

ORIGINAL ARTICLE

기상 데이터 분석 과정에서 나타나는 오류의 유형과 맥락 탐색

홍석영

(한국교원대학교부설고등학교 교사)

Exploration of Types and Context of Errors in the Weather Data Analysis Process

Seok-Young Hong

(Affiliated High school to Korea National University of Education)

ABSTRACT

This study explored the errors and context occurred during high school students' data analysis processes. For the study, 222 data inquiry reports produced by 74 students from 'A' High School were collected and explored the detailed error types in the data analysis processes such as data collection and preprocessing, data representation, and data interpretation. The results of study found that in the data interpretation process, students had a somewhat insufficient understanding of seasonal variations and periodic patterns about weather elements. And, various types of errors were identified in the data representation process, such as basic unit in graphs, legend settings, trend lines. The causes of these errors are the feature of authoring tools, misconceptions related to weather elements, and cognitive biases, etc. Based on the study's results, educational implications for big data education, a significant topic in future science education, were derived. And related follow-up studies were suggested.

Key words : weather, big data, data, error, context

I. 서론

미래사회 변화에 대한 전망 중에서는 셀 수 없이 생산되는 빅데이터에 대한 논의가 활발하다. 빅데이터를 활용한 유의미하고 새로운 지식의 생산을 강조하고 있기 때문이다. 우리나라의 2022 개정 교육과정에서도 공학적 도구나 센서를 통한 빅데이터 수집, 실생활의 데이터 기반 탐구 문제 발견, 데이터를 활용한 새로운 지식 생산 등을 다양한 교과 교육 목표로서 설정하

고 있다(교육부, 2022). 이는 빅데이터를 분석하는 과정이 다양한 측면의 역량을 함양하는 데 도움을 줄 수 있기 때문이다. 대표적으로 빅데이터를 수집하고 처리하는 과정을 통해 이와 관련한 컴퓨팅 기술을 이해하거나(정인진 등, 2023), 문제해결에 필요한 사고력을 기를 수 있다(김봉철 등, 2021; 김용민과 김종훈, 2017; 김재경과 손의성, 2021). 나아가 합리적인 추론과 논증을 위한 논거와 자료를 생산하는 것에도 도움을 줄 수 있다(송유경 등, 2021; Zagallo *et al.*, 2016)

과학 교과에서도 데이터를 활용한 다양한 탐구 활동

Received 28 June, 2024; Revised 25 July, 2024; Accepted 14 August, 2024

*Corresponding author: Seok-Young Hong, Affiliated High school to Korea National University of Education, 250 Taeseongtabyeon-ro, Gangnae-myeon, Heungdeok-gu, Cheongju-si, Chungbuk 28173, Korea
E-mail : gfdsl206@naver.com

© The Korean Society of Earth Sciences Education. All rights reserved.
This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

을 주목하고 있다. 데이터를 분석하는 일련의 과정이 과학의 기초·통합 탐구 과정과 유사하기 때문이다. 따라서 과학적 가설의 설정과 검증, 개념에 대한 모델링과 같은 인지적 측면의 발달에 도움을 줄 수 있다(김소현과 유선아, 2023; 정혜윤과 유선아, 2023; Mensan *et al.*, 2020). 전통적으로도 데이터를 분석하는 일련의 과정은 과학 교과와 중요한 탐구 기능으로 중요하게 다루어져 왔다. 특히 실험도구를 통해 측정된 자료(데이터)를 수집하고 변환하는 과정을 강조했으며, 이와 관련해 학생들이 활용하는 인지적 사고, 학습 전략 등을 탐색하는 시도가 있었다. 예컨대 데이터 분석 과정에서 나타나는 자료 변환 과정의 오류를 탐색하거나(오영열과 이윤경, 2022), 데이터를 해석하는 유형이나 데이터 해석에 영향을 끼치는 요인에 대한 연구(Rahmatina & Zaid, 2019) 등이 있다. 이러한 연구들은 학생들이 주어진 데이터를 표현·분석하는 방식을 살펴봄과 과학적 데이터를 다루는 학습에 대한 교수적 시사점을 제공했다는 점에서 의의가 있다. 한편, 지금까지의 연구들은 주로 실험실 상황에서 실험을 통해 생산되는 제한적인 데이터를 대상으로 하거나, 미리 교사가 데이터를 제시한 상황을 살펴보는 사례가 많았다. 하지만 학교 교육 현장에서 다루어야 하는 데이터가 단지 실험실 상황에서 측정을 통해 만들어진 것에 국한되지는 않는다. 따라서 학생들이 데이터를 분석하는 과정에서 나타나는 총체적인 학습의 양태를 살펴보는 데는 한계점을 갖는다.

그리고 기존의 연구에서는 학생들이 다루는 데이터와 관련한 학습 주제에 대한 탐색이 상대적으로 미비한 편이다. 특정한 과학 수업에서 활용하는 데이터는 학습 소재나 개념적 주제와 관련된 것이다. 따라서 특정한 데이터를 분석하는 과정은 데이터와 관련한 개념적 이해나 인지 구조에 영향을 받을 수 있다(Farmer, 2012; Prain & Tytler, 2012; Volkwyn *et al.*, 2020). 하지만 학생들이 데이터를 분석하는 데 활용하는 자원인 주제-특이적인 개념 요소, 오개념의 유형과 같은 특성에 대한 고려가 미비했다는 것이다(Chang *et al.*, 2024; Zagallo *et al.*, 2016). 이러한 요소들은 데이터 분석을 위한 디지털 도구 활용이나, 실시간 데이터 활용 교수 학습을 강조하는 학교 교육과 관련한 전환적 담론 속에서 실질적인 학생들의 학습 과정을 살피기 위해서 중요하게 다루어져야 한다.

한편, 지구과학 교과와 학습 소재인 기상요소와 관

련한 주제에서는 빅데이터를 활용한 다양한 교수학습 방법을 운영할 수 있다. 많은 지역에서 다양한 기상요소들을 관측하고 있으며 관측한 데이터를 지속적으로 누적하기 때문이다. 또한, 공공 누리집을 활용한다면 누구나 공신력 있는 기관의 빅데이터에 접근하는 것이 가능하다. 국내에서도 기후변화 정보 포털, 기상 자료 개방 포털과 같은 공공 기관의 빅데이터를 교육적 목적으로 활용할 수 있다. 또한, 기상요소가 시간에 따라 민감하게 변화하는 데이터이기 때문에 이를 통해 미래의 날씨와 기후변화를 예측하거나 과거의 날씨와 기후변동 등을 분석할 수 있다. 이와 같은 활동은 실제 과학자들의 연구 과정을 유사한 맥락에서 체험할 수 있는 기회를 제공할 수도 있다(Azzam *et al.*, 2013). 따라서 기상요소와 관련한 학습에서 빅데이터를 활용한 탐구는 학생들이 자신의 삶에 비추어 지식을 전이할 수 있는 경험을 제공하면서도 데이터와 관련한 기능적 측면의 이해를 돕는 교수학습방안이 될 수 있다.

이에 본 연구에서는 지구과학 교과와 날씨의 변화와 관련한 학생들의 빅데이터 탐구 과정을 조망하고, 기상 데이터를 분석하는 과정에서 나타나는 다양한 오류 유형을 살펴보고자 하였다. 문제해결과정에서 나타나는 오류는 학생들이 문제에 대해 접근하는 상이한 양식, 데이터 도구의 활용, 표상을 위한 메타적 사고 과정 등을 탐색하기 위한 기초적인 탐색자료가 될 수 있기 때문이다(김유정 등, 2009; Prain & Tytler, 2012; Wong, 2017). 이를 통해 빅데이터를 활용하는 지구과학 교수학습 활동 측면에서의 교육적 시사점을 도출하고자 했다.

II. 연구 방법

1. 연구 참여자

본 연구에서는 기상 데이터를 분석하는 활동에서 나타나는 오류의 유형을 탐색하고자 했다. 이를 위하여 지방 중·소도시에 위치한 A 고등학교 2학년 74명의 지구과학 I을 수강하는 학생을 연구 참여자로 선정하였다. 데이터 분석을 위해서는 이와 관련한 기초적 소양을 갖추고 있는 대상을 선정하는 것이 필요하기 때문이다. A 고등학교는 과학 교과 관련 교육과정 특화학교를

운영 중이기에 학생들이 수학, 과학, 정보 교과에 대한 다양한 교과를 이수할 수 있도록 하였다. 연구에 참여한 학생들은 1학년에서 정보 교과를, 2학년에서 수학 I, 수학 II 교과를 수강하면서 데이터 분석을 위해 필요한 데이터 수집 방법 및 기초적인 수학적 내용 요소에 대해 학습하였다. 또한, 지구과학 I 교과에서 날씨의 변화에 대한 내용을 학습한 학생들이다. 연구 참여 학생들의 학업 성취도는 주로 중상위권이며, 과학 교과 관련 활동에 적극적으로 참가하는 학생들이다.

