대된다. 마지막으로, 본 연구에서는 국가와 성별의 주효과만을 고려하여 차별기능문항 분석을 진행하였다. 주요 국가의 교육정책 및 교수 학습 내용에 관한 보다 충분한 자료가 확보된다면 주효과와 더불어 상호작용 효과를 확인하고 그 원인을 탐색하는 작업도 의미가 있을 것으로 보인다.

## 후 주

- 1) ICILS 2013은 총 18개국이 참여하였으며, 이중 4개국은 검사 참여율을 만족하지 못하여 국제비교 대상에서 제외됨.
- 2) 호주, 폴란드, 노르웨이, 대한민국의 CIL(Computer and Information Literacy) 차이는 통계적으로 유의미한 차이는 아님.
- 3) 1. 제목 설정, 2. 이미지 배치, 3. 텍스트 배치 및 설정, 4. 색 대비, 5. 색 일관성, 6. 정보 각색,

7. 정보의 완결성, 8. 설득력, 9. 전체 페이지의 활용

## 참고문헌

- 교육부(2015a). **2015 개정교육과정**. 교육부고시 제2015-74호(2015.9.23.).
- 교육부(2015b). **2015 개정교육과정**. 교육부고시 제2015-80호(2015.12.1.).
- 교육부(2015c). **2015 개정 정보과 교육과정**. 교육부 고시 제2015-74호 [별책 10].
- 교육부, 미래창조과학부(2015). SW중심사회를 위한 인재양성 추진계획. 보도자료(2015.7.21.).
- 김경성, 곽현석, 김중훈, 김혜숙, 서정희, 이수영, 전우천(2010). **2010 국가수준 초·중등학생 ICT 리터러시 수준 평가 연구**. 서울: 한국교육학술정보원.
- 김경성, 김중훈, 김혜숙, 전우천, 최성우(2009). **2009 국가수준 초·중등학생 ICT 리터러시 수준 평가 연구**. 서울: 한국교육학술정보원.
- 김경희, 시기자, 김미영, 옥현진, 임해미, 김선희, 정지영, 정송, 박희재(2010). **OECD 학업성취 도국제비교 연구(PISA 2009)**. 한국교육과정평가원 연구보고 RRE 2010-4-2.
- 김기석, 김성열, 김성식(2009). **학업 성취도에 대한 학생과 학교 변인들의 효과 분석-PISA 2000, 2003, 2006 자료를 중심으로**. 한국교육과정평가원 연구보고 RRE 2009-3.
- 김수진, 박지현, 전정희, 김미영, 이영준, 서지희, 김민정(2014). **국제 컴퓨터·정보 소양 연구: ICILS 2013 검사 결과보고서**. 서울: 한국교육과정평가원.
- 김수진, 이재봉, 박지현, 이문수, 이영준(2015). **우리나라 학생들의 컴퓨터·정보 소양 특성 및 교육맥락변인의 영향 분석**. 서울: 한국교육과정평가원.
- 김영애, 서정희(2011). **2010년 초·중등학생 ICT 리터러시 수준 측정의 결과와 시사점**. 한국 교육학술정보원 연구보고 RM 2011-2.
- 김종민, 안성훈, 이문수, 임현정, 김한성(2017). **초·중학생 ICT 리터러시 수준 심층 분석 (2007~2016년)**. 한국교육학술정보원 연구보고 KR 2017-6.
- 김종민, 이문수, 안성훈(2016). 2015년도 국가수준 초·중학생 ICT 리터러시 검사의 성별에 따른 차별기능문항 분석. **교육평가연구**, 29(2), 255-278.
- 김현철, 정순영, 김자미, 김홍래, 서정희(2011). **2011 국가수준 초·중등학생 ICT 리터러시 수준 평가 연구**. 서울: 한국교육학술정보원.
- 김혜숙, 진성희(2006). **미국 ETS의 ICT 리터러시 평가 현황 및 시사점**. 한국교육학술정보원 RM 2006-36.
- 김홍래, 이승진(2013). **외국의 정보(컴퓨터) 교육과정 현황 분석**. 한국교육학술정보원 RM

