

대학생의 데이터 리터러시 측정 도구 타당화 및 신입생 데이터 리터러시 진단 - A 대학 사례를 중심으로

정윤숙¹, 유지원²

¹제1저자, 가천대학교 교육성과관리센터 대우교수, yschung06@gachon.ac.kr

²교신저자, 가천대학교 교육대학원 부교수, uimagine@gachon.ac.kr

초록

4차 산업혁명 시대의 기술 중심에는 데이터가 있다. 데이터로 문제를 발견하고 의미를 해석하며 다른 사람들과의 데이터 기반으로 효과적인 의사소통과 합리적인 의사결정을 할 수 있는 데이터 리터러시가 현대 사회의 필수 소양으로 논의된다. 이러한 맥락에서 본 연구는 대학생의 데이터 리터러시를 측정할 수 있는 진단 도구를 타당화하고 신입생의 데이터 리터러시 수준을 살펴봄으로써 대학 교양교육으로서 데이터 리터러시 교육의 필요성을 검토하고 향후 관련 교육과정에 시사점을 제공하고자 수행되었다. 자료수집은 수도권 소재 A 대학의 신입생 3,510명으로부터 이루어졌으며, 진단 도구의 타당화를 위해 탐색적 요인분석과 확인적 요인분석을 시행하였다. 분석 결과 최종 19문항이 선정되었고, 도구는 양호한 신뢰도와 타당도를 보였다. 이후 기술통계로 신입생의 데이터 리터러시 수준을 분석하였으며, 데이터 리터러시가 성별, 계열별 차이가 있는지 t-test, ANOVA를 활용하여 분석하였다. 분석 결과, 신입생의 데이터 리터러시는 높지 않았고, 세부적으로 데이터 분석과 데이터 도구 활용 하위요인에서 특히 낮은 결과를 보였다. 성별 비교 결과에서는 유의한 차이가 없었으나, 계열 간에는 유의한 차이를 보였으며 이공계에 비해 인문·사회 및 예체능 계열이 낮게 나타났다. 연구 결과를 토대로 기초교양교육에서 데이터 리터러시 교육의 필요성을 논의하고 도구를 활용한 데이터 분석 역량 함양을 위한 제언을 하였다.

주제어: 데이터 리터러시, 정보처리역량, 양적추론, 측정 도구, 타당화, 기초교양교육

이 논문은 2022년도 가천대학교 교내연구비 지원에 의한 결과임.(GCU-202207190001)

이 논문은 2022년 9월 20일에 투고 완료되어
2022년 10월 04일 편집위원회에서 심사위원을 선정한 뒤
2022년 10월 11일까지 심사를 완료하여
2022년 10월 18일 편집위원회에서 게재가 결정된 논문임.

교양 교육 연구
Korean Journal of
General Education

1. 서론
2. 연구 방법
3. 연구 결과
4. 논의 및 결론

1. 서론

우리는 일상생활 속에서 이메일, SNS, YouTube, 전자상거래 등을 활용하면서 매일 많은 데이터를 만들어 내고 있으며 이렇게 수집된 빅데이터는 인공지능, 머신러닝, 딥러닝 등 4차 산업혁명 기술의 중심에 자리하고 있다. 현대는 데이터를 활용하여 새로운 가치를 창출하는 데이터 주도 경제 시대라 해도 과언이 아니다(Newman, 2011). 그러나 데이터 홍수 시대에서 축적된 데이터가 그 자체로 가치를 발하는 것은 아니다. 핵심은 무수한 데이터로부터 문제를 발굴할 수 있고, 패턴과 관계를 파악하며, 데이터의 의미를 통찰해내고 예측하는 등 데이터에 근거한 문제해결과 의사결정을 할 수 있는 능력을 갖추는 것이다. 이에 다수 기관과 학자들은 데이터를 이해하고 분석하며 활용할 뿐 아니라 데이터로 소통할 수 있는 역량을 강조하며, 이제는 글을 읽고 쓰는 전통적인 리터러시에서 나아가 데이터 리터러시가 현대인의 필수 소양이라고 논의한다(송유경, 송석리, 김예지, 임철일, 2021; 배화순, 2019; OECD, 2019). 특히 Macy와 Coates(2016)는 미래 인재는 데이터로부터 유용한 정보를 습득할 수 있는 역량을 갖추어야 하며, 대학에서 데이터 리터러시에 대한 교육을 시행해야 한다고 제안하였다. 우리나라 교육부(2020)도 데이터는 4차 산업 혁명 시대의 가장 중요한 자원이며, 정확한 문제해결, 원활한 커뮤니케이션, 가짜뉴스 판별 등을 위해 모든 시민이 데이터 리터러시를 갖추어야 함을 강조한 바 있다. 이러한 맥락에서 데이터 리터러시는 데이터 과학을 다루는 이공계 학생에게만 요구되는 역량이 아니며, 정치, 경제, 행정, 교육 등 전공을 막론하고 모든 대학생과 밀접하게 관련된 역량이라 할 수 있다.

현대 사회에서 데이터 리터러시의 중요성이 대두됨에 따라 다수 학자가 데이터 리터러시의 정의와 개념을 제시하고 있다. Carlson, Fosmire, Miller와 Nelson(2011)은 “도표를 제대로 읽는 법, 데이터로부터 올바른 결론을 도출하는 법, 데이터가 부적절하게 사용되고 있을 때 이를 인지하는 법 등을 포함하는, 데이터가 의미하는 바를 이해하는 것”(p.5)이라고 정의하였는데, 데이터를 다루는 기술적인 측면뿐 아니라 사회적인 측면도 다루고 있다는 점이 특징이다(김태영, 김철민, 윤혜경, 2020). Bhargava와 D'Ignazio(2015)는 “데이터를 읽고(read), 데이터로 일하고(work with), 데이터를 분석하며(analyze), 데이터로

논증(argue with)하는 능력”으로 데이터의 이해와 활용뿐 아니라 자신의 주장을 뒷받침하는 데 활용할 수 있는 능력까지 포함하는 정의를 제시하였다. Mandinach와 Gummer(2013)는 데이터 이해와 활용의 궁극적인 목표는 의사결정에 활용하기 위함이라고 강조하였으며, 배화순(2019)은 “데이터에 기반하여 정보를 추출하고, 데이터에 대한 질문을 생성하거나 실생활의 다양한 문제를 해결하기 위해서 데이터를 활용하며, 적절한 도구를 활용하여 데이터를 구성하고, 결론을 도출하는 일련의 과정을 의미 ... 상호 의사소통 과정에서 데이터를 사용하는 것까지도 개념적으로 포괄”(p. 97)하는 것이라고 정의하였다. 이처럼 학자마다 약간씩 차이를 보이지만, 데이터 리터러시에 대한 정의는 일반적인 통계 분석 능력과 달리, 데이터로 문제를 파악하고, 데이터로부터 의미 있는 가치를 추출하며, 데이터를 활용하여 다른 사람을 설득하고 효과적으로 의사소통하는 요소까지 포함한다.

데이터 리터러시를 정의하기 위한 시도와 함께 교육 현장에서는 학생들의 데이터 리터러시를 함양하기 위한 노력이 가시화되고 있다. OECD(2019)가 발표한 Learning compass 2030에서는 데이터 리터러시를 기초소양으로 명시하여 미래 사회의 핵심역량임을 분명히 하였고, 미국을 비롯한 해외 대학에서는 데이터에 근거한 비판적 사고를 다루는 양적 추론(quantitative reasoning) 교과목, 데이터 과학(data science) 교과목을 대학생의 교양 기초교육으로 운영하고 있다(김혜영, 2020; 김혜영, 이은하, 2019; HRA, 2015). 예로, 캐나다 델하우지대학교의 경우, 데이터 리터러시는 학생과 데이터 사용자뿐만 아니라 민주 시민이 갖추어야 할 핵심 역량임을 강조하며, 대학의 기초교양(core curriculum)에 포함해야 한다고 주장하였고, 여러 교과목이 연계된 프로젝트 기반 학습으로 교육할 것을 제안하였다(Risdale et al., 2015). 미국의 캘리포니아대학교(University of California, Berkeley; UC Berkeley)는 졸업생이 반드시 가져야 할 역량 중 하나로 데이터 리터러시를 강조하며 전교생을 대상으로 ‘데이터 과학의 기초(foundations of data science)’ 과목을 새로 개설하였다. 이 교과에서는 경제, 지리, 소셜 네트워크 분야의 실제 데이터를 활용하여 컴퓨터 프로그래밍, 통계 및 추론, 데이터 기반 의사결정의 기초를 학습할 기회를 제공한다(Kell, 2015). 또한 미국 매사추세츠주의 브랜다이스대학교(Brandeis University)와 뉴욕시립대학교(The City University of New

York) 소속의 리먼대학(Lehman College)은 언어학, 경제학, 정치학, 경영학, 생화학, 생물학, 수학, 데이터과학 등 학문 영역과 관계없이 입문 과정의 모든 교과과정에서 데이터를 이해 및 분석, 해석할 수 있는 능력을 갖추도록 데이터 리터러시 교육과정을 구성하고 있다(Brandeis University, 2022; Lehman College, 2022). 이처럼 대학들은 데이터 리터러시를 모든 학생이 습득할 역량으로 간주하면서 기초교양교과목으로 운영하기 시작하였으며, 이들 교과는 다양한 학문 분야와 융합된 교과목의 성격을 띤다.