2. 데이터 분석 활동 수행 및 학생 자료 수집

본 연구에서는 지구과학 I 교과 수업 중 개별 탐구활동으로 기상 빅데이터 분석 활동을 진행하고 학생들이 생산한 자료를 수집하였다. 이를 위해 학생들은 활동에 앞서 Google spreadsheet, Flourish studio(<https://fourish.studio>)와 같은 온라인 데이터 처리 및 시각화 저작 도구 활용 방법을 학습하였다. 또한, 기초적인 형태의 산점도 및 시계열 그래프의 작성 방법, 데이터에서 나타나는 상관관계 해석과 추세선 작성 방법을 학습하였다. 나아가 공공 기관의 데이터를 수집하는 다양한 방법을 소개했다.

이후 6차시 동안의 활동에서 공신력 있는 기상 데이터를 수집하고 기상요소의 시계열적 변화, 기상요소 데이터 간의 상관관계 등을 해석하는 활동을 진행하였다. 이 과정에서 온라인 저작 도구를 활용하여 직접 데이터를 정리하고 시각화 자료로 표현하도록 했다. 나아가 수집한 기상 데이터와 관련이 있을 것이라고 생각하는 다양한 종류의 데이터를 수집하고 여러 가지 형태로 표현하면서 기상요소와의 관계를 분석하도록 요청하였다. 학생들은 매 차시 활동을 진행하며 해당 차시에 수행한 내용을 탐구 노트에 정리했으며, 최종적으로 약 3~8페이지 분량의 기상 데이터 분석 보고서를 작성했다.

연구 참여 학생들이 작성한 내용은 연구 이외의 다른 용도로 활용되지 않는다고 알렸으며, 탐구 활동 이후 학생들의 오류에 대한 평정을 통해 오류가 나타난 부분을 다시 학습할 수 있는 기회를 제공하였다.

3. 자료 분석틀 개발 및 자료 분석

학생들이 생산한 기상 데이터 분석 결과와 탐구 노

트를 종합하여 오류의 유형을 분류하고자 하였다. 이를 위한 기초작업으로 교육 분야에서 데이터 분석, 온라인 도구를 활용한 데이터 처리 등과 관련한 국내·외 관련 문헌을 수집하여 데이터 분석 과정과 세부 절차를 파악하였다.

연구에 따라 학생들의 빅데이터 분석 과정을 다양한 수준으로 구분하고 있지만, 공통적인 과정으로 데이터 수집, 데이터 처리, 데이터 표현, 데이터 해석과 같은 과정에 주목하는 것이 특징이다(송유경 등, 2021; 지영명, 2020; Binali *et al.*, 2022; Chang *et al.*, 2024; Rahmatina & Zaid, 2019). 이후, 각 데이터 분석 과정에서의 학생 인식 유형, 오류, 편향, 어려움 등과 관련한 연구를 수집하며 학생들이 데이터 기반 탐구 활동에서 수행하는 주요 절차에 대해 포괄적으로 정리하였다.

이 과정에서 데이터 처리 과정에 해당하는 세부적인 절차가 데이터 수집 과정과 밀접한 관련성을 갖고 있으며 중복되는 절차가 많다는 것을 확인하였다. 따라서 두 과정을 ‘데이터 수집 및 처리’ 과정으로 병합하였다(Chai, 2020). 최종적으로 데이터 수집 및 처리, 데이터 표현, 데이터 해석과 같은 세 가지 데이터 분석 과정을 설정하고, 각 과정에 해당하는 세부 절차를 정리하였다(Table 1).

데이터 수집 및 처리 과정은 학생들이 온라인 도구나 앱, 센서 등을 활용하여 직접 탐구를 수행한 연구에서 주목한 과정이다. 이는 공신력 있는 데이터 출처 파악, 데이터 수집, 데이터의 측정 및 수집 신뢰도 확보, 데이터의 범주와 특성, 데이터 전처리 등의 세부 절차를 포함한다. 데이터 표현 과정은 다양한 데이터에 대한 적절한 시각화 유형 이해, 데이터 저작 도구 사용 방법 이해, 경향을 파악하기 위한 추세선·예측선의 활용, 시각화 자료의 세부적인 요소 표현, 데이터의 대푯값 표현 등과 관련한 요인과 같은 세부 절차로 구성된다. 데이터 해석 과정은 데이터의 기술적(양적) 해석, 경향성과 패턴 해석, 유의미한 정보 도출, 데이터에서 나타나는 해석 결과인 정보에 대한 이해, 데이터에 기반한 결론의 정당화 등의 세부 절차를 포함한다.

한편, 정리한 데이터 분석 과정은 본 연구에서 다루고 있는 수업 소재와 수업의 맥락적 특성을 반영하기 보다는 데이터를 수집·표현·해석하는 과정에서의 접근 방법이나 기능적 측면을 포괄적으로 제시한 것이다. 따라서 본 연구의 탐구 소재와 탐구 과정의 구체적인

Table 1. Inquiry procedures on the data analysis process and references

데이터 분석 과정	세부 탐구 절차	선행 연구
수집 및 처리	<ul style="list-style-type: none"> • 데이터 수집 • 데이터의 신뢰도 파악 • 데이터의 성격과 유형 이해 • 데이터 전처리 (결측치, 이상치 등의 정제) 	<ul style="list-style-type: none"> • 김명신 등(2003); 김유정 등(2009); 김지나와 변영찬(2007); 김형욱(2024); 오영열과 이윤경(2022); Börner et al.(2019); Chai (2020); Chang et al.(2024); Gobert et al.(2018)
표현	<ul style="list-style-type: none"> • 다양한 데이터의 시각화 유형 이해 • 데이터 저장 도구 사용 방법 이해 • 추세선, 예측선의 활용 • 시각화 자료(그래프, 표)의 요소 표현 • 데이터의 대푯값 표현 	<ul style="list-style-type: none"> • 김명신 등(2023); 김지나와 변영찬(2007); 김유정 등(2009); 김형욱(2024); 양수진과 장명덕(2012); 최한나와 이종희(2023); Åberg-Bengtsson & Ottosson(2006); Börner et al.(2019); Chang et al.(2024); Glazer(2011); Maltese et al.(2015); Pols et al.(2021); Rahmatina & Zaid(2019);
해석	<ul style="list-style-type: none"> • 데이터의 기술적(양적) 해석 • 데이터의 경향성 및 패턴 해석 • 유의미한 정보 도출 • 시각화 자료(결과, 정보)에 대한 이해 • 데이터에 기반한 결론의 정당화 	<ul style="list-style-type: none"> • 김지나와 변영찬(2007); 김유정 등(2009); 김형욱(2024); 마민영 (2018); 양수진과 장명덕(2012); 최한나와 이종희(2023); Chang et al.(2024); Glazer, 2011; Gobert et al.(2018); Maltese et al.(2015); Pols et al.(2021); Zagallo et al.(2016)

Table 2. Categories and types of errors in the data analysis process

데이터 분석 과정	범주	오류 유형
수집 및 처리	• (데이터 탐색을 위한) 잠정적 가설 설정	<ul style="list-style-type: none"> • 잘못된 개념을 활용한 가설 설정 • 편향적 가설 설정 • 가설의 타당성 미흡
	• 데이터 수집 (자료 활용)	<ul style="list-style-type: none"> • 분석 범위에 적합하지 않는 데이터 수집 • (기간, 지역 등) 조건이 다른 데이터 수집
	• 데이터 형식 이해 (데이터 범위 이해, 결측치, 이상치 등의 전처리)	<ul style="list-style-type: none"> • 범주형 데이터 처리 미흡 • 결측 처리 미흡 • 분석 단위 변환 미흡
표현	• 시각화 형식 (다양한 범주에 적절한 시각화 표현)	<ul style="list-style-type: none"> • 과도한 데이터 포함 • 물리량 표현에 부적절한 시각화 유형
	• 경향성 표현 (추세선, 예측선의 활용)	<ul style="list-style-type: none"> • 경향 및 추세 표현 누락 • 잘못된 추세선의 활용
	• 그래프 요소 (변수, 단위, 눈금 등)	<ul style="list-style-type: none"> • 그래프의 축 간격 • 변수 표현 • 단위 표현 • 변수의 특성 고려 미흡
	• 통계적 표현 (평균, 표준 편차 등)	<ul style="list-style-type: none"> • 기술적 표현 누락
해석	• 기술적(양적) 해석 (평균, 극값, 중위, 편차, 분산 등)	<ul style="list-style-type: none"> • 데이터의 특징을 정성적으로만 비교한 경우 • 자료와 다르게 해석한 경우
	• 경향성 및 패턴 해석 (내삽, 외삽, 추세, 주기성 등의 관계 해석)	<ul style="list-style-type: none"> • 상관관계와 비례관계 표현 혼동 • 함수관계에 대한 해석 미흡 • 진동 형태에 대한 해석 미흡 • 자료와 다르게 해석한 경우 • 경향 및 관계를 정성적으로만 해석한 경우
	• 유의미한 정보 도출 (과학적으로 타당한 해석)	<ul style="list-style-type: none"> • 인과관계와 상관관계의 혼동 • 인과적 관계에 있는 변수 파악 미흡 • 현재 추세, 특정한 값에 기반한 과도한 일반화
	• 정보 (그래프 해석 결과)에 대한 설명 (가설의 변경, 새로운 가설의 생성)	<ul style="list-style-type: none"> • 변수에 영향을 주는 요인에 대한 고려 미흡 • 이상적 상황(실험실 상황)과의 비교
	• 분석 결과 및 결론의 정당화 (논증)	<ul style="list-style-type: none"> • 잘못된 개념에 기반한 논증 • 잘못된 자료에 기반한 논증 • 데이터를 바탕으로 한 단정적 진술 • 연역적 정당화 및 논증 오류 • 자료를 제시하지 않고 주장

