

초·중·고등학생 대상 데이터 과학 교육 변인에 대한 메타분석

김선주*, 전석주**

A Meta-Analysis of the Variables in Data Science Education for K-12 Students

Sean-Joo Kim*, Seok-Ju Chun**

이 논문은 서울교육대학교 2022년도 교내연구비 지원에 의하여 연구되었음

요 약

본 연구의 목적은 일반교육으로서 데이터 과학 교육의 방향을 탐색하기 위함이다. 기존 연구들은 특정 대상 및 방법만을 사용하고 있어, 본 연구에서는 이를 종합적으로 분석하기 위해 '메타분석'을 사용하였다. 먼저, 국내에서 초·중·고등학생을 대상으로 한 데이터 과학 교육 연구들을 수집 및 분류하였다. 그 다음, 대상 연구들에서 보이는 데이터 과학 교육의 효과크기를 종합하고, 변인별 효과크기 비교를 통해 관련 요소들을 분석하였다. 그 결과, 58개의 연구가 선정되었고 데이터 과학 교육이 일반교육으로서도 유의미한 효과가 있는 것으로 나타났다. 특히, 데이터 과학 교육은 초등학교 4, 5학년 학생들에게 큰 효과가 있었으며, 언플러그드 방식과 EPL을 함께 사용하여 교육하는 것이 가장 효과적이었다. 또한, 데이터 과학 교육을 인공지능과 연계하여 학습하는 것도 효과적이었다.

Abstract

The purpose of this study is to explore the direction of data science education as general education. Existing studies only use specific subjects and methods, so in this study, 'meta-analysis' was used to comprehensively analyze them. First, we collected and selected data science education research targeting elementary, middle, and high school students in Korea. Then we synthesized the effect size of data science education seen in the target studies, and analyzed related factors by comparing the effect size for each variable. As a result, fifty-eight studies were selected and data science education was found to have a significant effect as a general education. In particular, data science education had a great effect on students in the 4th and 5th grade of elementary school, and education using the unplugged method and EPL together was most effective. In addition, learning by linking data science education with artificial intelligence was also effective.

Keywords

data science education, machine learning education, meta-analysis, K-12 students

* 서울교육대학교 컴퓨터교육과 박사과정

- ORCID: <https://orcid.org/0000-0002-4224-9117>

** 서울교육대학교 컴퓨터교육과 교수(교신저자)

- ORCID: <https://orcid.org/0000-0003-1299-1203>

· Received: Oct. 26, 2022, Revised: Nov. 30, 2022, Accepted: Dec. 03, 2022

· Corresponding Author: Seok-Ju Chun

Dept. of Computer Education, Teachers College, Seoul National University of Education, 96 Seochojungang-ro, Seocho-gu, Seoul, Korea

Tel.: +82-2-3475-2500, Email: chunsj@snu.ac.kr

I. 서 론

인공지능 기술의 발전과 더불어 그 근간이 되는 데이터는 4차 산업 혁명의 핵심 소재가 되었다. 비단 산업이나 전문분야뿐만 아니라 일반 시민들에게도 데이터를 읽고 쓸 줄 아는 능력은 점차 중요해지고 있다[1]. 현대사회에서 다량의 데이터가 축적되고 접근성이 용이해지면서 일반 시민들에게도 데이터의 활용도와 가치는 날로 증가하고 있기 때문이다[2]. 이에 일반교육의 일환으로 데이터 과학 교육을 실시해야 한다는 요구도 지속적으로 제기되고 있다[3].

이러한 사회적 요구에 맞춰 유아부터 초·중·고등학생들을 대상으로 한 데이터 과학 교육 연구가 국내에서도 점차 활발해지고 있다. 예를 들어, 이현국은 센서보드를 활용한 머신러닝 교육을 개발했고[4], 김민규는 파이썬을 활용한 데이터 시각화 교육 교재를 개발했다[5]. 또한, 김용민은 스프레드시트를 활용한 데이터 과학 교육이 초등학생 컴퓨팅 사고력에 미치는 효과에 대해 연구했고[6], 김정아 외는 언플러그드 방식을 활용한 초등학교 3학년 대상 데이터 시각화 교육 프로그램을 개발했다[7]. 이러한 연구뿐만 아니라 정규교육과정에도 사회적 요구를 반영한 데이터 관련 과목들이 늘어나고 있다. 현재, 초등 교육과정에서는 데이터 과학 교육을 직접적으로 다루고 있지 않으나, 중학교 공통과목 ‘정보’에는 ‘자료와 정보’단원이 있으며 2022 개정 교육과정에서는 고등학교 선택과목에 ‘데이터 과학’ 과목이 추가되었다[8]. 이처럼 일반교육으로서 데이터 과학 교육에 대한 수요와 연구는 점점 증가하고 있는 시점이다.