대학에서 교양 교육은 “사회적으로 요청되는 기본적인 소양을 함양하는 교육(강명구, 김지현, 2010, p.339)”을 지향하고, 지식기반사회에서 교양교육은 학생들이 새로운 아이디어와 가치를 창출하는 기회를 부여하고, 방대한 정보 속에서 유용한 정보를 판별할 수 있으며, 자신의 아이디어를 타인과 공유하고 소통하는 능력 등을 함양해야 한다고 논의된다(손동현, 2007; 이현주, 박성희, 2022). 따라서 대학의 교양교육은 학생들이 사회인, 민주시민, 평생학습자로 성공적인 삶을 영위할 수 있도록 문·이과 전공과 관계없이 데이터 리터러시 역량을 강화하는 학습 기회를 제공하는 것이 적절하다. 그러나 아직 우리나라 대학의 기초 교양교육과정은 대부분 글쓰기와 외국어 중심으로 구성되어 있으며, 수학 및 기초과학 교과는 이공계 학생들을 위한 전공기초 교과로 간주하는 경우가 많아 인문·사회 계열 학생의 경우 데이터와 관련된 교과를 접할 기회가 적은 실정이다. 최근 소프트웨어중심대학사업과 함께 학생들의 소프트웨어 능력이 강조되면서 소프트웨어 교과가 교양 교육과정에 새로이 편제되는 경향이 나타나지만, 그 비율은 높지 않으며 이공계 학생을 대상으로 한정적으로 개설된 경우가 대다수이다(정승원, 장현수, 김세준, 2020). 그마저 교과 내용이 앱 인벤터, C언어, 파이썬 등 프로그래밍 언어와 도구를 활용하는데 초점이 있어(김성현, 2020) 데이터 리터러시 함양과는 다소 거리가 있다고 지적된다(김혜영, 이은하, 2019).

데이터 리터러시 관련 선행연구도 비교적 최근에 국내에서 본격적으로 수행되는 추세이다. 관련 주제를 살펴보면, 교양교육으로 데이터 리터러시와 밀접한 양적 추론 교과의 개설 현황과 분석(김혜영, 이은하, 2019), 데이터 리터러시 모형 설계(한상우, 2020), 토론 기반 데이터 리터러시 수업모형 및 교수전략 개발(송유경 외, 2021), 교양교육으로서의 데이터 리터러시 교육의 방향 탐색(김슬

기, 김태영, 2022; 김태영 외, 2020) 등이 있다. 하지만 데이터 리터러시 연구가 아직 초기 단계라 데이터 리터러시 교육의 필요성과 시사점을 논의하거나 일부 교과목 수강생을 대상으로 데이터 리터러시 교육의 효과성을 분석하는 정도에 머무르고 있다. 이 중 데이터 리터러시에 대한 개념틀을 제시하고 고등학생을 대상으로 진단하는 연구가 일부 수행되었으나(송유경 외, 2021), 진단 도구를 실증적인 데이터로 타당화한 연구가 없고 데이터 리터러시 개념틀에 대한 논의도 충분하지 않다. 데이터 리터러시의 구인을 도출하고 진단 도구를 개발하고 타당화하는 것은 데이터 리터러시 측정을 위해 수반되어야 하는 노력이며, 관련 연구 활성화에 필수적이다. 또한 데이터 리터러시 함양을 위한 교육과정 개발이 체계적으로 이루어지고 교육에 대한 성과가 관리되기 위해서는 학생들의 데이터 리터러시를 객관적으로 측정하는 것이 선행되어야 하고, 학생들의 데이터 리터러시 수준을 살펴보면 시사점을 도출할 필요가 있다.

이에 본 연구에서는 선행연구 고찰을 토대로 대학생을 위한 데이터 리터러시 측정 도구를 구성하고, 국내 대학생을 대상으로 수집한 자료로 측정 도구를 실증적으로 타당화하며, 현재 대학생의 데이터 리터러시 수준을 진단하였다. 특히 대학 신입생을 대상으로 데이터 리터러시 수준을 측정함으로써 대학에서 데이터 리터러시 교육의 필요성을 검토하고 향후 교육과정 개발에 기초자료를 제공하고자 하였다. 이와 같은 논의를 바탕으로 도출한 본 연구의 연구 문제는 다음과 같다.

연구 문제 1: 대학생의 데이터 리터러시 측정을 위해 구성한 측정 도구는 타당한가?

연구 문제 2: 대학 신입생의 데이터 리터러시 수준은 어떠한가?

연구 문제 3: 대학 신입생의 데이터 리터러시 수준은 성별, 계열별에 따라 차이가 있는가?

2. 연구 방법

2.1. 연구 대상

본 연구는 기초교양교육의 일환으로 데이터 리터러시 필요성을 탐색하는 것을 목표로 한다. 기초교양과목의 경

우 대학 신입생 시기에 수강하는 경우가 대부분이므로 신입생을 대상으로 데이터 리터러시 수준을 파악하는 것이 필요하다고 판단하였고, 본 연구에서는 수도권 소재 A 대학의 신입생을 대상으로 수행하였다. A 대학은 4년제 종합대학이며 2021년 중앙일보 대학평가에서 30위권 내에 평가된 중위권 대학이다. A 대학은 다수 국가사업 및 국내외 기업과의 협력을 통해 대학발전과 평판을 높이고 있으며 그에 따라 입학하는 신입생의 학력 수준도 꾸준히 높아지고 있는 편이다. 자료 수집을 위해 2022년 3월 신입생을 대상으로 데이터 리터러시 진단을 시행하였고, 1학년 재학생의 70.6%인 3,667명이 참여하였다. 본 연구에서는 연구 참여에 동의한 학생의 응답 중 외국인, 재외국민, 불성실 응답을 제외한 후 3,510명의 데이터를 활용하였다. 연구에 참여한 응답자의 배경정보는 <표 1>과 같다.

총 3,510명의 참여자 중 남학생은 1,797명(51.2%), 여학생은 1,713명(48.8%)이었다. 계열별로 구분해보면, 인문·사회 계열 학생이 1,198명(34.1%), 자연 계열은 668명(19.0%), 공학 계열은 1,355명(38.6%), 예체능 계열은 289명(8.2%)이었다. 본 연구에서는 진단 도구 타당화를 위하여 탐색적 요인분석(EFA)과 확인적 요인분석(CFA)을 순차적으로 실시하였는데, 전체 데이터를 무작위로 2분할하여 탐색적 요인분석과 확인적 요인분석에 분리하여 사용하였다. EFA의 대상이 되는 1차 분석 표본($n_1=1,780$ 명)에는 남학생이 922명(51.8%), 여학생이 858명(48.2%)이었고, 인문·사회 계열이 639명(35.9%), 자연 계열이 342명(19.2%), 공학 계열이 645명(36.2%), 예체능 계열이 154명(8.7%)이었다. CFA에 사용된 2차 분석 표본($n_2=1,730$ 명)에는 남학생 875명(50.6%), 여학생 855명(49.4%)이 배치되었고, 인문·사회 계열 559명(32.3%), 자연 계열 326명(18.8%), 공학 계열 710명(41.0%), 예체

능 계열 135명(7.8%)으로 구성되었다.