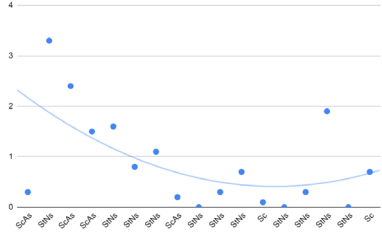
학생 자료	과정	범주	오류 유형
 <p>... 이 그래프를 보면 강수량과 구름 모양의 상관관계를 알 수 있다. 4월 15일 오전 9~20시까지, 16일 오전 5~9시까지 낮은 강수량이 나타나는데 이것이 앞에서 설명한 맑은 날씨에 영향을 주는 하층운과 상관관계가 있다. ...</p>	수집 및 처리	데이터 형식 이해	범주형 자료 처리 미흡
	표현	시각화 형식	물리량 표현에 부적절한 시각화 유형
		그래프 요소	단위 표현
	해석	기술적 해석	데이터의 특징을 정성적으로만 비교한 경우
		유의미한 정보 도출	인과관계와 상관관계의 혼동

Fig. 1. An example of an analysis about errors in the data analysis process

양상을 드러내지 못하는 절차가 포함되어 있었다. 대표적으로 전체 연구 자료에서 나타나지 않은 데이터의 신뢰도 확보와 관련한 절차가 있다. 본 수업 활동에서는 데이터를 수집하는 자료원으로 공신력 있는 국가 데이터 포털을 활용할 수 있도록 안내하였으며, 학생들도 데이터 제공 기관이나 단체 등 출처가 명확한 자료만을 사용하도록 요청하였기 때문이다. 따라서 세부 절차들을 반영하면서 학생들의 자료에서 나타나는 특징적인 요소를 표현할 수 있도록 Table 1에서 제작한 각 탐구 절차를 수정하여 오류 분석을 위한 분석틀로 재구성할 필요가 있었다.

이를 위해 각 과정별 탐구 절차를 참고하여 학생들의 오류 유형에 대해 귀납적으로 유목화하며 분석틀로 재구성하였다. 이 과정에서 분석틀의 타당도 확보를 위하여 과학 교육 전문가 1인, 데이터 기반 교육 연구 경력이 있는 지구과학교육 박사 1인, 지구과학교육 박사과정생 1인과의 동료 검토를 진행하였다. 동료 검토 과정에서 데이터 표현 과정에 해당하는 내용 중 데이터 표현 과정에서 세부 절차 간에 중복되는 내용이 있다는 것을 확인하고 이를 세부적으로 구분하였다. 또한, 기상요소와 관련한 잘못된 개념에 기반한 오류가 있다는 것을 확인하며 데이터 해석 과정에 해당하는 절차를 재구성하고 각 절차에 해당하는 오류 유형을 범주화하였다. 최종적으로 도출한 분석틀은 Table 2와 같다.

이후 제작한 분석틀에 맞추어 연구자가 학생들의 자료를 재코딩한 후 3명의 동료 연구자가 모두 의견을 합치한 사례에 해당하는 자료를 정리하였다. 분석틀을 활용한 분석 사례는 Fig 1과 같다.

Fig. 1에서 학생은 운형과 강수량의 비교를 위해 그래프의 각 축에 운형, 강수량을 표현한다. 다만, 운형은 범주형(categorical) 데이터에 해당하는 것으로 두 물리량 사이의 관계 분석을 위해서는 각각의 운형에 따른 평균적인 강수량을 제시하는 것이 적절하다. 이러한 요소를 범주형 자료 처리 미흡에 해당하는 오류 유형과 물리량 표현에 부적절한 시각화 유형을 활용한 오류로 판단하였다. 또한, “낮은 강수량이 나타”한다고 표현했지만, 강수량이 매우 적은 상황(1~5mm)에서 정량적인 값을 제시하지 않은 부분은 데이터의 특징을 정성적으로만 비교한 오류로 분류하였다. 또한, “맑은 날씨에 영향을 주는 하층운”과 같은 표현의 경우 전반적인 기상요소가 운형에 영향을 준다고 해석하는 것이 적절하기에 인과관계와 상관관계를 혼동한 사례에 해당하는 오류로 분류하였다.

연구를 위한 분석 단위는 탐구 보고서에서 각기 다른 데이터를 활용하여 활동을 수행한 내용에 해당한다. 예컨대, 기온의 시계열 변화, 기온과 습도의 상관관계, 기온과 사회적 데이터의 변동에 대해 탐구 활동을 진행한 보고서의 경우 각 내용을 주제 단위로 구분하였다. 이후 74명의 학생이 제작한 총 222개의 자료를 분석틀에 맞추어 분석하였다.

III. 연구 결과

1. 데이터 분석 과정에서 오류의 빈도

학생들의 탐구 보고서에서 분석 단위(주제별 탐구

Table 3. Results of the analysis of error types by category in the data analysis process

데이터 분석 과정	범주	빈도(%)*	합계(%)
수집 및 처리	잠정적 가설 설정	13(3.05)	40(9.39)
	데이터 수집	12(2.82)	
	데이터 형식 이해	15(3.52)	
표현	시각화 형식	17(3.99)	177(41.55)
	경향성 표현	42(9.86)	
	그래프 요소	94(22.1)	
	통계적 표현	24(5.63)	
해석	기술적 해석	30(7.04)	209(49.06)
	경향성 및 패턴 해석	66(15.49)	
	유의미한 정보 도출	25(5.87)	
	정보에 대한 설명	21(4.93)	
	분석 결과 및 결론의 정당화	67(15.7)	
계			426(100)

자료)인 222개의 자료에 대한 전체 오류의 빈도를 분석한 결과는 Table 3과 같다. 데이터 해석 과정에 대한 오류가 높은 빈도를 보이며 데이터 표현, 데이터 수집 및 처리 순으로 나타난다.

자료 해석 과정에서는 분석 결과 및 결론의 정당화(67회), 경향성 및 패턴 해석(66회)에서 높은 빈도를 보이는 것이 특징이다. 자료 표현 과정에서는 그래프의 요소와 관련한 항목이 전체 94회로 가장 높은 빈도를 보이며, 데이터에 나타나는 경향성 표현의 미비함(42회)도 높은 빈도를 보이는 것이 특징이다. 데이터 수집 및 처리 과정에서는 범주 간의 유사한 빈도가 나타났으며, 데이터 형식(15회), 데이터 수집을 위한 잠정적 가설 설정(13회), 데이터 수집(12회) 범주 순으로 나타났다. 다만, 학생들이 데이터를 분석하는 과정에서 수집 및 처리 과정을 명시적으로 기술하지 않은 경우가

있으며, 수집한 데이터의 종류와 특징에 따라 데이터 표현이나 해석을 위한 접근방식이 다소 상이하다. 따라서 이와 같은 결과가 동일한 과제에 대한 절대적인 빈도를 의미하진 않는다는 점에서 해석에 유의할 필요가 있다.

2. 기상 데이터 수집 및 처리 과정의 세부 오류 유형

기상 데이터 수집 및 처리 과정에서 나타나는 세부 오류 유형과 각 빈도는 Table 4와 같다.