이에 따라 기존 연구들을 종합하고 체계적으로 분석하여 앞으로의 방향성을 탐색해볼 필요가 있다. 그러기 위해서는 데이터 과학 교육에 영향을 미치는 다양한 변인들을 살펴보고 효과적인 교수 설계 방법을 탐색해야 한다. 이를 위해 본 연구에서는 기존 연구들을 ‘메타분석’하고자 한다. 메타분석이란 동일 주제에 대한 다양한 연구 결과들을 정량적이고 체계적으로 분석하는 통합 분석 방법이다[9]. 본 연구에서는 이를 활용하여 데이터 과학 교육이 교육적으로 얼마나 효과가 있는지 파악하고, 관련 변

인들에 대해 탐색적 분석을 하고자 한다.

II. 관련 연구

데이터 과학이란 데이터에 관해 이론을 개발하거나 과학적 관찰, 시스템적 분석 등을 하는 학문이다[10]. 데이터 과학 교육은 이러한 데이터 과학을 현실 문제 해결을 위해 활용하도록 돕는 교육을 말한다[11]. 최근, 데이터의 광범위한 축적 및 접근성 향상과 함께 일반교육에서도 데이터 과학 교육에 대한 연구가 다양한 형태로 전개되고 있다.

먼저, 교수학습 모형과 방법의 경우 SW와 유사한 형태로 진행되고 있는데, 이는 데이터 과학 교육이 주로 정보 교과에서 진행되고 있기 때문이다[8]. 구체적으로, 교수학습 모형은 한국교육개발원이 발표한 SW교육 교수학습 모형의 형태로 범주화할 수 있는데, 다음과 같은 유형들이 있다[12]. 교사의 설명과 시범을 듣고 학생들이 연습 및 적용하는 시연 중심 모델, 놀이를 통해 SW를 탐색하고 교사의 의도적인 변형으로 사고를 확장시켜 나가는 재구성 중심 모델, 학생들 스스로 SW를 탐구하고 알고리즘을 설계하여 산출물을 개발하는 개발 중심 모델, 프로젝트 학습법과 유사하게 문제 상황에 대한 요구를 분석하고 창의적으로 설계하여 산출물을 생산 및 공유하는 디자인 중심 모델, 마지막으로 컴퓨팅 사고력의 요소들을 문제해결과정 속에서 배우는 CT 중심 모델이 그것이다.

또한, 수업 방법의 경우 데이터 과학을 학습하는 매체에 따라 나눌 수 있는데, 신체활동이나 학습지를 사용하는 언플러그드, 교육용 프로그래밍 언어나 프로그램을 사용하는 EPL(Educational Programming Language), 로봇이나 센서보드를 사용하는 피지컬 컴퓨팅 방법이 있다[13]. 관련 연구를 살펴보면, 언플러그드 방법의 경우 김정아 외는 주사위 게임, 과자 등을[7], 이예지는 활동지를 사용하였다[14]. EPL의 경우 김용민은 스프레드시트를[6], 김민규는 파이썬을[5], 문우중 외는 머신러닝 포 키즈를 사용하였다[15]. 피지컬 컴퓨팅을 살펴보면 이현국은 마이크로비트[4], 안혜지는 햄스터봇[16], 이선진은 엡인벤티를 사용하였다[17].

한편, 데이터 과학 교육은 인공지능 교육과 연계하여 진행되는 경우가 많다. 이는 데이터 과학과 인공지능이 머신러닝(기계학습)이라는 공통항을 가지기 때문이다[10]. 예를 들어, 이은정은 인공지능 기반 데이터 교육이 고등학생에게 미치는 효과에 대해 연구하였고[18], 김봄술은 텍스트 데이터 라벨링 기반 머신러닝 교육 프로그램을 개발했다[19].

이처럼 데이터 과학 교육은 다양한 내용과 방법으로 연구되고 있다. 따라서 본 연구에서는 이상의 내용들을 바탕으로 변인을 설정하고 분석하여, 시사점을 도출하고자 한다.

III. 연구 방법

본 연구에서는 데이터 과학 교육 관련 변인을 살펴보기 위해 다음과 같이 연구를 진행하였다. 먼저, 관련 논문을 수집 및 선정하고, 변인 분석을 위한 코딩을 진행하였다. 그 다음 데이터 과학 교육의 평균 효과크기를 도출하고, 변인별 효과크기 비교를 통해 관련 변인들에 대한 탐색적 분석을 실시하였다. 마지막으로 출판 오류 분석을 통해 전체 결과의 타당성을 검증하였다.

3.1 메타분석 대상 선정

데이터 과학 교육에 관한 연구물을 수집하기 위해 한국학술정보(KISS)과 한국교육학술진흥원(RISS)에서 원문열람이 가능한 자료를 수집하였다. “(데이터 과학 or 데이터) 교육”을 키워드를 사용하였고, 데이터 과학 교육의 내용 중 기계학습에 관한 연구가 주로 인공지능 교육에 포함되어 시행되고 있는 점을 고려하여 “(인공지능 or AI or 기계학습 or 머신러닝) 교육” 또한 키워드로 포함 시켜 검색하였다. 검색 결과 1,509편이 수집되었고, 중복 및 연관 없는 연구를 제외한 뒤, 다음과 같은 선정기준을 바탕으로 최종논문을 선정하였다.

첫째, 효과크기를 계산하기 위한 양적 데이터가 제시되어야 한다. 사전-사후검사 결과의 피실험자수, 평균, 표준편차가 정확히 제시되어 있고, 단일집단 연구의 경우 검정통계치도 제시되어 있어야 한다.