2.2. 연구 절차 및 분석 방법

다음과 같은 절차로 연구를 수행하였다. 첫째, 문헌 고찰을 통해 데이터 리터러시의 개념적 틀을 도출하고 이를 바탕으로 예비 측정 문항을 구성하였다. 둘째, 구성된 문항을 활용하여 A 대학의 신입생을 대상으로 데이터 리터러시 역량 진단을 시행하고 자료를 수집하였다. 셋째, 구성된 개념틀과 측정 도구에 대한 신뢰도 및 타당도를 검증하기 위해 수집된 자료를 2분할하여 EFA와 CFA를 실시하였다. EFA에는 SPSS 25 프로그램을 사용하였으며, 최대우도법 직접 오블리민(direct oblimin) 회전을 통해 반복적으로 요인구조 분석을 수행하였다. 이후, EFA 결과를 반영하여 분할된 다른 표본을 대상으로 CFA를 수행하였다. CFA는 AMOS 25 통계 프로그램을 활용하였으며, 측정 모형 적합도, 요인부하량, 개념신뢰도, 평균분산추출 등의 값을 검토하여 도구의 타당도를 검증하였다. 넷째, 대학 신입생 데이터 리터러시 수준을 파악하기 위하여 진단 결과 전체를 대상으로 기술통계분석을 하고 학생들의 데이터 리터러시 수준을 분석하였다. 다섯째, 집단별 데이터 리터러시 수준 차이를 분석하였다. 성별 비교는 독립표본 t 검정을 활용하였고, 계열별 비교는 일원배치 분산분석 및 Duncan을 활용하여 사후분석을 실시하였다.

<표 1> 연구 참여자 배경 정보

구분		전체	1차 분석 표본(EFA 대상)	2차 분석 표본(CFA 대상)
		빈도((명), 비율(%))	빈도((명), 비율(%))	빈도((명), 비율(%))
성별	남	1,797 (51.2)	922 (51.8)	875 (50.6)
	여	1,713 (48.8)	858 (48.2)	855 (49.4)
계열	인문·사회 계열	1,198 (34.1)	639 (35.9)	559 (32.3)
	자연 계열	668 (19.0)	342 (19.2)	326 (18.8)
	공학 계열	1,355 (38.6)	645 (36.2)	710 (41.0)
	예체능 계열	289 (8.2)	154 (8.7)	135 (7.8)
계		3,510 (100.0)	1,780 (100.0)	1,730 (100.0)

3. 연구 결과

3.1. 데이터 리터러시 도구 검증

3.1.1. 데이터 리터러시 예비 측정 도구 구성

데이터 리터러시의 정의와 포함되는 하위요인은 학자마다 다양하게 제시하였다. 하지만 필요한 데이터를 선별하여 수집하는 능력과 수집한 데이터를 분석 또는 처리하는 데이터 가공 능력은 데이터 과학, 컴퓨터 교육, 정보 교육 등 분야를 막론하고 대부분의 연구에서 데이터 리터러시의 공통적인 구인으로 논의된다(김혜영, 2020; Fontichiaro & Oehrli, 2016; Otto, 2012). 예로 김혜영(2020)은 데이터 이해, 수집, 관리, 평가, 활용 역량으로 데이터 리터러시의 세부 요인을 제시하였고, Swan, Vahey, Rafanan, & Stanford (2009)는 데이터에 기반한 문제를 만들고 데이터에 근거한 해결안을 찾는 능력, 데이터를 적절하게 분석하고 표현할 수 있는 능력, 데이터에 기초한 추론과 설명을 끌어내고 평가할 수 있는 능력으로 개념화하였다. 최근에는 데이터 리터러시를 모든 사람에게 필요한 소양으로 간주하면서, 데이터를 비판적으로 판단하는 능력과 데이터에 근거한 의사결정 및 효과적으로 의사소통하는 역량이 데이터 리터러시의 구성요소로 강조된다(배화순, 2019; 송유경 외, 2021; Mandinach & Gummer, 2013). 송유경 외(2021)는 데이터 리터러시가 융합적 사고를 이끌어내고, 효과적

인 의사소통과 합리적인 의사결정에 중요하다고 강조하면서 데이터 이해, 데이터 수집 준비, 데이터 분석, 데이터 평가, 데이터 표현, 데이터 기반 의사소통 등의 6 요인으로 데이터 리터러시를 개념화하고 측정 문항을 제안하였다. 이 외에도 미국대학교육협의회(Association of American Colleges and Universities: AAC&U)는 대학생이 갖추어야 할 주요 역량의 하나로 정량적 문해력(Quantitative literacy)을 제시하였으며, 역량 수준 평가 기준인 VALUE rubrics에 수학적 형식의 정보 해석, 수학적 표현, 계산, 응용과 분석, 가설 수립, 데이터 기반 의사소통을 하위 요소로 제시하였다(Rhodes, 2010).

본 연구에서는 선행연구 고찰을 통해 하위요소를 분석하면서 <표 2>와 같이 데이터 리터러시의 개념틀을 구성하였다. 총 7개의 하위요인으로 구성되었고 하위요인별 설명은 다음과 같다. 첫째, ‘데이터 이해’는 문제를 해결하기 위한 데이터 유형과 내용을 파악하는 능력을 의미한다. 데이터 이해 요인은 ‘내 가설을 뒷받침할 수 있는 데이터가 무엇인지 판단할 수 있다’ 등의 2개의 문항으로 구성되었다. 둘째, ‘데이터 분석 계획 및 수집’은 데이터를 수집하고 분석하는 계획을 수립하여 데이터 수집을 실행할 수 있는 능력을 의미한다. 이 요인은 ‘문제 해결에 필요한 자료를 검색하여 적절한 형태로 수집할 수 있다’ 등의 3문항으로 구성되었다. 셋째, ‘데이터 분석’은 목적에 맞게 데이터를 정제하고 분석하는 능력을 의미하며, ‘나는 데

<표 2> 데이터 리터러시 개념틀

하위 요인	정의	Borner 외 (2016)	Gray 외 (2018)	Otto (2012)	Swan 외 (2009)	송유경 외 (2021)	유지원 (2016)	한상우 (2020)
데이터 이해	문제를 해결하기 위한 데이터 유형과 내용을 파악하는 능력			●		●		●
데이터 분석 계획 및 수집	데이터를 수집하고 분석하는 계획을 수립하여 데이터 수집을 실행할 수 있는 능력	●	●	●	●	●		●
데이터 분석	목적에 맞게 데이터를 정제하고 분석하는 능력	●	●	●	●	●		●
데이터 평가	데이터 분석 과정과 결과에 오류가 있는지 감지하고 데이터 분석 결과를 비판적 사고로 평가하는 능력		●		●	●		●
데이터 도구 활용	적절한 도구를 활용하여 효율적이고 정확한 데이터 분석과 관리를 할 수 있는 능력				○		●	○
데이터 표현	표와 그래프 등을 사용하여 데이터를 시각화하여 의미를 전달하는 능력	●	●			●		●
데이터 기반 의사소통	데이터에 근거하여 타인에게 설명하고 설득할 수 있는 능력		●			●		●

○: 도구 활용에 대한 언급은 있지만 명확하게 문항으로 제시하지 않음

이터로부터 표나 그래프를 만들기 위한 기준 항목을 찾아 낼 수 있다' 등의 4문항으로 구성되었다. 넷째, '데이터 평가'는 데이터 분석 과정과 결과에 오류가 있는지 감지 하고 데이터 분석 결과를 비판적 사고로 평가하는 능력을 의미한다. 데이터 평가는 '데이터를 분석할 때 오류가 있는지 점검하고, 이를 바로잡을 수 있다' 등 3문항으로 구성되었다. 다섯째 요인인 '데이터 도구 활용'은 데이터를 다루는 다양한 상황에서 적절한 도구를 활용하여 효율 적이고 정확한 데이터 분석과 관리를 할 수 있는 능력을 의미한다. '나는 계산기, 소프트웨어(엑셀, 통계 프로그램 등), 앱 등을 이용하여 정보해석의 정확도를 높일 수 있다' 등의 4문항으로 구성되었다. 기존 선행연구와는 달리, 본 연구에서 데이터 도구 활용을 독립된 하위요인으로 개념 틀에 포함하였다. 현대 사회에서 디지털 도구 없이 데이터를 수집 또는 관리하는 것이 불가능하고, 이러한 능력 수준 이 문제해결 또는 의사결정의 질에 영향을 줄 수 있기 때문이다(유지원, 2016). 여섯째, '데이터 표현'은 표와 그래프 등을 사용하여 데이터를 시각화하여 의미를 전달하는 능력을 뜻하며, '여러 시각화 방식 중 데이터의 내용과 전달 목적에 가장 적합한 방식을 선택할 수 있다' 등의

2문항으로 구성되었다. 마지막으로 일곱째, '데이터 기반 의사소통'은 데이터에 근거하여 타인에게 설명하고 설득 할 수 있는 능력을 의미한다. 데이터 기반 의사소통은 '나는 데이터를 활용하여 나의 주장을 객관적으로 뒷받침할 수 있다' 등의 3문항으로 구성되었다. 종합하자면 본 연구에서 구성한 데이터 리터러시 예비 측정 도구는 데이터 이해, 데이터 분석 계획 및 수집, 데이터 분석, 데이터 평가, 데이터 도구 활용, 데이터 표현, 데이터 기반 의사소통 7 요인 21문항으로 구성되었으며, 모든 문항은 5점 Likert 척도(매우 그렇지 않다~매우 그렇다) 형식으로 측정되었다.