첫째, 잠정적 가설 설정 범주에서는 잘못된 개념을 활용한 가설 설정과 관련한 오류가 가장 높은 빈도(6회)를 나타내는 것이 특징이다. 대표적인 사례로는 기상요소 변동이 계절성 기단의 영향이나 국소적인 기압

Table 4. Results of the analysis of error types in the ‘Data collection and preprocessing’ process

범주	오류 유형	빈도(%)*	합계(%)
잠정적 가설 설정	잘못된 개념을 활용한 가설 설정	6(15.0)	13(32.5)
	편향적 가설 설정	4(10.0)	
	가설의 타당성 미흡	3(7.5)	
데이터 수집	분석 범위에 적합하지 않는 데이터 수집	8(20.0)	12(30.0)
	(기간, 지역 등) 조건이 다른 데이터 수집	4(10.0)	
데이터 형식 이해	범주형 데이터 처리 미흡	5(12.5)	15(37.5)
	결측 처리 미흡	4(10.0)	
	분석 단위 변환 미흡	6(15.0)	
계			40(100)

배치에 따라 달라질 수 있다는 점을 고려하지 않은 채 자료를 수집한 경우가 있다. Fig. 2의 경우에도 우리나라 특정 지역에서 측정된 데이터를 수집하기 이전에 기온과 기압 사이의 관계가 반비례 형태일 것이라는 가설을 설정한다. 하지만 이는 일반적으로 다른 변수가 통제된 이상적인 상황에서의 관계이다. 실제 학생의 분석 결과에서도 수집한 기간 동안 기온과 기압이 점차 증가하는 경향이 나타났다.

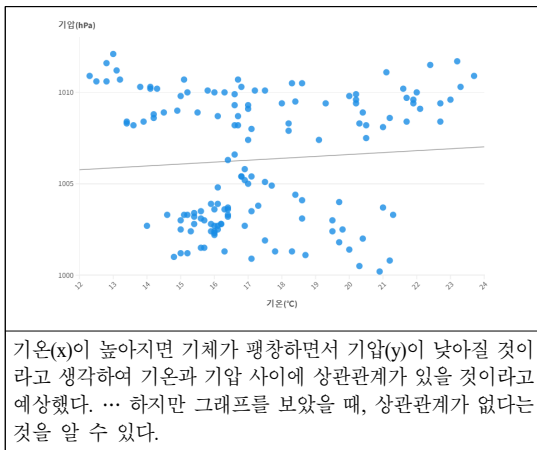


Fig. 2. An example of error type in the ‘Tentative hypothesis setting’ category

이외에도 특정한 조건에서만 성립하는 국소적이거나 다소 편향된 가설을 설정하는 사례(4회), 일상적인 경험이나 기분, 느낌 등에 기반하여 가설을 설정한 경우에 해당하는 타당성 미흡 측면(3회) 오류가 있었다.

둘째, 데이터 수집 범주에서는 분석 범위에 적합하지 않은 데이터를 수집하는 유형(8회)가 대표적이다. 예컨대, 지구온난화로 인해 기온이 증가할 것이라는

가설을 설정한 뒤, 단기간에 해당하는 기온 데이터를 수집하는 경우가 있었다. 이처럼 두 물리량 간의 상관관계가 나타날 것이라고 예상한 범위에 해당하는 데이터를 수집하지 않은 것에 해당한다. 또한, 조건이 다른 데이터를 수집하는 유형(4회)은 두 가지 이상 데이터 간의 관계를 분석하는 과정에서 데이터를 측정된 기간이나 지역을 고려하지 않는 오류에 해당한다. 예컨대, 기온 변화와 빙과류 판매량의 상관관계 분석에서 서로 다른 기간에 해당하는 기온의 일변화와 빙과류 판매량을 수집하여 비교한 사례가 있다. 마찬가지로 기온과 대중교통 이용량에 대한 분석에서 서로 다른 지역에서 측정된 물리량을 수집하는 사례가 있었다.

학생들은 잠정적으로 수립했던 가설에 기반하여 데이터를 수집한다. 하지만 데이터를 수집하는 과정에서 가설을 지지하는 데 가장 적합한 데이터만을 선별하는 사례가 있다는 것을 확인할 수 있었다. 이는 일종의 인지적 편향에 해당하는 것이다. 한편, 타당한 방법으로 수집한 데이터의 특징을 파악하여 더욱 적합한 새로운 가설을 설정하거나 가설을 정교화할 수 있다(김형욱, 2024). 이는 학생들의 탐구 과정 중에서도 데이터 수집이나 측정과 관련한 측면의 반성적 사고를 촉진할 필요성이 있다는 것을 시사한다(Chang et al., 2024).

셋째, 데이터 형식 이해 범주에서는 분석 단위 변환 미흡(6회), 범주형 데이터 처리 미흡(5회), 결측 처리 미흡(4회)이 유사한 빈도를 보이는 것이 특징이다. 대표적으로 단위 변환, 결측 처리 미흡의 경우에는 Fig. 3과 같은 사례가 있다. Fig 3의 좌측에 해당하는 자료에서는 날짜에 따른 태양광 발전량 변동을 살펴보기 위한 탐구를 계획하지만, 일별 시간에 따른 데이터를

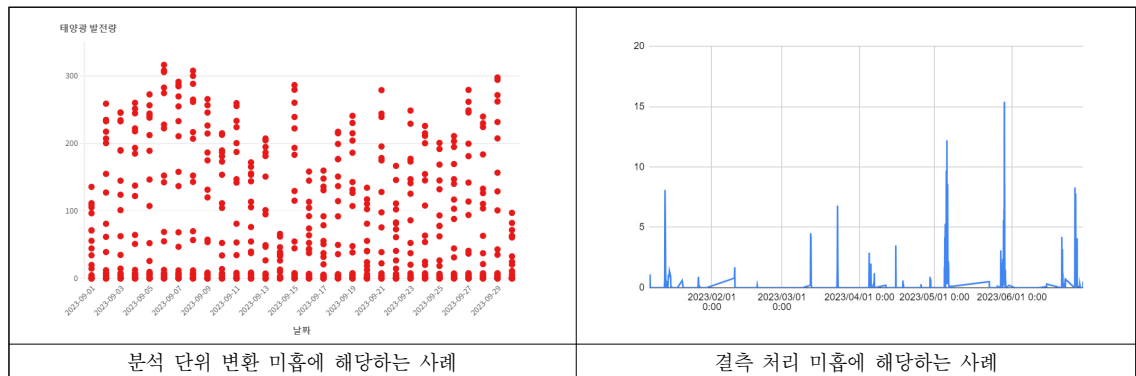


Fig. 3. An example of error type in the ‘Comprehension about data format’ category

수집한다. 따라서 이 학생이 표현한 자료에는 동일한 날짜의 여러 시점에 해당하는 값이 나타난다. Fig 3의 우측 자료에 해당하는 것은 일별 강수량 변동을 나타낸 그래프로, 중간 관측값이 없는 경우에도 자료가 표현되어 있어 이를 선형적인 그래프로 표현한 경우 결측값과 관측값 사이의 변화가 그래프로 표현된다.

이처럼 기상요소 중에서는 결측 구간이 나타나는 경우가 있다. 특히 상대적으로 긴 기간의 평균을 활용하는 월평균, 연평균 자료의 경우에는 결측이 거의 나타나지 않지만, 시간(hour)을 기준으로 하는 자료에서는 일부 구간에서 결측이 나타난다. 일부 학생들의 경우에는 결측이 나타나는 경우 분석하고자 했던 기상요소를 다른 요소로 대체하거나, 결측 구간이 나타나지 않는 시점으로 분석 범위를 변경하는 사례가 있었다. 하지만 결측에 대한 처리 절차를 거치지 않은 채로 결측치를 고려하지 않거나, 결측치에 대한 값을 단순한 추세선에 의해 판단하는 경우가 있다(Chang et al., 2024). 이러한 측면에서 학생들이 결측치를 처리하는 방법에 대한 이해가 다소 미흡하다고 해석할 수 있다.

이 범주는 기상요소와 관련한 추가적인 데이터를 수집하는 과정에서 양화하기 어려운 범주형 자료를 불충분하게 수집한 사례, 추가적인 데이터를 분석에 적합하지 않은 형태로 처리한 경우 등을 포함한다. 이러한 사례는 학생들이 기상요소 이외의 데이터를 수집하는 과정에서 언론 보도 자료, 이미지 검색 자료 등을 활용한 경우에 나타난 것이다. 이러한 결과는 학생들이 더욱 신뢰할 수 있으면서도 정교한 상관관계 분석을 위해 필요한 다양한 학습 자원을 안내할 필요가 있음을 시사한다(김명신 등, 2023).

3. 기상 데이터 표현 과정의 세부 오류 유형

기상 데이터 표현 과정에서 나타나는 세부 오류 유형과 각 빈도는 Table 5와 같다.