- 2013-17.
- 노르웨이 교육부(2016). Coding is elective in 146 schools. Regjeringen.no. Retrieved July 8, 2018, from <https://www.regjeringen.no/no/aktuelt/koding-blir-valgfag-pa-146-skoler/id2481962/>
- 백순근, 김동일, 김리량, 김혜숙, 박소화, 유예림, 김세원, 김미림(2008). **ICT 리터러시 검사도구 개발 연구 - 중등학생용**. 서울: 한국교육학술정보원.
- 서순식, 민경석, 황경현, 장운정, 김혜숙(2009). **초등학생용 ICT 리터러시 검사도구 타당화 연구**. 서울: 한국교육학술정보원.
- 성태제(1994). 1994학년도 제1차 대학수학능력시험의 성별에 따른 차별기능문항 추출. **교육평가연구**, 7(2), 87-101.
- 송미영(2001). 수행평가 도구의 성별에 따른 차별기능문항 추출 및 추출방법 비교. 박사학위 논문, 이화여자대학교.
- 안성훈, 김성식, 김혜원, 남창우, 양혜경, 김운정, 조규복(2014). **2014 국가수준 초·중학생 ICT 리터러시 수준 측정 연구**. 대구: 한국교육학술정보원.
- 안성훈, 김성식, 남창우, 김종민, 김혜원(2015). **2015 국가수준 초·중학생 ICT 리터러시 수준 측정 연구**. 대구: 한국교육학술정보원.
- 이대용, 김석우, 길임주(2014). 성별에 따른 디자인 진로적성 검사의 차별기능문항군 탐색. **교육평가연구**, 27(4), 945-964.
- 이원규, 김영기, 김현철, 서순식, 전우천, 한선관, 김영애, 김혜숙(2007). **ICT 리터러시 검사도구 개발 연구-초등학생용-**. 서울: 한국교육학술정보원.
- 이 정, 이순목(2009). 다면적 성격검사 척도 반응의 성차 분석: 맨틀-헨젤(Mantel-Haenszel) 통계방법을 통해 본 차별적 문항 기능(DIF). **한국심리학회지: 일반**, 28(1), 263-281.
- 최숙기, 박기범(2012). 읽기 성취에 대한 남학생과 여학생 간 성별차이 분석 및 교육적 방안 탐색 - PISA 2009 읽기 관련 변인을 중심으로. **교육과정평가연구**, 15(1), 239-266.
- ACER (2010). *National Assessment Program - ICT Literacy Year 6 & 10 Report*. Australia: MCEECDYA.
- Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal*, 21(4), 495-508.
- Cai, L. (2008). SEM of another flavor: Two new applications of the supplemented EM algorithm. *British Journal of Mathematical and Statistical Psychology*, 61, 309-329.
- Cai, L. (2013). flexMIRT<sup>®</sup> version 2.00: A numerical engine for flexible multilevel multidimensional item analysis and test scoring [Computer software]. Chapel Hill, NC: Vector Psychometric Group.
- Cai, L., Thissen, D., & du Toit, S. (2011). *IRTpro for Windows*. [Computer software]. Lincolnwood, IL:



Scientific Software International.