3.1.2. 탐색적 요인분석(EFA)

첫 번째 표본이 EFA를 수행하기에 적합한지 살펴보기 위하여 Kaiser-Meyer-Olkin(KMO)와 Barlett의 구형성 검정을 실시한 결과, KMO는 .93, Barlett 구형성 검정치는 18,359.79($df=171$, $p<.001$)로 적합한 것으로 나타났다. 최대우도법과 직접 오블리민 회전 방법을 적용하여 반복적으로 요인구조를 탐색하였고, 요인분석 결과는 <표 3>과 같다. 분석 결과 개념틀에서 제안한 '데이터 이해'와 '데이터 분석 계획 및 수집'이 하나의 요인으로 통합됨에 따라

<표 3> 탐색적 요인분석 결과

측정변수	요인					
	1	2	3	4	5	6
[문항1] 데이터 이해 및 수집1				.63		
[문항2] 데이터 이해 및 수집2				.81		
[문항3] 데이터 이해 및 수집3				.65		
[문항5] 데이터 이해 및 수집4				.44		
[문항6] 데이터 분석1	.53					
[문항7] 데이터 분석2	.89					
[문항8] 데이터 분석3	.55					
[문항9] 데이터 분석4	.46					
[문항10] 데이터 평가1						.62
[문항11] 데이터 평가2						.68
[문항12] 데이터 평가3						.71
[문항13] 데이터 도구 활용1			-.61			
[문항14] 데이터 도구 활용2			-.87			
[문항15] 데이터 도구 활용3			-.85			
[문항16] 데이터 도구 활용4			-.84			
[문항17] 데이터 표현1		-.97				
[문항18] 데이터 표현2		-.56				
[문항20] 데이터 기반 의사소통1					-.74	
[문항21] 데이터 기반 의사소통2					-.52	

※ 요인부하량 중 .40 이하는 생략

두 개 요인을 통합하여 ‘데이터 이해 및 수집’으로 명명하였다. 문항부하량이 .30 미만이거나 여러 요인에 .30 이상의 요인부하량이 동시에 나타날 때 문항 제거기준으로 삼아 문항 4(‘나는 주어진 데이터에서 수집 내용, 수집 방법, 대상, 목적 등을 이해할 수 있다’)와 문항 19(‘나는 데이터에 기반하여 합리적인 의사결정을 내릴 수 있다’)를 차례대로 삭제하였다. 6 요인으로 분석된 데이터 리터러시 측정 도구의 설명력은 62.4%로 나타났다.

EFA를 통해 도출된 6요인 간 상관을 분석하였을 때(<표 4> 참조), 상관계수는 유의수준 .05에서 모두 유의한 것으로 분석되었다. 요인간 상관은 .37-.73으로 적절한 상관관

계를 보였다고 해석할 수 있으며, 신뢰도(Cronbach' α)는 모두 .60 이상으로 수용 또는 양호한 수준인 것으로 분석되었다.

3.1.3. 확인적 요인분석(CFA)

탐색적 요인분석에서 최종 선정된 19개 문항에 대해 2차 분석 표본($n_2=1,730$)을 활용하여 확인적 요인분석을 수행하였다. 하위요인별 기술통계와 신뢰도, 상관분석 분석 결과는 <표 5> 및 <표 6>과 같다. 관찰변수의 기술통계 값을 살펴보면 평균은 최소 3.05에서 최대 4.05로 나타났고, 표준편차는 .68-1.03 사이에 분포하였다. 왜도의 절대

<표 4> 데이터 리터러시 하위요인에 대한 기술통계 및 상관분석 결과 ($n_1=1,780$)

하위요인	1	2	3	4	5	6
1. 데이터 이해 및 수집	-					
2. 데이터 분석	.52*	-				
3. 데이터 평가	.54*	.73*	-			
4. 데이터 도구 활용	.41*	.56*	.44*	-		
5. 데이터 표현	.47*	.45*	.38*	.47*	-	
6. 데이터 기반 의사소통	.47*	.57*	.58*	.37*	.37*	-
<i>M</i>	3.78	3.57	3.71	3.15	3.75	3.78
<i>SD</i>	.56	.73	.71	.85	.72	.73
신뢰도	.63	.87	.82	.89	.78	.74

* $p < .05$

<표 5> 문항별 기술통계($n_2=1,730$)

측정변수	최솟값	최댓값	<i>M</i>	<i>SD</i>	왜도	첨도
데이터 이해 및 수집1	1	5	3.66	.83	-.25	-.20
데이터 이해 및 수집2	1	5	3.98	.68	-.54	.91
데이터 이해 및 수집3	1	5	4.05	.68	-.45	.60
데이터 이해 및 수집4	1	5	3.53	.88	-.20	-.39
데이터 분석1	1	5	3.60	.86	-.47	.21
데이터 분석2	1	5	3.50	.94	-.42	-.17
데이터 분석3	1	5	3.72	.85	-.61	.63
데이터 분석4	1	5	3.37	.89	-.18	-.10
데이터 평가1	1	5	3.72	.90	-.68	.42
데이터 평가2	1	5	3.63	.83	-.63	.67
데이터 평가3	1	5	3.67	.87	-.55	.36
데이터 도구 활용1	1	5	3.16	1.02	-.08	-.53
데이터 도구 활용2	1	5	3.05	1.00	.08	-.49
데이터 도구 활용3	1	5	3.15	.99	-.09	-.43
데이터 도구 활용4	1	5	3.05	1.03	.09	-.55
데이터 표현1	1	5	3.76	.84	-.60	.47
데이터 표현2	1	5	3.77	.78	-.43	.29
데이터 기반 의사소통1	1	5	3.85	.82	-.95	1.63
데이터 기반 의사소통2	1	5	3.71	.83	-.54	.49

값은 .08-.95 사이에 분포하고 첨도의 절대값은 .10-1.63 사이에 분포하여 정규성 가정에 문제가 없다고 판단하였다(Curran, West, & Finch, 1996). <표 6>의 상관분석 결과, .37-.72 사이에 분포하여 요인간 적절한 상관을 나타냈으며, 유의수준 .05에서 모두 통계적으로 유의하였다.

CFA 수행에 따른 모형적합도 분석 결과는 다음 <표 7>과 같다. $\chi^2=756.42$, $df=137$, $p<.001$ 로 유의하였으나, χ^2 는 표본 크기에 민감하므로 RMSEA, TLI, CFI 등 다른 적합도 지수를 함께 고려하는 것이 바람직하다(Hair, Black, Babin, & Anderson, 2009). RMSEA=.05, TLI=.96, CFI=.97로 양호한 것으로 분석되었다.

다음 단계로 측정 문항의 요인부하량, 개념신뢰도, 분산추출지수를 분석하여 측정 도구의 구인타당도를 검토하였다(<표 8> 참조). 표준화된 요인부하량은 모두 .50 이상으로, 유의수준 .05에서 유의하였다. 개념신뢰도 분석 결과, .75-.89 사이로 나타나 일반적인 판단 기준인 .70 이상을 충족시켜 양호한 것으로 판단하였다(Anderson & Gerbing, 1988). 평균분산추출(AVE)은 .50 이상이면 양호한 것으로 판단하는데(Fornell & Larcker, 1981), 본 연구에서는 모두 기준을 충족시켜 적합한 것으로 판단하였다. 이와 같은 분석을 바탕으로 6개 요인, 19개 문항으로 구성된 진단 도구는 데이터 리터러시 측정에 타당하다고 판단하였으며, 최종 문항은 <표 9>와 같다.

3.2. 대학 신입생의 데이터 리터러시 수준 분석

A 대학 신입생의 데이터 리터러시 수준 진단 결과에 대한 기술통계 분석 결과는 <표 10>과 같다. 분석 결과 대학 신입생의 데이터 리터러시의 평균은 3.62로 보통 수준을 나타내었고, 응답의 최솟값은 1.08, 최댓값은 5.00이었다. 데이터 리터러시 수준의 분포 경향을 이해하기 위해 사분위 값을 살펴보았으며, 25분위, 50분위, 75분위 값은 각각 3.24, 3.63, 3.96으로 모두 보통 수준에 있음을 알 수 있다.