첫째, 시각화 형식 범주에서는 분석이 어려운 수준의 과도한 데이터를 제시하는 경우(5회), 또는 물리량 표현에 부적절한 시각화 유형(형식)을 활용하는 경우(12회) 등이 있다. 학생들은 일상적이고 친숙한 그래프를 선택하는 경향이 있는데(Chang et al., 2024), 쉽게 접할 수 있는 선형적 그래프나 막대 그래프 등을 선택하는 사례가 많다. 이때 경향성을 파악하기 어려운 수준으로 시각화 자료를 제작하는 사례가 있었다.

둘째, 경향성 표현 범주에서는 잘못된 추세선을 활용하거나(26회), 추세를 파악하기 어려운 범위에 해당하는 자료를 활용하여 추세를 분석하는 유형이 있었다. 학생들이 활용한 스프레드시트나 온라인 저작 도구에서는 추세선을 다항함수나 지수, 로그 함수 등을 활용하여 표현할 수 있다. 하지만 추세선을 표현하는 과정에서 분석의 단위가 되는 기준을 설정하지 않은 경우가 있다. Fig. 4와 같이 짧은 기간에 나타나는 풍속 변동은 국소적인 기압 배치의 영향을 받게 되며 뚜렷한 패턴을 보이지 않는 것이 특징이다. 이 데이터에서는 12월 17~18일 사이에 풍속의 최댓값이 나타나고 있지만, 이 구간에서 추세선은 아래로 블록한 형태로 그려지게 된다. 이러한 표현은 뒤이어 제시하는 자료 해석에도 영향을 주는 중요한 요인이 된다.

이처럼 추세를 크게 벗어나는 결측치나 이상치(outlier)와 같이 타당도가 낮은 자료가 있음에도 불구하고 온라인 저작 도구에서 제공하는 추세선을 단순히

Table 5. Results of the analysis of error types in the ‘Data representation’ process

범주	오류 유형	빈도(%)	합계(%)
시각화 형식	과도한 데이터 포함	5(2.82)	17(9.60)
	물리량 표현에 부적절한 그래프 활용	12(6.78)	
경향성 표현	경향 및 추세 표현 누락	16(9.04)	42(23.73)
	잘못된 추세선의 활용	26(14.7)	
그래프 요소	그래프의 축 간격	12(6.78)	94(53.11)
	단위 표현	56(31.6)	
	변수 표현	20(11.3)	
	변수의 특성 고려 미흡	6(3.39)	
통계적 표현	기술적 표현 누락	24(13.6)	24(13.56)
계			177(100)

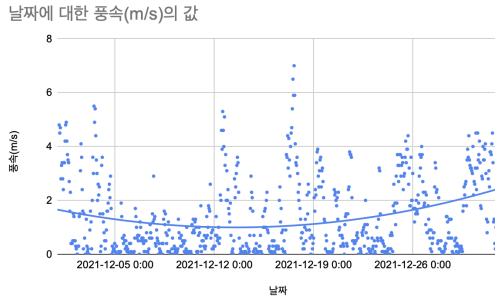


Fig. 4. An example of error type in the 'Trend expression' category

적용한 경우가 있었다. 또한, 추세가 나타나지 않거나, 추세선의 상관계수가 타당하지 않은 수준의 자료에서도 특정한 저작 도구의 추세선을 표현한 경우가 있다. 김유정 등(2009)은 추세선이 일종의 경향성을 보여주는 방법 중 하나이지만, 학생들이 추세선을 표현하는 과정에서 다양한 유형의 오류가 나타나는 사례가 있다는 것을 지적한 바 있다. 한편, 본 연구에서 나타난 오류는 저작 도구가 추세선의 타당도를 확인할 수 있는 기능을 제공하지 않으며, 학생들이 상관 함수와 상관 계수의 의미 이해에 어려움이 있었기 때문에 발생한 오류라고 해석할 수 있다.

셋째, 그래프 요소 범주에서 나타나는 오류는 주로 데이터의 단위(56회), 범례와 변수(20회) 등을 구체적으로 표현하지 않은 유형이 있다. 김유정 등(2009)의 연구에서는 학생들이 그래프의 변수나 단위 등을 올바르게 표현하지 못하는 사례가 있음을 지적한 바 있는데, 본 연구에서도 이러한 측면의 오류 빈도가 높은 편이 특징이다. 학생들이 온라인 도구를 활용하며 범례가 설정되어 있는 데이터를 활용하였기에 간단한 조작으로도 축의 단위와 눈금을 조정할 수 있다. 하지만 여전히 많은 학생들이 생산한 자료에서는 이와 관련한 요소가 명확하게 나타나지 않는 사례가 있었다.

그중 변수의 특성 고려 미흡(6회)에 해당하는 사례로는 Fig. 5가 있다. 이는 학생들이 활용하는 도구에 따라서 스프레드시트에 입력한 수치 데이터가 텍스트 데이터의 형식으로 처리된 경우에 나타난 사례이다. 이러한 오류를 방지하기 위해서는 데이터를 수집하고 처리하는 과정에서 데이터가 입력된 셀의 특성을 지정해 줄 필요가 있다. 하지만 학생들이 저작 도구가 생산한 그래프 자료를 검토하지 않고 그대로 제시하는 과정에서 이러한 오류가 발생하였다.

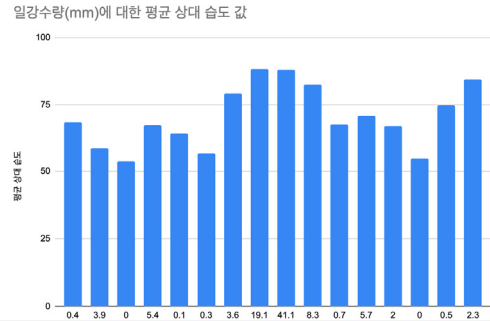


Fig. 5. An example of error type in the 'Elements of graph' category

그래프의 축 간격(12회)과 관련한 유형은 데이터에서 나타나는 경향을 분석하기 어려울 정도로 축을 과도하게 축소하거나 확장한 사례에 대한 것이다. 특히, 물리량의 단위가 다르고 값에서 큰 편차가 나타나는 두 가지 이상의 데이터를 표현하는 과정에서 양적 변화가 크지 않은 기상요소의 경향성이 드러나지 않게 되는 사례가 있다. 이는 이러한 오류를 나타낸 학생들이 저작 도구 활용 측면에서 서로 다른 축을 설정하는 방법을 숙지하지 못했다는 것을 의미한다.

넷째, 통계적 표현 범주는 대푯값을 비교하는 분석 사례에서 데이터의 평균, 최댓값, 최솟값 등을 제시하지 않은 경우에 해당한다(24회). 데이터를 해석하는 과정에서 데이터의 기술적 통계와 관련한 요인을 설명했지만, 이와 관련하여 별도의 값을 제시하지 않은 경우가 있었다.

이처럼 데이터 표현 과정에서는 데이터 해석에 영향을 주는 도구적 측면의 요인을 살펴볼 수 있다. 또한, 학생 중에서는 데이터를 수집하는 과정의 어려움, 데이터 표현을 위한 도구 사용의 어려움을 탐구 노트에 작성한 사례가 있었다. 이처럼 도구의 특성이나 도구 사용의 어려움으로 인해 데이터 표현에 해당하는 오류가 나타날 수 있으며 이는 데이터 해석에도 영향을 줄 수 있다. 디지털 도구가 수업을 구성하는 인공지능자 수업-배치(assembly)의 구성 물질이라는 점에서 직접적인 도구 활용 과정과 도구 활용에 영향을 주는 요인을 탐색하는 것이 중요하다(서민주 등, 2023; Hillmayr *et al.*, 2020). 또한, 학생들이 표현한 자료는 수집하고 처리한 데이터의 특성에 영향을 받는다. 그리고, 학생들은 생산한 시각화 자료에 기반해서 정보를 추론하게 된다. 이 과정에서 자신의 가설을 수정하거나, 해석을 위한 새로운 가설을

마련하게 된다(김익균 등, 2016; Gilbert, 2005). 따라서 학생들이 명료한 형태의 시각화 자료를 표상하는 것은 매우 중요하다.

4. 기상 데이터 해석 과정의 세부 오류 유형

기상 데이터 해석 과정에서 나타나는 세부 오류 유형과 각 빈도는 Table 6과 같다.

첫째, 기술적 해석 범주에 해당하는 유형으로는 데이터에 나타나는 극값, 평균값 등에 대한 구체적인 표현 없이 정성적인 수준에서 비교한 사례가 있다(27회). 이 유형에는 추세선을 벗어나는 큰 오차에 대해 해석을 하지 않은 경우, 표준 편차가 상대적으로 큰 값들이 포함되어 평균의 타당도가 낮음에도 불구하고 단순히 평균값만을 사용하여 해석한 사례 등을 포함한다(Glazer, 2011). 이는 데이터 표현 과정에서 나타나는 기술 통계량 표현이 미흡한 유형과 밀접한 관계가 있는 것이다. 이처럼 정성적인 접근을 통해 데이터를 분석한 경우에 특이값들을 올바르게 파악하지 못하고 값을 다르게 해석하는 경우(3회)가 발생하기도 했다.