둘째, 검증된 평가도구를 사용한 연구여야 한다. 개별 연구의 오류로 인한 전체 결과의 오류를 방지하기 위해 평가도구의 타당화 및 신뢰도 검증에 대한 내용이 제시되어 있어야 한다.

셋째, 초등학생이나 중학생, 고등학생을 대상으로 한 연구여야 한다. 본 연구는 전공교육이 아닌 기본 교육으로서의 데이터 과학 교육의 효과를 분석하기 위함으로 성인 대상 연구는 제외하였다.

이와 같은 기준으로 32편의 논문이 최종선정되었다. 이때, 한 편의 논문에서 2개 이상의 사례에 대해 보고하고 있는 경우, 각각의 사례를 독립된 분석대상으로 분류하여 58개의 사례를 최종 분석대상으로 선정하였다.

3.2 자료의 코딩

본 연구에서는 데이터 과학 교육에 미치는 변인을 추출하기 위해 2인 이상의 연구자가 함께 변인에 대해 코딩하고 검증하였다. 컴퓨터 교육 전공 박사과정 1인과 석사과정 1인이 교차 검증하였고, 불일치 항목에 대해서는 컴퓨터 교육 박사 1인의 자문을 통해 협의하였다.

표 1. 연구변인에 대한 코딩
Table 1. Coding for variables

Category		Item
Learner factor	Age	Elementary school(1st-3rd), Eementary school(4th-6th), Middle school, High school
		Demonstration, Reconstruction, Development, Design, CT elements
Common teaching and learning factors	Teaching model	
	SW education method	Unplugged, EPL, Physical computing
Teaching and lLearning factors related to data education	Data collection method	Teacher-provided, Searched, Self-produced
	Artificial intelligence connection	Relevant, not relevant

도출된 변인들은 크게 학습자 요인, 일반적 교수 학습 요인, 데이터 교육 관련 교수학습 요인으로 나뉘며, 하위 요인들은 다음과 같다. 학습자 요인 중 학습자 연령은 초등학생을 1-3학년과 4-6학년으로 세분화하였다. 일반적 교수학습 요인의 교수학습 모형과 방법은 데이터 과학 교육이 주로 정보교육의 영역에서 이뤄지고 있는 점을 감안하여 한국교육개발원이 개발한 SW교육 교수학습 모형[12]과 방법을 기준으로 범주화하였다. 데이터 교육 관련 변인은 데이터 과학 교육 시 교육 효과에 차이를 유발할 것이라고 예상되는 데이터 수집방법과 인공지능 연계 여부로 범주화하였다. 세부사항은 표 1과 같다.

3.3 자료 분석

3.3.1. 효과크기 산출

효과크기(ES, Effect Size)는 변수 간 관계의 크기 또는 프로그램의 효과 크기를 표현한 값이다[20]. 본 연구에서는 데이터 과학 교육이 학생들에게 미치는 교육적 효과를 의미한다. 이는 표준편차의 단위로 표시된 값이기 때문에 서로 다른 도구로 측정된 연구들을 비교하는데 유용하다[21].

본 연구에서는 각 연구들의 효과크기를 구하기 위해 표준화된 평균차이(Cohen's d)를 교정한 Hedges's g를 사용하여 계산하였다. Hedges's g는 표본의 크기가 작을 경우 효과 크기가 과대추정 되는 문제를 교정하기 위한 방법이다. Hedges's g(g)는 Cohen's d(d)에 교정지수(J)를 곱하여 구할 수 있는데, Cohen's d(d)는 실험집단의 평균(\bar{X}_1)에서 통제 집단의 평균(\bar{X}_2)을 뺀 값을 통합표준편차(S_p)로 나눈 값이고, 교정지수(J)는 자유도(df) 값을 이용하여 구한 값이다. 그 식은 아래의 (1)과 같다[20].

$$d = \frac{\bar{X}_1 - \bar{X}_2}{S_p}, J = \left(1 - \frac{3}{4df - 1}\right) \quad (1)$$

$$g = J \times d$$

한편, 본 연구의 대상 논문들은 실험집단과 비교 집단을 대상으로 한 연구와 단일집단만을 대상으로 구성되어 있는데, 두 경우 효과크기를 구하는 방법

에 다소 차이가 있다. 전자의 경우 두 집단의 사후 검사 결과를 위 식에 대입하여 효과크기를 구할 수 있다. 반면, 단일집단의 경우 검정통계치(t)로 사전-사후검사 결과의 상관계수(r)를 구한 다음, 이를 이용해 통합표준편차(S_p)를 계산한 뒤 효과크기를 구할 수 있다. 그 방법은 아래의 식 (2)와 같다[22].

$$r = \sqrt{\frac{t^2}{t^2 + df}} \quad (2)$$

$$S_p = \sqrt{\frac{S_1^2 + S_2^2 - (2 \times r \times S_1 \times S_2)}{2(1-r)}}$$

본 연구에서는 이렇게 개별 연구에 대한 효과크기로 전체 연구의 효과크기, 즉 평균 효과크기를 구하였다. 평균 효과크기를 구할 때는 각 사례들을 균등하게 계산하는 ‘고정 효과 모형’과 표본 수가 큰 연구에 더 많은 가중치를 부여하는 ‘랜덤 효과 모형’을 사용할 수 있는데, 본 연구에서는 후자를 사용하였다[23]. 그 이유는 본 연구에서 다른 개별 연구들은 지역, 대상, 연구 기간 등의 차이로 모집단의 효과크기가 다름을 가정하였고, 이 경우 표본이 큰 연구가 신뢰성이 더 높게 나타나기 때문이다[24].