하위 요인별로 분석한 결과를 살펴보면, 데이터 이해 및 수집과 데이터 기반 의사소통의 평균이 3.78로 가장 높았고, 뒤이어 데이터 표현이 3.76, 데이터 평가 3.69, 데이터 분석 3.56, 데이터 도구 활용 3.13 순으로 나타났다. 분위 값을 살펴보면, 데이터 이해 및 수집, 데이터 분석, 데이터 평가의 경우 25분위 값과 50분위 값이 보통, 75분위 값이 높은 수준을 나타냈다. 또한 데이터 이해 및 수집의 경우 25분위 값이 가장 높았으며, 25분위 값과 75분위 값의 차이가 가장 적은 요인이기도 했다. 데이터 도구 활용 요인의 경우 평균값도 낮을 뿐 아니라 6개 요인 중 유일하게 25분위 값이 2.50으로 매우 낮았고, 75분위의 값도 3.75로 비교적 낮은 수준이라 특별한 관심을 가질 필요가 있다. 데이터 도구 활용 요인은 25분위 값(2.50)과 75분위 값(3.75)의 차이가 1.25로 6가지 요인 중 가장 큰 차이를 나타냈다. 데이터 표현의 경우 25분위 값은 3.00으로 보통

<표 6> 데이터 리터러시 잠재변인별 기술통계 및 요인 간 상관분석 결과($n_2=1,730$)

잠재변인	1	2	3	4	5	6
1. 데이터 이해 및 수집	-					
2. 데이터 분석	.55*	-				
3. 데이터 평가	.52*	.72*	-			
4. 데이터 도구 활용	.44*	.60*	.42*	-		
5. 데이터 표현	.52*	.46*	.37*	.45*	-	
6. 데이터 기반 의사소통	.52*	.59*	.58*	.42*	.40*	-
평균	3.78	3.55	3.67	3.10	3.76	3.78
표준편차	.57	.74	.75	.87	.73	.74
신뢰도	.64	.86	.83	.88	.78	.77

* $p<.05$

<표 7> 데이터 리터러시 모형적합도 분석 결과

구분	χ^2	df	p	RMSEA (90%신뢰구간)	TLI	CFI
모형	756.42	137	<.001	.05 (.05-.06)	.96	.97
참고기준	-	-	.05이상	.06이하	.95이상	.95이상

수준이었지만 50분위 값과 75분위 값이 4.00으로 높은 의사소통의 경우 25분위 값이 3.24, 50분위 값이 3.63, 수준을 나타냈다. 데이터 표현의 경우 50분위 값에서 높은 75분위 값이 3.96으로 모두 보통 수준을 나타내었다. 음의 수준을 나타낸 유일한 하위 요인이다. 데이터 기반

〈표 8〉 확인적 요인분석 결과와 요인별 개념 신뢰도, 평균분산추출(AVE)

잠재변인	측정변수	B	β	S.E.	C.R.	개념신뢰도	AVE
데이터 이해 및 수집	데이터 이해 및 수집1	1.00	.71	-	-	.80	.50
	데이터 이해 및 수집2	.93	.80	.03	28.63***		
	데이터 이해 및 수집3	.87	.75	.03	27.42***		
	데이터 이해 및 수집4	.84	.56	.04	20.86***		
데이터 분석	데이터 분석1	1.00	.82	-	-	.87	.62
	데이터 분석2	1.08	.81	.03	38.09***		
	데이터 분석3	.97	.80	.03	37.31***		
	데이터 분석4	.89	.71	.03	31.71***		
데이터 평가	데이터 평가1	1.00	.75	-	-	.83	.62
	데이터 평가2	1.02	.84	.03	33.76***		
	데이터 평가3	1.00	.78	.03	31.52***		
데이터 도구 활용	데이터 도구 활용1	1.00	.60	-	-	.89	.67
	데이터 도구 활용2	1.44	.87	.05	27.14***		
	데이터 도구 활용3	1.47	.90	.05	27.63***		
	데이터 도구 활용4	1.48	.88	.05	27.20***		
데이터 표현	데이터 표현1	1.00	.76	-	-	.78	.64
	데이터 표현2	1.02	.84	.04	26.75***		
데이터 기반 의사소통	데이터 기반 의사소통1	1.00	.77	-	-	.75	.60
	데이터 기반 의사소통2	1.04	.78	.04	27.22***		

*** $p < .001$

〈표 9〉 최종 데이터 리터러시 측정 도구

번호	하위요인	문항
1	데이터 이해 및 수집	자료 수집 전에 자료 수집을 통해 확인해야 하는 내용을 가설의 형태로 정리할 수 있다.
2		내 가설을 뒷받침 할 수 있는 데이터가 무엇인지 판단할 수 있다.
3		문제 해결에 필요한 자료를 검색하여 적절한 형태로 수집할 수 있다.
4		나는 자료에서 이상치, 누락된 값이 있을 경우 잘 발견해내는 편이다.
5	데이터 분석	나는 데이터로부터 표나 그래프를 만들기 위한 기준 항목을 찾아낼 수 있다.
6		나는 다양한 방법(평균/분포 계산, 표/그래프로 변형 등)을 활용하여 데이터를 탐색할 수 있다.
7		나는 표, 차트, 그래프 등으로 표현된 자료의 의미를 정확히 해석할 수 있다.
8		나는 데이터를 다양한 측면에서 해석하여 또다른 의미를 발견할 수 있다.
9	데이터 평가	나는 데이터의 출처와 수집과정을 확인하여 믿음직한 자료인지 비판적으로 판단한다.
10		데이터를 분석할 때 오류가 있는지 점검하고, 이를 바로잡을 수 있다.
11		나는 제시된 자료에서 해석의 왜곡이나 과장이 없는지 비판적으로 평가할 수 있다.
12	데이터 도구 활용	나는 계산기, 소프트웨어(엑셀, 통계 프로그램 등), 앱 등을 이용하여 정보해석의 정확도를 높일 수 있다.
13		나는 계산기, 소프트웨어(엑셀, 통계 프로그램 등), 앱 등을 이용하여 많은 양의 정보를 빠르게 해석할 수 있다.
14		나는 계산기, 소프트웨어(엑셀, 통계 프로그램, 앱 등)를 이용하여 자료의 의미를 도출해 낼 수 있다.
15		나는 컴퓨터 소프트웨어(엑셀, 통계 프로그램 등)를 이용하여 많은 양의 정보를 효율적으로 관리할 수 있다.
16	데이터 표현	나는 표, 차트, 그래프 등 여러 시각화 방식의 특징을 이해하고 있다.
17		나는 여러 시각화 방식 중 데이터의 내용과 전달 목적에 가장 적합한 방식을 선택할 수 있다.
18	데이터 기반 의사소통	나는 데이터를 활용하여 나의 주장을 객관적으로 뒷받침할 수 있다.
19		나는 데이터를 이용하여 나와 생각이 다른 사람을 효과적으로 설득할 수 있다.

3.3. 집단에 따른 데이터 리터러시 차이 검증

3.3.1. 성별에 따른 데이터 리터러시

성별에 따라 데이터 리터러시에 차이가 있는지 알아보기 위해 독립표본 t-검정을 실시하였고 결과는 <표 11>과 같다.

남학생의 데이터 리터러시 평균은 3.60, 여학생의 평균은 3.64로 여학생의 데이터 리터러시가 남학생보다 근소하게 높았으나 이 차이는 통계적으로 유의하지 않았다. 하위요인별로 살펴보았을 때 데이터 이해 및 수집, 데이터 표현, 데이터 기반 의사소통의 경우 여학생이 남학생보다 통계적으로 유의하게 높은 수준을 보였으며, 데이터 도구 활용의 경우 남학생이 여학생보다 통계적으로 유의하게 높은 것으로 분석되었다.

3.3.2. 계열에 따른 데이터 리터러시

계열에 따라 데이터 리터러시 수준에 차이가 있는지 알아보기 위하여 일원배치분산분석을 실시하였고 결과는 <표 12>와 같다. 계열에 따라 데이터 리터러시 전체 평균과 하위요인 평균에 통계적으로 유의한 차이가 있었다. 데이터 리터러시 전체 평균은 자연 계열과 공학 계열 학생의 평균이 3.68로 가장 높았으며 뒤이어 인문·사회 계열 학생의 평균이 3.56이었고, 예체능 계열 학생의 평균이 3.44로 가장 낮았다.