둘째, 경향성 및 패턴 해석 범주는 전체 66회의 높은 빈도를 보이는 것이 특징이다. 이러한 범주의 대표

적인 유형으로는 경향 및 관계를 정성적으로만 해석하는 유형(33회)이 있다. 두 물리량 사이의 상관관계를 분석하는 과정에서 물리량을 변수로 활용하여 그래프를 그리지 않고, 시간에 따른 추세선을 정성적으로 비교하는 경우가 대표적이다. Fig. 6과 같은 사례에서는 평균 상대습도와 평균 기온의 물리량에 대한 서로 다른 추세선을 그리고 추세선 사이의 관계를 해석한다. 두 물리량 사이의 관계를 구체적으로 파악하기 위해서는 서로 다른 물리량을 두 변인으로 두어 관계를 분석해야 하지만, 시계열적 경향만을 비교했기에 해석에서 미흡한 부분이 나타났다.

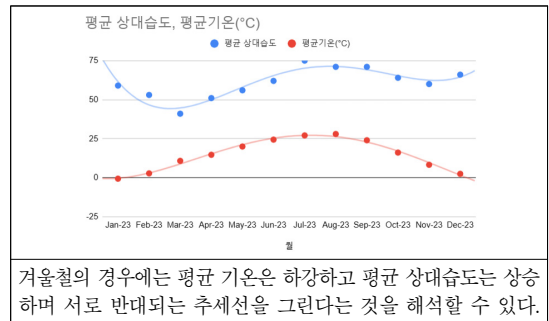


Fig. 6. An example of error types in the ‘trend and pattern interpretation’ category

Table 6. Results of the analysis of error types in the ‘Data interpretation’ process

범주	오류 유형	빈도(%)	합계(%)
기술적 해석	데이터의 특징을 정성적으로만 비교한 경우	27(12.92)	30(14.35)
	자료와 다르게 해석한 경우	3(1.44)	
경향성 및 패턴 해석	상관관계와 비례관계 표현 혼동	20(9.57)	66(31.58)
	함수관계에 대한 해석 미흡	5(2.39)	
	진동 형태에 대한 해석 미흡	4(1.91)	
	자료와 다르게 해석한 경우	4(1.91)	
	경향 및 관계를 정성적으로만 해석한 경우	33(15.79)	
유의미한 정보 도출	인과관계와 상관관계의 혼동	10(4.78)	25(11.96)
	인과적 관계에 있는 변수 파악 미흡	7(3.35)	
	현재 추세, 특정한 값에 기반한 과도한 일반화	8(3.83)	
정보에 대한 설명	변수에 영향을 주는 요인에 대한 고려 미흡	17(8.13)	21(10.05)
	이상적 상황(실험실 상황)과의 비교	4(1.91)	
분석 결과 및 결론의 정당화	잘못된 개념에 기반한 논증	24(11.48)	67(32.06)
	잘못된 자료에 기반한 논증	3(1.44)	
	데이터를 바탕으로 한 단정적 진술	20(9.57)	
	연역적 정당화 및 논증 오류	6(2.87)	
	자료를 제시하지 않고 주장	14(6.70)	
계			209(100)

또한, 높은 빈도가 나타나는 상관관계와 비례관계 해석에 대한 혼동(20회)의 유형이 있다. 이는 x축의 변수가 증가할 때 y축의 변수가 감소하는 경향을 보이는 경우를 반비례 관계로 표현한 사례가 다수이다. 이외에도 함수관계, 진동 형태에 대한 해석 미흡에 대한 유형은 자료의 분포나 추세선의 형태를 수학적 용어로 표현하는 과정에서 나타난 오류가 있다. 대표적으로 주기함수와 이차곡선 등에 대한 혼동이 있다. 이처럼 학생들은 데이터를 수집한 기간에 한정하지 않고 그 경향성을 친숙한 함수적 표현을 통해 표현하거나, 측정 구간 이외에서도 이와 같은 추세가 나타날 것이라고 해석하는 경우가 있었다.

셋째, 유의미한 정보 도출 범주는 학생들이 표현한 데이터에서 나타나는 패턴에 대한 인과적 해석에 대한 범주이다. 대표적으로 나타나는 오류 유형은 인과관계와 상관관계에 대한 잘못된 해석(10회)이 있다. 이는 두 변수가 인과성을 갖지 않는 것임에도 불구하고 상관관계를 인과관계로 혼동하여 해석하는 경우가 대부분이다(Rohrer, 2018). 또한, 변수에 영향을 주는 다양한 요인이 있음에도 불구하고 특정한 기상요소의 변화를 한가지 기상요소에 의한 변동에 의한 것으로 한정하는 오류 유형(7회)이 있다. 대표적으로 불쾌지수와 관련한 탐구에서 불쾌지수가 온도와 상대습도에 영향을 받는 물리량임에도 불구하고, 상대습도를 고려하지 않고 불쾌지수를 온도에 비례하는 물리량으로 해석하는 사례가 있었다. 그리고 미래나 과거의 값을 예상하는 과정에서 현재 수집한 구간의 결과를 과도하게 일반화하는 유형(8회)이 있다. 이 같은 유형은 짧은 기간의 국소적 변동을 수집한 사례, 계절 변동 패턴을 파악하지 못하고 특정 구간의 경향성이 지속될 것이라고 판단한 사례 등을 포함한다. 대표적으로 Fig. 7에서 학생은 강수에 의해 특정 기간의 기온이 낮아진 시점을 분석하면서 “기온이 점차 낮아질 것이라고 예측할 수 있다”는 일반화된 결과를 도출하였다. 하지만 학생이 관측을 수행한 구간의 일부 영역에서는 계절적 요인에 의해 기온의 최댓값이 점차 상승하는 구간을 포함하는 것이었다. 따라서 이러한 해석에는 다소 단정적인 판단이 포함된 것이다. 이는 학생들이 제작한 자료에 나타나는 평균적인 경향을 통해 일반화된 진술을 이끌어낸 사례가 있음을 발견한 김형욱(2024)의 연구 결과와도 유사하다. 한편, 이러한 오류는 학생들이 그래프에

나타나는 경향성을 추세선이나 평균값을 제시하지 않고 정성적으로 비교하여 해석하는 경우나, 자료 표현 과정에서 통계적 기술과 관련한 표현이 미흡한 경우의 사례와도 관련된 것이다.

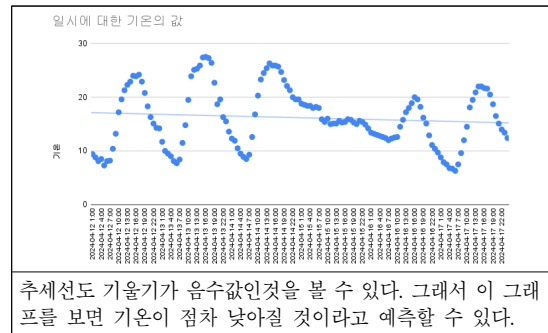


Fig. 7. An example of error types in the ‘deriving information’ category

넷째, 도출한 정보에 대한 설명 과정에서 나타나는 오류가 있다. 이는 데이터의 특징을 통해 두 물리량의 관계를 해석하기 위한 모델을 구성하고 활용하는 단계에 해당한다(김지나와 변영찬, 2007). 대표적인 유형으로는 변수에 영향을 주는 다양한 요인이 있지만 이에 대한 영향력을 고려하지 않은 유형(17회), 통제된 이상적인 상황에서 적용 가능한 범칙으로 자연적 현상을 설명하는 경우(4회)가 있었다. 그중에서 3회 이상의 빈도를 보였던 사례로는 기온과 포화수증기량에 대한 관계에 대한 잘못된 해석이 있다.

기온과 포화수증기량 사이에는 기온이 높아지면 포화수증기량이 증가하는 양의 상관관계가 나타나는데 상대습도는 포화수증기량이 증가할수록 낮아지는 음의 상관관계를 가지므로 기온과 상대습도 사이에서 음의 상관관계를 확인할 수 있다.