마지막으로, 이렇게 구한 효과크기를 해석하기 위해 본 연구에서는 Cohen이 제시한 해석 기준[25]을 따른다. Cohen은 효과크기가 0.2 이하는 작은 효과크기, 0.2 초과 0.8 미만은 중간 효과크기, 0.8 이상은 큰 효과크기로 해석하였다(표 2).

표 2. Cohen의 효과크기 해석
Table 2. Cohen's effect size interpretation

≤0.2	0.2 ~ 0.8	≥0.8
small effect	medium effect	great effect

3.3.2. 이질성 검증

이질성 검증이란, 각 연구가 추정하는 모집단의 효과크기에 유의미한 차이가 있는지를 검증하는 방법이다. 이는 효과크기의 이질성을 나타내는 지수인 I^2 지수를 사용하는데, 이것은 총 분산에 대한 실제 분산 비율을 의미한다. 일반적으로 I^2 이 50%이상이고, 이질성 검증의 유의확률이 $p < 0.1$ 인 경우 효과크기의 이질성이 크다고 판단한다[26].

이 경우 메타-ANOVA 분석을 활용하여 조절 효과 분석을 실시할 수 있다. 구체적으로, 이질성 지수(I^2)를 구하는 방법은 다음 식 (3)과 같다[26].

$$I^2 = \frac{Q - df}{Q} \times 100\% \quad (3)$$

Q =총분산(실제분산 + 표집오차분산), df =모집단의 효과크기 동일을 가정 한 기대분산

3.3.3. 출판 오류 분석

출판 오류 분석이란, 개별 연구들의 편향된 출간으로 인해 메타분석 전체 연구 결과의 신뢰성과 타당성 문제가 발생하였는지를 분석하는 것이다. 이는 대체로 작은 표본의 연구는 유의미 할 경우에만 출판되는 경향이 있어[27], 작은 표본 연구들이 큰 효과크기를 보이는지 확인하여 오류 여부를 분석한다. 이를 위해 본 연구에서는 Egger의 회귀 분석을 사용하였다. Egger의 회귀분석은 연구의 표준오차와 효과크기의 관계를 회귀식으로 설명하는 방법이다[28]. 회귀분석 결과, 표준오차와 효과크기가 통계적 유의성이 없다($p>.05$)면, 출판 오류가 없음을 의미한다.

IV. 연구 결과

본 연구에서 선정한 58편의 연구들은 초, 중, 고등학생을 대상으로 하며, 다양한 교수학습 모형과 방법으로 데이터 과학 교육을 실시한 사례들이었다. 또한, 대상 연구들은 데이터 과학 교육이 학생들에게 미치는 교육적 효과에 대해 밝히고 있었다. 본 연구에서는 이를 효과크기라는 통합적이고 정량적인 수치로 변환하여 평균 효과크기 및 변인별 효과크기를 구하여 분석하였다. 다만, 초등학생(4-6)학년 대상 연구가 전체 연구의 72%를 차지하고 있어 해석의 주의가 필요하다.

4.1 효과크기 분석

4.1.1. 평균 효과크기

국내에서 시행된 데이터 과학 교육들의 평균 효

과크기는 표 3과 같다. 평균 효과크기(ES)는 0.756, 95% 신뢰구간은 0.638~0.874로 나타났고, 통계적으로 유의($p<.001$)하게 나타났다. 이는 Cohen의 효과크기 해석 기준[25]에 따라 큰 효과 범위(≥ 0.8)에 근접한 수치로, 중간이상의 효과를 보였다고 할 수 있다. 그리고 각 연구들의 효과크기에 대한 이질성 검증 결과, 이질성 지수(I^2)는 84.5%로 이질성이 매우 크게 나왔다. 따라서 메타-ANOVA 분석을 통해 대상 연구들에 대해 조절 효과 분석을 실시하였다.

표 3. 평균 효과크기 및 이질성 검사 결과

Table 3. Mean effect size and heterogeneity test results

k	ES(g)	95% CI		Q (df)	p	I ²
		Lower	Upper			
58	0.756	0.638	0.874	367.04 (57)	<.0001	84.5%

주. k=사례 수, ES(g)=교정된 효과크기(hedges's g), 95% CI=95% 신뢰구간, Q=동질성 검증 통계치, df=자유도, p=유의확률, I²=이질성 지수

4.1.2. 조절 변인에 따른 효과크기

표 4는 연구대상을 연령에 따라 분류한 하위그룹의 효과크기를 보여주고 있다. 고등학생과 초등학생(1~3학년)은 0.8 이상으로 큰 효과크기를 보였고, 그 외의 연령에도 중간 효과크기를 보였다. 그 차이는 통계적으로 유의하였다($p=.001$).