계열 간 차이에 대해 Duncan의 사후분석을 실시한 결과, 자연 계열과 공학 계열의 점수가 다른 계열에 비해 유의하게 높았고, 그다음으로 인문·사회 계열 점수가 높았으며, 예체능 계열이 다른 계열 대비 데이터 리터러시 수준이 유의하게 낮게 분석되었다. 하위요인별로 살펴보면, 데이터 이해 및 수집, 데이터 분석, 데이터 기반 의사소통 요인이 경우 점수는 자연 계열(b), 공학 계열(c), 인문·

<표 10> 신입생 데이터 리터러시 수준 기술통계 분석 결과 (n=3,510)

구분	M	SD	최솟값	최댓값	25분위	50 분위	75 분위
데이터 리터러시(전체)	3.62	.55	1.08	5.00	3.24	3.63	3.96
데이터 이해 및 수집	3.78	.56	1.25	5.00	3.50	3.75	4.00
데이터 분석	3.56	.74	1.00	5.00	3.00	3.50	4.00
데이터 평가	3.69	.73	1.00	5.00	3.33	3.67	4.00
데이터 도구 활용	3.13	.86	1.00	5.00	2.50	3.00	3.75
데이터 표현	3.76	.73	1.00	5.00	3.00	4.00	4.00
데이터 기반 의사소통	3.78	.73	1.08	5.00	3.24	3.63	3.96

<표 11> 성별에 따른 대학 신입생 데이터 리터러시 수준 분석 결과

구분	N		M	SD	평균차이(남-여)	t
데이터 리터러시(전체)	남	1,797	3.60	.57	-.04	-1.93
	여	1,713	3.64	.53		
데이터 이해 및 수집	남	1,797	3.75	.58	-.07	-3.86*
	여	1,713	3.82	.55		
데이터 분석	남	1,797	3.56	.76	.00	.14
	여	1,713	3.56	.72		
데이터 평가	남	1,797	3.68	.74	-.03	-1.27
	여	1,713	3.71	.71		
데이터 도구 활용	남	1,797	3.17	.87	.09	2.95*
	여	1,713	3.08	.84		
데이터 표현	남	1,797	3.70	.76	-.12	-5.31*
	여	1,713	3.82	.68		
데이터 기반 의사소통	남	1,797	3.75	.76	-.07	-2.87*
	여	1,713	3.82	.70		

* $p < .05$

〈표 12〉 계열별 대학 신입생 데이터 리터러시 수준 사후 분석 결과

구분	N		M	SD	F	Duncan
데이터 리터러시(전체)	인문·사회(a)	1,198	3.56	.57	22.51*	d<a<b,c
	자연(b)	668	3.68	.53		
	공학(c)	1,355	3.68	.54		
	예체능(d)	289	3.44	.56		
데이터 이해 및 수집	인문·사회(a)	1,198	3.78	.58	7.34*	d<a,b,c
	자연(b)	668	3.81	.54		
	공학(c)	1,355	3.80	.56		
	예체능(d)	289	3.64	.56		
데이터 분석	인문·사회(a)	1,198	3.67	.75	5.02*	d<a,b,c
	자연(b)	668	3.74	.72		
	공학(c)	1,355	3.72	.71		
	예체능(d)	289	3.57	.72		
데이터 평가	인문·사회(a)	1,198	3.47	.77	24.83*	d<a<b,c
	자연(b)	668	3.64	.72		
	공학(c)	1,355	3.65	.69		
	예체능(d)	289	3.33	.74		
데이터 도구 활용	인문·사회(a)	1,198	2.97	.86	37.15*	a,d<b<c
	자연(b)	668	3.15	.82		
	공학(c)	1,355	3.30	.85		
	예체능(d)	289	2.93	.85		
데이터 표현	인문·사회(a)	1,198	3.70	.75	12.96*	d<a<b,c
	자연(b)	668	3.86	.69		
	공학(c)	1,355	3.79	.70		
	예체능(d)	289	3.59	.76		
데이터 기반 의사소통	인문·사회(a)	1,198	3.75	.76	10.08*	d<a,b,c
	자연(b)	668	3.87	.71		
	공학(c)	1,355	3.80	.71		
	예체능(d)	289	3.60	.75		

* $p < .05$

사회 계열(a), 예체능 계열(d) 순으로 나타났고, 사후분석 결과 예체능 계열이 타계열보다 통계적으로 유의하게 낮은 것으로 분석되었다($d < a, b, c$). 데이터 평가와 데이터 표현의 경우, 사후분석 결과 예체능 계열이 타계열에 비해 유의하게 가장 낮았고, 인문·사회 계열이 자연 계열 및 공학 계열보다 유의하게 낮았다($d < a < b, c$). 데이터 도구 활용의 경우, 사후분석을 실시한 결과, 공학 계열이 3.30으로 타 계열보다 유의하게 높게 나타났고, 자연 계열이 그 다음으로 3.15이었으며, 인문·사회 계열(2.97)과 예체능(2.93)이 공동으로 타계열보다 유의하게 낮았다($a, d < b < c$).

4. 논의 및 결론

본 연구는 데이터 리터러시 개념틀 및 측정 도구의 타당성을 검증하고 대학 신입생의 데이터 리터러시를 측정하여 교양교육에서의 함의점을 탐색하는 데 목적을 두고 수행되었다. 이를 위해 먼저 데이터 리터러시에 관한 문헌 고찰을 통해 데이터 리터러시 개념틀을 구조화하고 그에 따라 측정 도구를 구성하였다. 이를 바탕으로 수도권 소재 A 대학 신입생을 대상으로 자료를 수집하여 개념틀과 도구의 신뢰성, 타당성을 통계적으로 검증하였다. 그리고 진단 결과에 대한 기술통계 분석과 성별, 계열별 평균 비교를 통하여 대학 신입생의 데이터 리터러시 특성을 살펴보았다. 연구 문제에 관한 주요 연구 결과와 논의는 다음과 같다.

첫째, 선행연구 고찰을 바탕으로 데이터 이해, 데이터 분석 계획 및 수집, 데이터 분석, 데이터 평가, 데이터 도구 활용, 데이터 표현, 데이터 기반 의사소통 등 7개 요인으로 구성된 데이터 리터러시 개념틀과 21개의 예비문항을 개발하였다. 연구 결과, ‘데이터 이해’ 요인과 ‘데이터 분석 계획 및 수집’ 요인이 통합되어 ‘데이터 이해 및 수집’으로 명명하였고, 최종적으로 ‘데이터 이해 및 수집’, ‘데이터 분석’, ‘데이터 평가’, ‘데이터 도구 활용’, ‘데이터 표현’, ‘데이터 기반 의사소통’ 등 6개 요인, 19개 문항의 데이터 리터러시 진단 도구를 구성하였다. 모형 적합도, 요인부하량, 개념신뢰도, 평균분산추출 등의 값이 우수하게 나타나, 본 연구에서 구성한 개념틀과 측정 도구는 데이터 리터러시 측정에 타당한 도구로 판단하였다. 특히 도출된 6개의 하위요인은 데이터 리터러시가 데이터를 다루는 수리·통계학의 영역을 넘어 데이터 이해, 수집, 분석과 평가, 활용, 의사소통까지 아우르는 융합적 개념임을 실증적으로 뒷받침하는 결과이다. 이는 데이터 리터러시 교육이 소프트웨어 교육이나 통계 이론과 기술로만 이루어져서는 부족함을 시사한다. 앞에서 언급한 것과 같이 다수 해외 대학은 실제 사회의 문제를 활용한 학제 간 주제로 융합된 맥락에서 데이터 리터러시 교육을 시도하고 있으며, 이를 실현하기 위한 구체적인 교수학습 전략으로 학생들이 관련 주제를 발굴하고 문제해결 과정을 체험하는 문제중심학습 또는 프로젝트기반학습 방식을 활용하고 있다(Brandeis University, 2022; Kell, 2015; Lehman College, 2022). 따라서 기초교양교육에서의 데이터 리터러시 교육과정은 복합적인 실제 문제를 중심으로 비판적 사고방식과 의사소통, 수리·통계적 기술과 도구 활용까지 포괄하는 접근으로 개발할 필요가 있다.

둘째, 본 연구의 경우 선행연구에서 주로 논의된 하위요인 외에 ‘데이터 도구 활용’ 요인을 추가하여 데이터 리터러시 개념틀을 구성하였다는 데 차별점이 있다. 데이터 리터러시 관련 선행연구에서는 데이터 도구 활용 능력을 독립적인 구인으로 제시하기보다는 ‘데이터 도구를 이용하여’라는 표현을 포함하여 데이터 분석 또는 활용을 설명하되(한상우, 2018) 명시적이지는 않았다. 그러나 취업 등이 주요 과제인 대학생에게 있어 적절한 도구를 활용하여 정확하고 효율적인 데이터 분석과 활용 능력을 함양하여 실무 역량을 갖추는 것은 대학생 신분으로서 매우 중요한 발달과제라 할 수 있다. 현재까지 개발된 데이터

리터러시 진단 도구는 중·고등학생을 대상으로 연구가 수행되어 도구 활용에 대한 중요성이 간과되었다는 판단 하에 본 연구에서는 데이터 도구 활용을 독립된 하위요인으로 제시하였다. 실증적으로 본 연구에서는 데이터 도구를 활용하는 능력이 수집, 분석, 평가 요인과 분리되는 독립된 구인임을 요인분석을 통해 확인하였다. 따라서 개발된 도구는 대학생의 데이터 분석 및 관리에 해당하는 기술적 측면의 데이터 리터러시 역량을 보완했다는 데 의의가 있다.