이는 기온과 습도의 관계를 해석하기 위해 기온에 따른 포화수증기압 곡선을 활용하여 현상을 설명하는 것이다. 하지만 포화수증기량이 증가하는 상황에서 상대습도가 점차 낮아지는 것은 전체 공기 중에 포함된 수증기량이 통제된 상황에 적합한 설명이다. 하지만 실제 우리나라 여름철 해양성 기단의 영향으로 기온과 습도가 모두 상승하게 되는 것을 고려하지 않은 채로 해석한 사례에 해당한다. 이러한 사례의 경우 학생은

데이터 표현 결과가 선개념에 대해 불일치하는 사례를 경험하게 된다. 어느 학생의 경우 이러한 결과가 자신이 알고 있는 개념과 다른 상황임을 설명하면서 과학적 이론이 모든 상황에 적용되는 것은 아니라는 진술을 이끌어 낸 경우가 있었다. 이 같은 측면에서 선개념은 가설을 설정하는 과정이나 데이터를 표현하고 해석하는 과정에 총체적으로 영향을 끼치는 요인이 된다고 할 수 있다(김익균 등, 2018, 김지나와 변영찬, 2007; Glazer, 2011).

마지막으로 분석 결과 및 결론의 정당화 범주에서는 잘못된 개념에 기반한 논증(24회), 데이터를 바탕으로 한 단정적 진술(20회) 등을 포함한다. 대표 유형으로는 과학적으로 부정확한 개념어를 활용하거나 활용, 오개념을 활용한 해석의 사례가 있다. 어느 학생의 경우 강수량과 미세먼지 양(농도), 풍속과 미세먼지 양(농도)간의 관계에 대해 분석한 후 다음과 같이 탐구의 결론을 진술한다.

따라서 강수량이 줄어드는 1월쯤에 미세먼지 양은 높고, 강수량이 많아지는 7월쯤에는 미세먼지양이 낮게 나타나는 것이다. ... 청주의 여름철에는 저기압으로 인해 운량과 강수량이 많아지게 되기 때문에 미세먼지양은 적게 나타나는 것이다.

이러한 사례에서는 미세먼지의 농도가 우리나라의 계절적 풍계에 영향을 받는 요인이라는 것을 고려하지 못한 것을 확인할 수 있다. 또한, 학생의 진술에서 "...여름철에는 저기압..."과 같은 오개념이 포함되어 있는 것을 확인할 수 있다. 이처럼 학생들은 기상요소에 영향을 미치는 요인에 대해 잘못된 개념을 활용하거나, 분석 범위에 해당하는 데이터를 바탕으로 탐구의 결과를 과도하게 일반화하고 이를 통해 결론을 도출하는 경우가 있었다. 이러한 측면에서는 데이터의 주제와 관련한 개념적 학습에 대한 중요성에 대해 고찰할 필요가 있다. 기상요소의 구체적인 변동 특징이나 패턴을 올바르게 해석하기 위해서는 과학적 개념에 대한 정교한 이해가 필수적이기 때문이다.

IV. 결론 및 제언

본 연구에서는 기상 데이터 분석 과정에서 나타나는 오류의 유형을 분석하고 오류가 발생하는 맥락에 대해 탐색하였다. 이를 통해 빅데이터 활용 지구과학 수업에 대한 시사점을 탐색하고자 하였다. 연구 결과와 논의를 바탕으로 하는 연구의 결론은 다음과 같다.

첫째, 학생들은 기상 데이터를 해석하는 과정에서 다양한 유형의 오류를 보였다. 특히, 계절적 패턴과 주기적 변동에 대한 불충분한 이해를 바탕으로 해석하여 오류가 나타나는 사례가 대표적이었다. 그리고 이는 불충분한 논증의 사례로 이어지기도 한다. 이러한 측면에서 데이터 표현에 대한 단순한 읽기가 아닌 데이터와 관련한 구체적인 맥락(context)을 정교하게 학습하는 것이 중요하다. 따라서 과학 교과의 학습에서는 데이터의 내용에 담긴 의미를 구체적으로 이해하고, 데이터가 기존의 지식과 맺는 관계를 명확하게 파악하도록 하는 것을 강조해야 한다. 또한, 학생들의 데이터 해석이 학생들이 가지고 있는 선개념이나 인지적 모델에 영향을 받을 수 있다는 것을 확인하였다. 이러한 측면에서 실제 생활에서 나타나는 데이터와 모델 사이의 불일치를 통해 모델의 한계점을 밝히고 이를 정교하게 교정하는 모델 기반 수업을 진행할 수 있다. 이를 통해 내용 요소에 대한 맥락적 이해를 촉진하는 학습 기회를 제공할 수 있을 것이다.

둘째, 기상 데이터를 표현하는 과정에서의 오류로는 추세선을 잘못 활용하거나 기술적 통계에 대한 표현의 미비함 등의 요소가 있었다. 이는 데이터 해석을 단순한 정성적 비교에 국한하게 되는 요인이 되었다. 따라서 정교한 데이터 분석을 위해서는 통계적, 수학적 기초 학습이 필수적이다. 다만, 이러한 내용 요소가 중등학교의 수학 교과 교육과정에 포함되어 있다는 점에서 학습에 대한 전이를 높이기 위한 방안이 필요하다. 이를 위해서는 실생활과 관련한 다양한 차원의 경험이 필요할 것이다. 또한, 학생들이 수집한 기상요소는 서로 독립적인 물리량이기보다는 다양한 변인들과 맞물려 상호작용하는 것이다. 따라서 다양한 형태의 시각화 방법을 사용하는 수업을 진행하거나 다변수 함수 회귀 분석, 데이터 군집화와 같은 더욱 정교한 형태의 교수방안을 마련할 수 있을 것이다.

셋째, 오류 발생 맥락으로는 데이터의 특성과 분석

범위, 데이터를 처리하고 표현하는 도구 등이 영향을 준다. 학생들이 외부의 빅데이터를 활용할 수 있으며, 빅데이터를 분석을 위한 다양한 도구를 활용할 수 있다는 것은 더욱 포괄적인 형태의 연구가 필요하다는 것을 시사한다. 이러한 관점에서는 도구의 예측하지 못했던 새로운 생성, 도구의 활용 과정에서 (무)의식적으로 변화한 인식 등을 연구의 대상으로 포함할 수 있다. 도구와의 얽힘이 만들어 내는 새로운 유형의 사고와 추론 규칙과 같이 분산적이고 확장된 관점을 바탕으로도 학습을 바라볼 수 있을 것이다. 나아가 데이터 자체가 가지고 있는 한계와 불명료함에 대해서는 과학적 지식의 생산 과정에 대한 인식론적 측면의 수업이 가능하다는 것을 시사한다.

넷째, 데이터 분석 과정에서 오류의 유형은 각각의 유형과 맥락적 발생이 독립적이기보다는 상호 관계 속에서 발생한다. 예컨대, 데이터를 잘못 처리하는 경우 데이터 표현과 해석에 영향을 준다. 이러한 측면에서 데이터의 수집과 처리, 표현, 해석의 요인들이 매우 유기적으로 관계 맺고 있다는 것을 살펴볼 수 있다. 따라서 데이터 분석이라는 문제해결과정을 더욱 총체적인 관점에서 조망할 필요가 있다. 다만, 본 연구에서는 학생들의 오류를 탐색하기 위하여 각각의 질차적 범주에 해당하는 유형으로 구분하여 제시하였다. 따라서 문제해결과정에 영향을 주는 다양한 맥락 간의 구조적 관계를 밝혀내는 것에는 한계점이 있다. 따라서 문제해결과정에서 나타나는 인지적 전략을 파악할 수 있는 추적 자료 등을 활용한다면 오류가 발생하는 더욱 구체적인 맥락과 맥락 사이의 관계성을 살펴보는 데 도움을 줄 수 있을 것이다. 나아가, 학생들의 구체적인 학습 수준이나 데이터 표상 능력이 문제해결과정에 영향을 줄 수 있다는 점에서 오류의 발생과 관계를 갖는 학습자 특성을 탐색하는 연구를 진행할 수 있을 것이다.

데이터는 결코 정합적인 채로 완결된 것이 아니다. 현상은 객체적인 진실로 이미 세계 내에 놓여있기보다는 새로운 데이터들과 함께 공동-생산된다. 지금까지는 복잡한 자연계의 요소들을 환원하며 추상적 지식을 생성하는 것이 과학자가 전유하던 것이었다면, 이제는 학생들이 새로운 생산자가 되어야 한다. 따라서 미래 사회 교육 목표 달성을 위해서는 데이터 기반 교수학습활동에서 나타나는 구체적인 학습의 양태를 살펴보는 작업을 선행해야 한다. 이를 통해 학생들이 데이터

기반의 학습과 관련한 다양한 측면의 역량을 강화할 수 있도록 하는 교수학습 방법과 교수학습 맥락을 지원할 수 있는 요인들을 발견할 수 있을 것이다.