표 5는 일반적 교수학습 요인에 따라 효과크기를 분류하여 분석한 것으로서 교수학습 모형 및 SW교육 방법을 살펴보았다. 그 결과 교수학습 모형의 경우 시연중심 모형, 재구성중심 모형, CT요소중심 모형은 0.8 이상으로 큰 효과크기를 보였고, 개발중심 모형과 디자인중심 모형은 중간 효과크기를 보였다. 이와 관련한 효과크기의 차이는 통계적으로 유의하였다($p=.049$). 반면, SW교육 방법에 따른 차이는 통계적으로 유의하지 않았다. 하지만 언플러그드 & EPL 방법(1.029)이 평균 효과크기(0.756)에 비해 상당히(35%) 높게 나왔다.

표 6은 데이터 교육 관련 교수학습 요인 중 데이터 수집 방법과 데이터 분석 방법에 따라 분류하여 분석한 결과이다. 데이터 수집 방법의 경우 대체로 학생이 직접 제작하거나 검색하는 방법(제공 & 검색 & 제작, 제작, 검색)의 경우 효과크기가 크게 나

16 초·중·고등학생 대상 데이터 과학 교육 변인에 대한 메타분석

왔고, 교사가 제공하는 방법(제공, 제공 & 검색)의 경우 중간 효과크기를 보였다. 그 차이는 통계적으로도 유의하였다($p=0.027$). 인공지능 연계 여부의 경

우 데이터 인공지능과 연계한 경우(0.982)가 그렇지 않은 경우(0.538)보다 더 효과적인 것으로 나왔고, 통계적 차이 역시 유의하였다($p<0.0001$).

표 4. 학습자 요인에 따른 효과크기

Table 4. Effect size according to learner factor

Category	Item	<i>k</i>	<i>ES(g)</i>	95% <i>CI</i>		<i>Q_b</i>	<i>df</i>	<i>p</i>
				<i>Lower</i>	<i>Upper</i>			
Age	High school	5	0.908	0.577	1.239	16.76	3	0.001
	Elementary school(1st-3rd)	5	0.802	0.320	1.284			
	Elementary school(4th-6th)	42	0.779	0.632	0.926			
	Middle school	6	0.449	0.338	0.560			

주. *k*=사례 수, *ES(g)*=교정된 효과크기(hedges's *g*), 95% *CI*=95% 신뢰구간, *Q_b*=하위집단 간 동질성 검증 통계치, *df*=자유도, *p*=유의확률

표 5. 일반적 교수학습 요인에 따른 효과크기

Table 5. Effect size according to general teaching and learning method

Category	Item	<i>k</i>	<i>ES(g)</i>	95% <i>CI</i>		<i>Q_b</i>	<i>df</i>	<i>p</i>
				<i>Lower</i>	<i>Upper</i>			
Teaching model	Demonstration-centered	13	0.850	0.637	1.062	9.55	4	0.049
	Reconstruction-centered	15	0.833	0.582	1.083			
	CT element-centered	11	0.825	0.447	1.198			
	Development-centered	3	0.662	0.423	0.900			
	Design-centered	16	0.535	0.411	0.658			
SW education method	Unplugged & EPL	11	1.029	0.748	1.311	5.27	3	0.153
	Physical computing	9	0.792	0.445	1.138			
	EPL	34	0.668	0.534	0.821			
	Unplugged	4	0.616	0.229	1.003			

주. *k*=사례 수, *ES(g)*=교정된 효과크기(hedges's *g*), 95% *CI*=95% 신뢰구간, *Q_b*=하위집단 간 동질성 검증 통계치, *df*=자유도, *p*=유의확률

표 6. 데이터 교육 관련 교수학습 요인에 따른 효과크기

Table 6. Effect size according to data education related teaching and learning method

Category	Item	<i>k</i>	<i>ES(g)</i>	95% <i>CI</i>		<i>Q_b</i>	<i>df</i>	<i>p</i>
				<i>Lower</i>	<i>Upper</i>			
Data collection method	Teacher-provided & searched & self-produced	6	0.952	0.601	1.302	15.79	7	0.027
	Self-produced	21	0.833	0.590	1.076			
	Searched	14	0.819	0.592	1.045			
	Searched & self-produced	2	0.660	0.016	1.304			
	Teacher-provided & self-produced	8	0.609	0.382	0.836			
	Teacher-provided	2	0.432	0.230	0.700			
	Teacher-provided & searched	4	0.422	0.193	0.650			
Artificial intelligence connection	Relevant	29	0.982	0.811	1.152	16.94	1	<.0001
	Not relevant	29	0.538	0.413	0.663			

주. *k*=사례 수, *ES(g)*=교정된 효과크기(hedges's *g*), 95% *CI*=95% 신뢰구간, *Q_b*=하위집단 간 동질성 검증 통계치, *df*=자유도, *p*=유의확률

4.2 전체 결과의 타당성 검증

본 연구에서는 출판 오류 분석을 통해 전체 결과의 타당성을 검증하였다. 이를 위해 연구들의 표준 오차와 효과크기의 관계를 회귀식으로 나타내는 Egger의 회귀분석을 사용하였다. 그 결과, 본 논문의 대상 연구들로 표준오차와 효과크기가 통계적으로 유의미한 관계가 없음($p=.269$)을 확인할 수 있었다. 즉, 대상 연구들의 편향된 출간으로 인한 메타 분석 전체 연구 결과의 신뢰성과 타당성에 문제가 없음을 확인하였다.