셋째, 대학 신입생의 데이터 리터러시 수준을 진단한 결과, 신입생의 데이터 리터러시 수준이 비교적 높지 않음을 알 수 있었다. 데이터 리터러시 전체 및 6개 하위 요인에서 모두 보통 수준이었으며, 상대적으로 높은 수준의 학생이 위치한 75분위 값에서도 4점을 초과하는 요인을 찾아볼 수 없었다. 이는 대학 신입생을 대상으로 한 데이터 리터러시 교육이 시급하다는 것을 시사한다. 특히 데이터 도구 활용 요인의 경우 평균이 가장 낮고, 분위별 평균값도 다른 요인에 비해 낮으며, 25분위 값과 75분위 값의 차이가 6 요인 중 가장 크게 나타났다. 특히 데이터 도구 활용을 측정하기 위해 사용한 문항(최종 문항 12-16에 해당)에서 데이터 도구의 예시가 통계 프로그램이나 프로그래밍 언어와 같은 전문 도구가 아니라 엑셀과 같은 보편적 도구를 제시했다는 점을 상기할 필요가 있다. 따라서 대학 신입생은 데이터 도구 활용 경험과 능력이 미흡하다고 해석할 수 있으며, 편차도 상대적으로 다른 요인보다 크게 나타나 관련 교육과정에서는 데이터 도구를 활용하여 문제를 해결하는 경험을 제공하는 동시에, 다양한 수준의 학생 역량을 고려하여 과제를 제시하는 등의 노력이 필요하다.

넷째, 성별 데이터 리터러시 분석 결과, 전체 데이터 리터러시의 경우 성별 차이가 유의하지 않았으나, 하위요인에 따라서 유의한 차이를 발견할 수 있었다. 데이터 평가, 데이터 표현, 데이터 기반 의사소통 요인의 경우 여학생이 남학생보다 유의하게 높은 결과를 보였다. 이는 일반적인 의사소통역량 연구에서 여학생이 남학생보다 높은 수준을 나타내는 것과 유사한 현상으로 해석할 수 있다. 데이터 리터러시는 데이터의 분석과 활용을 토대로 하여 타인과 소통하고 설득하는 것을 포함하는 개념이며, 의사소통역량은 정보 활용 능력에 크게 영향을 미치는 요인이라는 것이 선행 연구에서 밝혀진 바 있다(한상우,

김완중, 2016). 그러므로 데이터에 대한 통계적 접근뿐만 아니라 데이터를 활용하여 설명, 반박, 설득하는 것을 훈련할 수 있는 토론 기반 교수-학습 전략을 활용한 교육과정이 데이터 기반 표현 및 의사소통 향상에 도움을 줄 수 있을 것이다.

다섯째, 신입생의 데이터 리터러시 수준 분석 결과 계열별 차이가 드러났다. 자연 계열과 공학 계열의 수준이 대체로 높았으며, 인문·사회와 예체능 계열 신입생의 수준이 낮은 경향이 있었다. 이 현상은 자료 수집이 대학 입학 이후 전공 교과 학습이 이루어지지 않은 상태인 1학년, 3월에 실시되었다는데 주목할 필요가 있다. 이는 대학 입학 초기부터 인문·사회 및 이공계열 간의 차이를 확인한 것으로, 고등학교의 문과와 이과 간 차이를 반영한 결과로 해석된다. 대학에서 학년이 높아지면서 전공 수업을 이수하게 되면, 이공계열과 인문·사회, 예체능 계열 간 격차가 심화될 것으로 예측할 수 있다. 데이터 리터러시가 계열과 무관하게 요구되는 핵심역량임에도 불구하고(교육부, 2020; OECD, 2019), 인문·사회 및 예체능 계열 학생들을 위한 교양교과로서의 데이터 관련 과목을 찾아보기 어렵다. 또한 비이공계열 학생의 경우, 데이터 관련 교과목 수강에 대한 흥미와 가치를 쉽게 인식하기 어렵고, 수리 과목에 대한 두려움 등으로 수강 동기가 낮을 수 있다. 실제로 대학 신입생의 교양교과 수강현황을 비교·분석한 연구에 따르면, 인문·사회 계열 학생은 주로 인문·사회 계열 관련 교양강좌를 중심으로 수강하지만, 이공계열 학생들은 인문계열 교양강좌와 이공계열 교양강좌를 수강하는 특성이 있다고 보고된다(심태은, 이송이, 2015). 그러므로 대학에서의 데이터 리터러시 교육은 문과·이과 학생들이 모두 수강하는 기초 교양교육으로 제공하되, 학생들의 수강 동기와 배경을 고려하여 맞춤형 교육과정을 개발하고, 계열 간 격차가 심해지지 않도록 방안을 모색해야 한다. 또한 신입생 시기뿐만 아니라 매년 데이터 리터러시를 측정하고 관리함으로써 데이터 리터러시 교육의 성과를 분석하고, 교육과정 개발 및 개선에 지속해서 반영할 필요가 있다.

이상의 논의를 통해 교양교육에서 데이터 리터러시 교육을 위한 다음과 같은 시사점을 도출하였다. 첫째, 대학 신입생의 데이터 리터러시 수준을 고려해 보았을 때 데이터 리터러시 교육이 시급하며, 계열별 편차가 유의하므로 데이터 리터러시 교과를 기초교양교과로 편제하여 계열

과 무관하게 학생들이 데이터 리터러시 학습 기회를 얻을 수 있도록 지원해야 한다. 특히 숫자와 관련이 적은 전공의 학생에게도 학생의 관심사, 생활, 전공과의 관련성이 높은 주제와 사례를 제시함으로써 데이터 리터러시 학습의 유용성과 자신과의 관련성을 인식하고 관련 교과에 대한 흥미와 학습 동기를 촉진하는 것이 필요하다. 둘째, 데이터 리터러시는 비판적 사고, 데이터 관련 기술, 데이터 도구 활용, 의사소통 등이 융합적으로 관련된 역량이다. 그러므로 데이터 과학의 이론과 기술을 넘어서는 접근을 시도할 필요가 있다. 이론이 아닌 “현실 세계의 데이터를 이해하고 활용하기 위한 수단 ... 다양한 관점의 새로운 렌즈로 세상을 바라보며 흥미 있는 이슈를 탐구하는 과정에 필요한 도구로서 통계가 필요한 것”과 같이(김혜영, 2020, p.154) 현실 데이터를 수집, 분석, 평가하여 문제를 해결하는 문제기반 또는 프로젝트 기반 학습을 적용한 교육과정 개발을 고려할 수 있을 것이다. 이와 같은 경험 중심 학습은 데이터 리터러시 하위요인 중 데이터 도구 활용이 저조한 현상도 극복하는 데 이바지할 것으로 예상된다.

본 연구의 제한점과 이를 보완하는 후속 연구는 다음과 같다. 첫째, 본 연구에서 개발된 데이터 리터러시 측정 도구가 새로운 표본을 활용한 후속 연구에 활용된다면 측정 도구의 신뢰도와 타당도를 지속적으로 확인하는데 도움이 될 것이다. 둘째, 신입생의 데이터 리터러시 수준과 논의에 A 대학 신입생 자료를 활용하였기 때문에 결과 해석에 유의할 필요가 있다. 대학마다 신입생의 구성과 특성이 다르므로 후속 연구를 통해 다양한 대학에서 신입생을 대상으로 데이터 리터러시 수준 진단과 논의가 활성화되기를 기대한다. 또한 학생의 고등학교 성적이나 활동 내용 등의 배경변인을 포함한다면 신입생의 데이터 리터러시 수준을 해석하고 이해하는데 도움이 될 수 있을 것이다. 셋째, 본 연구는 신입생에게 국한하여 데이터 리터러시를 측정함으로써 학년별, 학년에 따른 계열별 차이는 분석하지 못하였다. 향후 연구에서 재학생 전체를 대상으로 데이터 리터러시를 주기적으로 진단하여 연구를 수행한다면 대학생의 데이터 리터러시를 이해하는 데 이바지할 것이다. 또한 종단적으로 데이터를 구축하여 데이터 리터러시 추이를 분석하고 관련 교육의 효과성을 검증하는 연구를 제안한다.