국문요약

본 연구에서는 고등학생들의 기상 데이터 분석 과정에서 나타나는 오류와 오류의 발생적 맥락에 대해 탐색하였다. 연구를 위해 A 고등학교 74명의 학생들이 제작한 222개의 기상 데이터 탐구 활동 내용을 수집하고, 데이터 수집 및 해석, 데이터 표현, 데이터 해석 과정에서 나타난 세부 오류 유형을 탐색하였다. 연구의 결과로 데이터 해석 과정에서는 학생들이 기상요소의 계절적 변동과 주기적 패턴에 대한 이해가 다소 미흡하다는 것을 확인하였다. 또한, 데이터 표현 과정에서는 그래프의 단위 표현, 범례 설정, 추세선 활용 등에서 다양한 유형의 오류가 있음을 확인하였다. 오류가 발생하는 원인으로는 저작 도구의 특징, 오개념과 인지적 편향 등이 있었다. 연구 결과를 바탕으로 미래 사회 과학 교육의 중요한 화두인 빅데이터 활용 교육에 대한 교육적 시사점을 도출하였다. 그리고 이와 관련한 후속 연구를 제안하였다.

주제어: 기상, 빅데이터, 데이터, 오류, 맥락

References

- 교육부(2022). 2022 개정교육과정 총론 과학과 교육과정 (고시번호 2022-33, [별책 9]). 세종: MOE.
- 김명신, 조영환, 금선영(2023). 탐구학습에서 가설 생성과 평가를 위한 데이터 시각화의 효과와 난점. *교육정보미디어연구*, 29(1), 175-199.
- 김봉철, 김재준, 문우중, 서영호, 김정아, 오정철, 김용민, 김종훈(2021). 데이터 과학 교육이 초등학생의 컴퓨팅 사고력에 미치는 효과: 마이크로비트의 센서 기능을 중심으로. *정보교육학회논문지*, 25(2), 337-346.
- 김소현, 유선아(2023). 데이터 기반 모델링 활동을 통한 과학적 탐구 질문 생성과 과학 개념의 변화. *학습자 중심교과교육연구*, 23(3), 747-764.

- 김용민, 김종훈(2017). 스프레드시트를 활용한 데이터 과학 교육 프로그램이 초등학생의 컴퓨팅 사고력 향상에 미치는 효과. *정보교육학회논문지*, 21(2), 219-230.
- 김유정, 최길순, 노태희(2009). 고등학생들의 과학 그래프 작성 및 해석 과정에서 나타난 오류. *한국과학교육학회지*, 29(8), 978-989.
- 김익균, 이인선, 박종원(2016). 과학영재학교 학생들의 변인간의 관계를 찾는 자료변환과정 분석. *새물리*, 66(7), 853-863.
- 김익균, 이인선, 박종원(2018). Effects of Students' Background Knowledge and Methodological Belief on the Process of Finding the Relationship between Measured Data. *새물리*, 68(4), 387-396.
- 김재경, 손의성(2021). 비전공자 대상의 데이터과학 교과목 개발과 컴퓨팅 사고력 효과 분석. *컴퓨터교육학회논문지*, 24(3), 23-31.
- 김지나, 변영찬(2007). 물리실험에서 선개념과 불일치한 실험 데이터에 대한 학생들의 해석 유형. *한국과학교육학회지*, 27(5), 404-411.
- 김형욱(2024). 초등학교 과학영재 학생을 대상으로 한 데이터 시각화 학습 프로그램 개발 및 효과: Tableau 프로그램 활용을 중심으로. *초등과학교육*, 43(1), 18-34.
- 마민영(2018). 그래프 표현과 해석에서 드러나는 두 중학생의 일정한 변화율의 이해에 대한 사례 연구. *중등교육연구*, 66(2), 439-466.
- 서민주, 서유민, 정혜윤, 이경화(2023). 표집 시뮬레이션을 활용한 비형식적 통계적 추리의 교수-학습: 문화역사적 활동이론의 관점에 따른 분석. *한국학교수학회논문집*, 26(1), 21-47.
- 송유경, 송석리, 김예지, 임철일(2021). 데이터 리터러시 향상을 위한 데이터 기반 토론 수업 모형 및 교수전략 개발 연구. *교육공학연구*, 37(4), 943-982.
- 양수진, 장명덕(2012). 초등학생들의 과학 선 그래프 작성 및 해석 과정 분석. *초등과학교육*, 31(3), 321-333.
- 오영열, 이윤경(2022). 데이터 기반 통계교육을 위한 초등학생의 데이터 관련 오류적 인식 사례 분석. *한국초등교육*, 33(4), 91-108.
- 정인진, 박경애, 이지은(2023). 중등학교 교과서 기후변화 관련 온실 기체 데이터 시각화 수업 활동 자료 개발. *현장과학교육*, 17(4), 416-430.
- 정혜윤, 유선아(2023). 복잡한 데이터를 활용한 모델 기반 탐구에서 나타난 예비교사의 과학적 상상력 발현 및 인식. *학습자중심교과교육연구*, 23(4), 171-187.
- 지영명(2020). 비형식적 통계적 추리 지도를 위한 예비초등교사의 통계지식 연구. *서울대학교 대학원 박사학위논문*.
- 최한나, 이종희(2023). 중학교 수학과 사회, 과학 교과서의 그래프 활동 및 이해 수준 분석. *교과교육학연구*, 27(1), 108-125.
- Åberg Bengtsson, L., & Ottosson, T. (2006). What lies behind graphicacy? Relating students' results on a test of graphically represented quantitative information to formal academic achievement. *Journal of Research in Science Teaching*, 43(1), 43-62.
- Azzam, T., Evergreen, S., Germuth, A. A., & Kistler, S. J. (2013). Data visualization and evaluation. In T. Azzam & S. Evergreen (Eds.), *Data visualization, part 1. New Directions for Evaluation*, 139, 7-32.
- Binali, T., Chang, C. H., Chang, Y. J., & Chang, H. Y. (2022). High school and college students' graph-interpretation competence in scientific and daily contexts of data visualization. *Science & Education*, 1-23.
- Börner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, 116(6), 1857-1864.
- Chai, C. P. (2020). The importance of data cleaning: Three visualization examples. *Chance*, 33(1), 4-9.
- Chang, H. Y., Chang, Y. J., & Tsai, M. J. (2024). Strategies and difficulties during students' construction of data visualizations. *International Journal of STEM Education*, 11(11), 1-22.
- Farmer, S. (2012). Real graphs from real data: Experiencing the concepts of measurement and uncertainty. *School Science Review*, 346, 81-84.
- Gilbert, J. K. (2005). Visualization: A metacognitive skill in science and science education. In Gilbert, J. K.(Eds.), *Visualization in science education*. (pp. 9-27). Dordrecht: Springer Netherlands.
- Glazer, N. (2011). Challenges with graph interpretation: A review of the literature. *Studies in Science Education*, 47(2), 183-210.

- Gobert, J. D., Moussavi, R., Li, H., Sao Pedro, M., & Dickler, R. (2018). Real-time scaffolding of students' online data interpretation during inquiry with Inq-ITS using educational data mining. *Cyber-physical Laboratories in Engineering and Science Education*, 191-217.
- Hillmayr, D., Ziemwald, L., Reinhold, F., Hofer, S. I., & Reiss, K. M. (2020). The potential of digital tools to enhance mathematics and science learning in secondary schools: A context-specific meta-analysis. *Computers & Education*, 153, 103897, 1-25.
- Maltese, A. V., Harsh, J. A., & Svetina, D. (2015). Data visualization literacy: Investigating data interpretation along the novice-expert continuum. *Journal of College Science Teaching*, 45(1), 84-90.
- Mensan, T., Osman, K., & Majid, N. (2020). Development and validation of unplugged activity of computational thinking in science module to integrate computational thinking in primary science. *Education Science Education International* 31(2), 142-149.
- Pols, C. F. J., Dekkers, P. J. J. M., & De Vries, M. J. (2021). What do they know? Investigating students' ability to analyse experimental data in secondary physics education. *International Journal of Science Education*, 43(2), 274-297.
- Prain, V., & Tytler, R. (2012). Learning through constructing representations in science: A framework of representational construction affordances. *International Journal of Science Education*, 34(17), 2751-2773.
- Rahmatina, D., & Zaid, N. M. (2019). Students' misconceptions in interpreting the mean of the data presented in a bar graph. *International Journal of Insights for Mathematics Teaching*, 2(1), 57-74.
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27-42.
- Volkwyn, T. S., Airey, J., Gregorcic, B., & Linder, C. (2020). Developing representational competence: Linking real-world motion to physics concepts through graphs. *Learning: Research and Practice*, 6(1), 88-107.
- Wong, V. (2017). Variation in graphing practices between mathematics and science: Implications for science teaching. *School Science Review*, 98(365), 109-115.
- Zagallo, P., Meddleton, S., & Bolger, M. S. (2016). Teaching real data interpretation with models (TRIM): Analysis of student dialogue in a large-enrollment cell and developmental biology course. *CBE-Life Sciences Education*, 15(2), 1-18.