V. 결론 및 논의

본 연구는 데이터 과학 교육 연구 58편의 효과크기를 종합하고 분석하여 데이터 과학 교육 효과에 영향을 미치는 변인들에 대해 살펴보았다. 연구 결과와 시사점은 다음과 같다.

첫째, 국내에서 초·중·고등학생을 대상으로 진행된 데이터 과학 교육의 평균 효과크기(ES)는 0.756으로 중간이상의 효과를 보였다(~ 0.3 : 작은 효과, $0.3\sim 0.8$: 중간 효과, $0.8\sim$: 큰 효과). 이를 통해 데이터 과학 교육이 일반교육으로서도 학생들에게 교육적으로 유의미한 효과가 있음을 확인할 수 있었다.

둘째, 모든 연령에게 데이터 과학 교육은 효과가 있는 것으로 나타났다. 특히, 고등학생과 초등학교(1-3학년)의 경우 큰 효과(각 0.908, 0.802)를 보였다. 데이터 과학 교육이 초등학교 저학년부터 시행되어도 무리가 없음을 알 수 있다.

셋째, 모든 교수학습 모형이 교육적으로 효과가 있었으나, 모형별 효과크기의 차이가 존재하는 것으로 나타났다. 교수학습 모형별 차이를 알아보기 위해서는 구체적인 종속 변인들과의 관계에 대해 연구해 볼 필요가 있다.

넷째, SW교육 방법에 따른 차이는 통계적으로 유의하지 않았으나, 언플러그드 & EPL 방법의 효과(1.029)가 평균 효과크기(0.756)에 비해서 매우 큰 것으로 나타났다. 데이터 과학 교육에 있어서 언플러그드 방법과 EPL 방법을 병행했을 때 보다 효과적인 교육이 이뤄진 원인에 대해 탐색해볼 필요가 있다.

다섯째, 데이터 수집 방법의 경우 교사가 제공해주는 데이터를 사용하는 것보다 직접 제작하거나 검색하여 수집하는 것이 더 효과적이었다. 데이터 수집 과정은 실제 데이터 과학에서도 중요한 부분이며, 이를 직접 경험함으로써 교육적 효과도 향상됨을 시사한다.

마지막으로, 인공지능과 연계하여 교육하는 경우가 그렇지 않은 경우 보다 효과적이었다. 인공지능 기술을 사용해 데이터를 분석해 보는 것이 데이터 과학의 교육적 효과를 높이는 방법이 될 수 있다.

이상으로 데이터 과학 교육에 영향을 미치는 변인들에 대해 분석하고 시사점을 도출하였다. 본 연구는 국내의 데이터 과학 교육을 종합적으로 살펴볼 수 있는 기초자료를 제공한다. 또한, 효과적인 교수학습 설계를 위한 탐색적 근거자료로 사용될 수 있다. 다만, 대상 연구의 상당수가 초등학교 고학년이었던 점을 고려하여, 연령 별 분석도 이뤄진다면 보다 명확한 해석이 가능할 것이다.

References

- [1] L. Pangrazio and N. Selwyn, "Personal data literacies: A critical literacies approach to enhancing understandings of personal digital data", *New Media & Society*, Vol. 21, No. 2, pp. 419-437, Sep. 2019. <https://doi.org/10.1177/1461444818799523>.
- [2] H. W. Lee and S. H. Han, "An analysis of data science curriculum in Korea", *Journal of the Korean Society for Library and Information Science*, Vol. 54, No. 1, pp. 365-385, Feb. 2020. <http://dx.doi.org/10.4275/KSLIS.2020.54.1.365>.
- [3] D. H. Koo and D. J. Kim, "Data science education program based on problem solving learning", *The Journal of Korea Elementary Education*, Vol. 31, No. 102, pp. 203-215, Aug. 2020. <http://dx.doi.org/10.20972/Kjee.31.S.202008.S203>.
- [4] H. G. Lee and I. H. Yoo, "Development and application of supervised learning-centered machine learning education program using micro:bit",