- Chambers, R., Kokic, P., Smith, P. and Cruddas, M. (2000). *Winsorization for identifying and treating outliers in business surveys*. Proceedings of the Second International Conference on Establishment Surveys, 717-726, American Statistical Association Alexandria, Virginia.
- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: issues and practice*, 17(1), 31-44.
- DeMars, C. E. (2009). Modification of the Mantel-Haenszel and Logistic Regression DIF Procedures to Incorporate the SIBTEST Regression Correction. *Journal of Educational Behavioral Statistics*, 34(2), 149-170.
- ETS (2005). *Measuring college-level information and communication technology proficiency*. An issue report. Available at: [www.ets.org/ictliteracy](http://www.ets.org/ictliteracy).
- Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). *Preparing for Life in a Digital Age: The IEA International Computer and Information Literacy Study International Report*. Cham, Switzerland: Springer.
- Fraillon, J., Schulz, W. & Ainley, J. (2013). *International Computer and Information Literacy Study: Assessment framework*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).
- Gardiner, W. L.(1998) *Can Computers Teaching Inside-Out, Transform Education, and Redefine Literacy?*, Kuber Robert(eds), New Brunswick & London; Transaction Publisher.
- Gurbiel, E., Hardt-Olejniczak, G., Kolczyk, E., Krupicka, H., & Syslo, M. (2005). *Informatics and ICT in Polish Education System*. Informatics in Secondary Schools Evolution and Perspectives. Austria.
- Holland, P. W., & Wainer, H. (2012). *Differential item functioning*. Routledge.
- Kim, S., & Cohen, A. S. (1995). A comparison of Lord's chi-square, Raju's area measures, and the likelihood ratio test on detection of differential item functioning. *Applied Measurement in Education*, 8, 291-312.
- Kokic, P. N. and Bell, P. A. (1994). Optimal winsorizing cutoffs for a stratified finite population estimator. *Journal of Official Statistics*, 10, 419-435.
- Langer, M. (2008). *A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation* (Unpublished doctoral dissertation). University of North Carolina, Chapel Hill.
- Li, Y., Brooks, G. P., & Johanson, G. A. (2012). Item discrimination and type I error in the

- detection of differential item functioning. *Educational and Psychological Measurement*, 72, 847-861.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum.
- Magis, D., Raïche, G., Béland, S., & Gérard, P. (2011). A generalized logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing*, 11(4), 365-386.
- Organisation for Economic Co-operation and Development (OECD). (2009). *Equally prepared for life?: How 15-year-old boys and girls perform in school*. OECD.
- Penfield, R. D. (2001). Assessing differential item functioning among multiple groups: A comparison of three Mantel-Haenszel procedures. *Applied Measurement in Education*, 14(3), 235-259.
- Wald, A. (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is larger. *Transactions of the American Mathematical Society*, 54, 426-482.
- Woods, C., Cai, L., Wang, M. (2012). The Langer-Improved Wald Test for DIF Testing With Multiple Groups: Evaluation and Comparison to Two-Group IRT. *Educational and Psychological Measurement*, 72(3), 1-16.
- World Economic Forum. (2016). The future of jobs: Employment, skills and workforce strategy for the fourth industrial revolution. *In World Economic Forum*.

© 논문접수: 2018. 08. 05 / 수정본 접수: 2018. 09. 08 / 게재승인: 2018. 09. 14

— 저 자 소 개 —

· 이문수 : UCLA에서 교육측정 및 평가 전공으로 박사학위를 취득하였으며, 현재 한국교육과정  
평가원 부연구위원으로 재직 중임. 연구 관심 분야는 문항반응이론, 척도개발과 동등  
화, 컴퓨터 적응검사, 다층모형 연구 등임. mslee9@kice.re.kr

〈ABSTRACT〉

**Differential Item Functioning in the aspects of  
Country and Gender Differences in the Computer and  
Information Literacy Test: Based on ICILS 2013**

Moonsoo Lee

Korea Institute for Curriculum and Evaluation

The purpose of this study was to investigate the differential item functioning (DIF) according to the countries and genders of Korea, Norway, and Poland with similar achievements in the ICILS 2013. The results of the Wald-test on the 19 items of module A showed that 10 items for boys and 7 items for girls were beneficial to two European countries. One of the item was likely to function differently according to the language. As a result of examining whether there were DIF items according to gender by country, Korea showed high achievement of girls in item 15 and it seemed to be beneficial to girls with excellent language ability. In ICILS 2013, a large number of DIF items were found in the construct response items.

*Keywords : ICT literacy, ICILS, multiple group DIF, Wald test*