참고문헌

- 강명구, 김지현(2010). “한국 대학의 학사구조 변화와 기초교양 교육의 정체성 확립의 과제”, *아시아교육연구* 11(2), 327-361.
- 김성현(2020). “비전공 대학생의 코딩교육 현황과 프로그램에 대한 요구조사”, *송실대학교 교육대학원 석사학위논문*.
- 김슬기, 김태영(2022). “예비교사의 데이터 리터러시 역량 증진을 위한 빅데이터 분석 교양강좌의 개발 및 적용”, *정보교육학회논문지* 26(2), 141-151.
- 김태영, 김철민, 윤혜경(2020). “교양강좌로서의 데이터 리터러시 가능성 탐색”, *한국자료분석학회* 22(6), 2475-2487.
- 김혜영(2020). “21세기 핵심 역량 강화를 위한 대학 교양기초교육으로서의 데이터 리터러시(Data Literacy) 분석”, *교양교육연구* 14(6), 147-159.
- 김혜영, 이은하(2019). “대학 교양의 기초교육으로서의 양적 추론(Quantitative Reasoning) 개설 현황과 분석”, *교양교육연구* 13(6), 29-55.
- 배화순(2019). “데이터 리터러시의 사회와 교육적 함의”, *시민교육연구* 51(1), 95-120.
- 손동현(2007). “새로운 교육수요와 교양기초교육”, *교양교육연구* 1(1), 107-123.
- 송유경, 송석리, 김예지, 임철일(2021). “데이터 리터러시 향상을 위한 데이터 기반 토론 수업 모형 및 교수전략 개발 연구”, *교육공학연구* 37(4), 943-982.
- 심태은, 이송이(2015). “인문 · 사회 · 이공계열 대학 신입생의 교양교과과정 수강현황과 학업역량 및 진로적응능력에 관한 연구”, *학습자중심교과교육연구* 25(12), 343-362.
- 유지원(2016). “디지털 시대에 최적화된 협력적 문제해결역량 측정 도구 개발”, *학습자중심교과교육연구* 16(7), 185-214.
- 이현주, 박성희(2022). “대학생의 창의역량 함양을 위해 개발된 교양강좌의 효과분석”, *학습자중심교과교육연구* 22(14), 705-722.
- 정승원, 장현수, 김세준(2020). “한국교양기초교육원의 표준모델을 적용한 4년제 대학 교양기초교육의 현황과 시사점”, *교양교육연구* 14(5), 83-95.
- 한상우(2018). “인문학 기반 데이터 리터러시 개념에 대한 연구”, *정보관리학회지* 35(4), 223-236.
- 한상우(2020). “인문학 기반 데이터 리터러시 모형 설계에 관한 연구”, *정보관리학회지* 37(1), 179-195.
- 한상우, 김완중(2016). “대학생의 의사소통역량이 정보활용능력에 미치는 영향에 관한 연구”, *한국문헌정보학회지* 50(1), 377-394.
- Anderson, J., & Gerbing, D. W.(1988). “Structural equation modeling in practice: A review and recommended two-step approach”, *Psychological Bulletin* 103(3), 411-423.
- Bhargava, R., & D'Ignazio, C.(2015, June). *Designing Tools and Activities for Data Literacy Learners*, Oxford, UK: Webscience: Data Literacy Workshop.
- Börner, K., Maltese, A., Balliet, R. N., & Heimlich, J.(2016). “Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors”, *Information Visualization* 15(3), 198-213.
- Carlson, J., Fosmire, M., Miller, C., & Nelson, M. S.(2011). “Determining data information literacy needs: A study of students and research faculty”, *Libraries Faculty and Staff Scholarship and Research*, Paper 23. Retrieved from https://docs.lib.purdue.edu/lib_fsdocs/23/
- Curran, P. J., West, S. G., & Finch, J. F.(1996). “The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis”, *Psychological Methods* 1(1), 16-29.
- Fontichiaro, K., & Oehrli, J. A.(2016). “Why data literacy matters”, *Knowledge Quest* 44(5), 21-27.
- Fornell, C., & Larcker, D. F.(1981). “Evaluating structural equation models with unobservable variables and measurement error”, *Journal of Marketing Research* 18(1), 39-50.
- Gray, J., Gerlitz, C., & Bounegru, L.(2018). “Data infrastructure literacy”, *Big Data & Society* 5(2), 1-13.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E.(2009). *Multivariate Data Analysis* (7th ed.), Upper Saddle River, NJ: Pearson Education, Inc.
- Hart Research Associates(2015). *Falling short? College learning and career success*, Association of American Colleges and Universities(AAC&U).
- Macy, K. V., & Coates, H. L.(2016). “Data information literacy instruction in business and public health: Comparative case studies”, *International Federation of Library Associations and Institutions Journal* 42(4), 313-327.
- Mandinach, E. B., & Gummer, E. S.(2013). “A systemic view of implementing data literacy in educator preparation”, *Educational Researcher* 42(1), 30-37.
- Newman, D.(2011). *How to Plan, Participate and Prosper in the Data Economy*, Stamford, CT: Gartner.
- OECD(2019). *OECD Learning Compass 2030: A Series of Concept Notes*, Paris: OECD.
- Otto, J. L.(2012). “Assessing and improving data literacy: A study with urban and regional planning students”, *PNLA Quarterly* 76(4), 5-23.
- Rhodes, T.(2010). *Assessing Outcomes and Improving Achievement: Tips and Tools for using Rubrics*, Washington, DC: Association of American Colleges and Universities.
- Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., & Wuetherick, B.(2015). *Strategies and best practices for data literacy education*, Dalhousie University. Retrieved from <https://dalspace.library.dal.ca/handle/10222/64578>
- Swan, K., Vahey, P. J., Rafanan, K., & Stanford, T.(2009). “Challenges to cross-disciplinary curricula: Data literacy and divergent disciplinary perspectives”, In *Annual Conference of the American Educational Research Association*, San

- Diego, CA. Retrieved from <http://www.sri.com/work/publications/challenges-cross-disciplinary-curricula-data-literacy-and-divergent-disciplinary-p>.
- 교육부(2020). “데이터를 해석하는 능력, 데이터 리터러시를 소개합니다”, Retrieved from <https://www.moe.go.kr/boardContents/view.do?boardID=340&boardSeq=81359&lev=0&m=011101>
- Brandeis University(2022, September 23). “The brandeis core”, Retrieved from <https://www.brandeis.edu/arts-sciences/core/requirements/foundational-literacies.html>
- Kell, G.(2015, August 26). “Are you data literate? New courses debut this fall”, *Berkely News*, Retrieved from <https://news.berkeley.edu/2015/08/26/data-science-class/>
- Lehman College(2022, September 23). “Quantitative reasoning”, Retrieved from <https://www.lehman.edu/office-academic-programs/quantitative-reasoning.php>

Validation of Data Literacy Scale for University Students and Analysis of Freshmen Data Literacy

– The Case of A University

Chung, Yoonsook¹, You, Ji Won²

¹Research Professor, Research Center for Higher Education Achievement, Gachon University

²Associate Professor, Gachon University

Abstract

The key technology in the 4th Industrial Revolution era is related to data. Data literacy, which can uncover various problems from raw data, interpret meanings, lead to more effective communication, and allow for people to make more rational decisions, is discussed as an essential skill in modern society. In this context, this study was conducted to consider the need for data literacy education, particularly in the field of liberal arts education, in order to validate a measurement tool for college students, to analyze the data literacy level of freshmen, and to provide implications for the future curriculum. Data were collected from 3,510 freshmen from A University in the metropolitan area, and the validity of the diagnostic tool was verified through exploratory factor analysis and confirmatory factor analysis. As a result, 19 items were finalized, and the instrument showed good reliability and validity. Following this, the level of data literacy of freshmen was analyzed with descriptive statistics, and a t-test and ANOVA were used to see if there was a significant difference in data literacy according to gender and colleges. The results showed that the data literacy of freshmen was not high, and the results were particularly low in data analysis and the utilization of the data analysis tool. There was no significant difference between genders, but there were significant differences among colleges. Data literacy of students in the arts, physical education, the humanities and the social sciences was lower than those of students in the natural sciences and engineering. Based on these findings, the necessity of data literacy education was discussed in liberal arts education, and suggestions were made to develop data analysis capabilities using certain tools.

Key Words: Data Literacy, Information and Communication Technology (ICT), Quantitative Reasoning, Questionnaire, Validation, Liberal Arts Education