- Journal of The Korean Association of Information Education, Vol. 25, No. 6, pp. 995-1003, Dec. 2021. <http://dx.doi.org/10.14352/jkaie.2021.25.6.995>.
- [5] M. G. Kim, "Development and application of data visualization education using python and their effects : for sixth grade in elementary school", Master Thesis, Jeju National University, Feb. 2020.
- [6] Y. M. Kim and J. H. Kim, "Effect of data science education program using spreadsheet on improvement of elementary school computational thinking", Journal of The Korean Association of Information Education Vol. 21, No. 2, pp. 219-230, Apr. 2017. <http://dx.doi.org/10.14352/jkaie.2017.21.2.219>.
- [7] J. A. Kim, B. S. Kim, T. H. Kim, Y. M. Kim, and J. H. Kim, "The effect of education data visualization using unplugged program on the computational thinking of third grade students", Journal of The Korean Association of Information Education, Vol. 23, No. 4, pp. 283-292, Aug. 2019. <http://dx.doi.org/10.14352/jkaie.2019.23.4.283>.
- [8] Ministry of Education, "Summary of the 2022 Revision Curriculum", Nov. 2021.
- [9] Fred. M. Wolf, "Meta-Analysis: Quantitative Method for Research Synthesis", USA: SAGE Publications, 1986.
- [10] H. V. Jagadish, "Big data and science: Myths and reality", Big Data Research, Vol. 2, No. 2, pp. 49-52, Jun. 2015. <https://doi.org/10.1016/j.bdr.2015.01.005>.
- [11] Y. K. Song, S. R. Song, Y. J. Kim, and C. I. Lim, "A developmental study of an instructional model and strategies for Data-Driven Debate (DDD) to improve data literacy", Journal of Educational Technology, Vol. 37, No. 4, pp. 943-982, Dec. 2021. <http://dx.doi.org/10.17232/KSET.37.4.943>.
- [12] Korea Educational Development Institute, "SW education teaching and learning model development research", CR 2015-35, Dec. 2015.
- [13] Ministry of Education, "Software training operating guidelines", Feb. 2015.
- [14] Y. J. Lee, "Development and application of artificial intelligence convergence education program for elementary school students based on machine learning : focusing on 3rd graders in elementary school", Master Thesis, Chinju National University of Education, Feb. 2022.
- [15] W. J. Moon, J. H. Lee, B. C. Kim, Y. H. Seo, J. A. Kim, J. C. Oh, Y. M. Kim, and J. H. Kim, "Effect of block-based machine learning education using numerical data on computational thinking of elementary school students", Journal of The Korean Association of Information Education, Vol. 25, No. 2, pp. 367-375, Apr. 2021. <http://dx.doi.org/10.14352/jkaie.2021.25.2.367>.
- [16] H. J. Ahn, "Development of AI education program using robots for the middle grades of elementary school", Master Thesis, Jeju National University, Feb. 2022.
- [17] S. J. Lee, "Development and application of artificial intelligence Education program for elementary school students based on recommendation systems", Master Thesis, Jeju National University, Aug. 2018.
- [18] E. J. Lee, "The effects of AI-based data analysis education on convergent thinking ability and data literacy of general high school students", Master Thesis, Jeju National University, Feb. 2022.
- [19] B. S. Kim, "Development and application of the machine learning educational program based on text data labeling", Master Thesis, Jeju National University, Aug. 2021.
- [20] M. L. Smith and G. V. Glass, "Meta-analysis of psychotherapy outcome studies", American Psychologist, Vol. 32, No. 9, pp. 752-760. Oct. 1977. <https://doi.org/10.1037/0003-066X.32.9.752>.
- [21] P. Beeson and R. Robey "Evaluating single-subject treatment research: Lessons learned from the aphasia literature", Neuropsychological Review, Vol. 16, No. 4, pp. 161-169, Dec. 2006. <https://doi.org/10.1007/s11065-006-9013-7>.

- [22] S. D. Hwang, "Meta-analysis using R", Seoul: Hakjisa, Jul. 2015.
- [23] L. V. Hedge and I. Olkin, "Statistical methods for meta-analysis", New York: Academic Press, Jul. 1985.
- [24] M. Borenstein, L. V. Hedges, J. P. Higgins, and H. R. Pothstein, "Introduction to Meta-Analysis Chichester", UK: Wiley, Mar. 2009.
- [25] J. Cohen, "Statistical power analysis for the behavioral sciences", New York and London: Routledge, Jul. 1988.
- [26] Cochrane Handbook for Systematic Reviews of Interventions, <http://handbook-5-1.cochrane.org>. [accessed: Oct. 25, 2022]
- [27] J. A. Sterne, D. Gavaghan, and M. Egger, "Publication and related bias in meta-analysis: power of statistical tests and prevalence in the literature", *Journal of clinical epidemiology*, Vol. 53, No. 11, pp. 1119-1129, Nov. 2000. [https://doi.org/10.1016/S0895-4356\(00\)00242-0](https://doi.org/10.1016/S0895-4356(00)00242-0).
- [28] M. Egger, D. G. Smith, M. Schneider, and C. Minder, "Bias in meta-analysis detected by a simple, graphical test", *BMJ*, Vol. 315, No. 7109, Sep. 1997. <https://doi.org/10.1136/bmj.315.7109.629>.
- [29] M. B. Kim, "The effect of education data visualization using Google spreadsheet program on improvement of creativity : for forth and fifth grade students", Master Thesis, Jeju National University, Feb. 2022.
- [30] M. J. Kim, S. C. Lee, and T. Y Kim, "The effect of software education including data literacy on computational thinking and the creative problem-solving ability of middle school students", *Korean Journal of Teacher Education*, Vol. 37, No. 1, pp. 167-184, Jan. 2021. <https://doi.org/10.14333/KJTE.2020.37.1.08>.
- [31] B. C. Kim, B. S. Kim, E. J. Ko, W. J. Moon, J. C. Oh, and J. H. Kim, "Development of machine learning education program for elementary students using localized public data", *Journal of The Korean Association of Information Education*, Vol. 25, No. 5, pp. 751-759, Oct. 2021. <http://dx.doi.org/10.14352/jkaie.2021.25.5.751>.
- [32] Y. M. Kim, "Data science education program to improve computational thinking and creativity", PhD Thesis, Jeju National University, Aug. 2018.
- [33] Y. M. Kim, "Effect of data science education program using spreadsheet on improvement of elementary school computational thinking", *Journal of The Korean Association of Information Education*, Vol. 24, No. 6, pp. 551-562, Dec. 2020. <http://dx.doi.org/10.14352/jkaie.2019.23.4.283>.
- [34] J. A. Kim, "Computer education program based on data visualization to improve computational thinking and creativity", PhD Thesis, Jeju National University, Aug. 2021.
- [35] H. R. Kim and S. Y. Choi, "Development and application of artificial intelligence STEAM program for real-time interactive online class in elementary science - focused on the unit of 'Life of plant' -", *Journal of Korean Elementary Science Education*, Vol. 40, No. 4, pp. 433-442, Nov. 2021. <https://doi.org/10.15267/keses.2021.40.4.433>.
- [36] Y. M. Do, "Development and application of artificial intelligence literacy education program for vocational high school students", Master Thesis, Korea National University of Education, Feb. 2022.
- [37] J. H. Moon, "A development and application of EPM learning model for artificial intelligence education in elementary schools", Master Thesis, Seoul National University of Education, Aug. 2021.
- [38] H. W. Moon, "The effects of data literacy-based SW convergence education on elementary school student's computational thinking ability", Master Thesis, Korea National University of Education, Feb. 2022.
- [39] S. H. Park, "The effects of data science education using korean Python on elementary students' computational thinking", Master Thesis, Seoul National University of Education, Aug. 2022.

- [40] J. H. Park and J. B. Song, "Case of operating a curriculum-linked artificial intelligence education camp program", *Journal of Digital Contents Society*, Vol. 23, No. 1, pp. 49-56, Jan. 2022. <https://doi.org/10.9728/dcs.2022.23.1.49>.
- [41] J. S. Shin, "Development and implementation of an activity-based AI convergence education program for elementary school student", Master Thesis, Cheongju National University of Education, Feb. 2021.
- [42] S. H. Sim, "A study on the development and effectiveness of python-based artificial intelligence education program in elementary schools: focusing on machine learning", Master Thesis, Daegu National University of Education, Aug. 2021.
- [43] J. S. Lee, "Development of AI education program using robots for the middle grades of elementary school", Master Thesis, Gwangju National University of Education, Feb. 2022.
- [44] C. I. Lim, J. C. Lee, Y. K. Song, U. G. Lee, H. G. Lee, and E. S. Lim, "A study on the development and application of a data science education program for elementary school students", *The Journal of Educational Information and Media*, Vol. 28, No. 2, pp. 319-345, Jun. 2022. <http://dx.doi.org/10.15833/KAFEIAM.28.2.319>.
- [45] D. J. Jang, "Development of a data education program for elementary students using Gummer's data literacy framework", Master Thesis, Daegu National University of Education, Feb. 2022.
- [46] K. M. Jung, "Development of explanatory artificial intelligence education program based on artificial intelligence literacy", Master Thesis, Gyeongin National University of Education, Aug. 2021.
- [47] I. H. Choi and K. H. So, "The effect of a fine dust environmental education program using artificial intelligence-based tools on environmental sensitivity and self-effectiveness of elementary school students", *Journal of Korean Elementary Science Education*, Vol. 41, No. 3, pp. 468-479, Aug. 2022. <https://doi.org/10.15267/keses.2022.41.3.468>.
- [48] M. Y. Han, "Development and application of chatbot-based AI education program", Master Thesis, Gyeongin National University of Education, Feb. 2021.
- [49] S. Y. Hong, S. Han, and H. B. Kim, "Analysis of problem-solving processes through data-based steam education: focusing on atmospheric circulation and surface currents", *Journal of Korean Society of Earth Science Education*, Vol. 13, No. 3, pp. 330-343, Dec. 2020. <https://doi.org/10.15523/JKSESE.2020.13.3.330>.
- [50] S. J. Hong, "The effects of design thinking based artificial intelligence education programs on middle school students' computational thinking and creative problem solving ability", Master Thesis, Korea National University of Education, Feb. 2022.

저자소개

김 선 주 (Sean-Joo Kim)



2015년 8월 : 서울교육대학교
컴퓨터교육과(석사)
2021년 3월 ~ 현재 : 서울교육
대학교 컴퓨터교육과 박사과정
관심분야 : 컴퓨터교육, 특수교육,
프로그래밍 언어

전 석 주 (Seok-Ju Chun)



2002년 2월 : 한국과학기술원
컴퓨터공학과(공학박사)
2004년 3월 ~ 현재 : 서울교육
대학교 컴퓨터교육과 교수
관심분야 : 컴퓨터교육, 프로그래밍
방법, 데이터마이닝, 멀티미디어DB