J. Korean Soc. Math. Ed. Ser. D. (2024) 27(4), 431–447
https://doi.org/10.7468/jksmed.2024.27.4.431
ISSN 1226-6191
Online ISSN 2287-9943

RESEARCH ARTICLE

From missteps to insights: Reflecting on failed secondary analyses

Jeong-su Ahn[1], Min-kyeong Ji[2], Yeonseok Ka[3], Ki-jeong Bae[4], Seung-jin Cha[5], Jihyun Hwang[6]

[1] Teacher, Sejong Academy of Science and Arts
[2,3] Graduate Student, Department of Mathematics Education, Korea National of University of Education
[4] Teacher, Sang-Gyung Middle School
[5] Teacher, Chung-buk Cheong-won High School
[6] Professor, Department of Mathematics Education, Korea National of University of Education

Received: June 3, 2024 / Revised: July 13, 2024 / Accepted: August 19, 2024

Abstract

Secondary analysis of existing data provides unique opportunities for researchers to conduct large-scale studies with enhanced efficiency of resources and time, a concept increasingly utilized in educational research. This study aimed to provide crucial insights into the alignment between original and secondary theoretical frameworks, the creation and interpretation of variables, and the careful handling of data structures. This study explored the specific challenges and opportunities of using secondary analysis through two illustrative cases involving the Trends in International Mathematics and Science Study (TIMSS) and the Seoul Educational Longitudinal Study (SELS). The case with the TIMSS data highlighted the complexities of applying theoretical frameworks and interpreting newly generated variables when the original data lacked certain measurements needed for current research questions. A significant challenge identified in the SELS data was the non-uniformity of ID structures over years, which complicates the understanding of longitudinal trends and demands meticulous attention to detail. Our cases underscored the necessity for rigorous theoretical alignment, precise operational definitions, and nuanced statistical interpretations when conducting secondary analysis and international comparison. These are essential to unlock the full potential of large-scale public datasets in educational research, ensuring that findings are both scientifically robust and practically relevant.

Keywords: secondary analysis, longitudinal studies, theoretical alignment, TIMSS, SELS

• Corresponding Author: Min-Kyeong Ji, email: wasylove@hanmail.net

I. INTRODUCTION

In mathematics education, quantitative research has contributed significantly to decision-making and theory development related to teaching and learning by utilizing various data sources. Particularly, the advancements in artificial intelligence, machine learning, and various statistical techniques are providing more research opportunities that can contribute to the development of mathematics education. At the same time, increasing numbers of large-scale data open to public provide other changes for quantitative analysis.

Secondary analysis involves utilizing previously collected data, offering easier accessibility and higher validity compared to traditional methods where researchers collect the data themselves. It is the reuse of existing data to answer new research questions (Tripathy, 2013), facilitating research at a scale that individual researchers cannot achieve alone and optimizing the efficiency of resources and time (Castle, 2003). Researchers attempt to analyze already collected data in their field of interest to derive statistically significant results. This approach provides the full value of quantitative research. However, because it differs from traditional research methodologies, it requires thorough consideration of research methods, implementation, and interpretation.

Common practices in secondary analysis differ significantly from traditional approaches that attempt to understand phenomena based on pre-established theoretical models and first-hand collected data. Particularly, these exploratory methods, including latent profile analysis and exploratory factor analysis with secondary data, often involve naming unobserved latent variables. While this can facilitate interpretation and make results easier to understand, it is a practice that should be avoided. Respect for the latent and unobserved nature of variables is a fundamental principle in interpretation (Henson & Roberts, 2006). Furthermore, the frequent confusion between correlation and causation in exploratory research is a critical issue (Altman & Krzywinski, 2015). Specifically, in secondary analysis, where researchers are working with data not collected by themselves, interpretations of causality need to be approached with great caution. While correlation indicates a statistical association between two variables, it does not imply causation (Pearl, 2009). Many researchers misinterpreted observed correlations as causal relationships, leading to misuse of research findings. To counteract this, researchers can employ experimental or longitudinal research designs, or enhance interpretations through statistical methods. However, in quantitative research in mathematics education, the most crucial element for claiming causality is a sound theoretical discussion.

In the field of mathematics education, secondary analysis research using public education data such as PISA and TIMSS is actively conducted (Hwang & Shin, 2021). These public education data afford researchers the opportunity to analyze large amounts of information that would be difficult for private researchers to collect, and to understand achievement and contextual variables in different contexts. In addition to international-scale datasets such as PISA and TIMSS, Seoul and other local education offices in Korea are engaged in the construction of large-scale datasets on local education. Among these are longitudinal studies of education, such as the Seoul Education Longitudinal Study (SELS). This not only allows for the investigation of local characteristics and the efficacy of education policies pursued by local education boards, but also provides an opportunity to explore a variety of research questions in a region of interest.

A recent example of studies in the field of mathematics education using secondary analysis of educational public data is Jeong et al. (2023). The authors conducted an international comparative analysis of students' mathematics literacy based on PISA 2012 data using latent profile analysis. They aims to assess and compare student achievement, instructional practices, and curricula across different countries. Another example of study is Lindström et al. (2024) where structural equation modeling (SEM) was used to analyze the correlation between teacher competence and student achievement in Sweden.

Despite significant advantages of using these datasets for secondary analysis, researchers should consider several factors in analysis. First of all, the transparency of the source of secondary data, the collection methods, and the processing steps significantly affect the accuracy and reliability of the research results. If researchers conduct analyses without considering the original context of the data or without sufficient theoretical discussion, it can lead to errors in interpreting the results.

Furthermore, it is essential to consider the impact of cultural differences when conducting international comparisons through secondary analysis. Hatano and Inagaki (1998) argued that it is impossible to improve an educational system simply by importing technologies and beliefs developed in another culture. This highlights the difficulty of directly transferring educational practices from one context to another without considering cultural differences. The analysis and the interpretation should be based on acknowledgement on all educational systems in comparison.

Without this interpretation, international comparison research can be misused in a way to impose a global curriculum. Forcing a standardized global curriculum on participating countries can overlook local educational needs and cultural values, leading to ineffective educational reforms. Another misuse is the appropriation of the research agenda by dominant countries or organizations, which might skew the research to reflect their own priorities, sidelining the needs of other participants. Furthermore, the exploitation of research results can occur when findings are misinterpreted or misapplied, resulting in inappropriate and ineffective educational changes (Holliday & Holliday, 2003).

This study presents two cases illustrating the difficulties and dilemmas experienced by researchers during the analysis of TIMSS and SELS data. Through these cases, the study aims to provide points of caution and implications for conducting secondary analysis and international comparison through secondary analysis. A recent literature review showed that most studies focused on successful outcomes, and reflective analysis of failed studies was rare (Banks et al., 2012). However, learning from failures could be considered an important experience (Pflugler et al., 2018), and critical review of failed actions can broaden fundamental understanding (Shepherd et al., 2014). Therefore, sharing our failure cases in the data analysis process will have significant academic value.

II. CASE 1: SECONDARY ANALYSIS WITH TIMSS 2019 DATA

As the first case, the research team conducted a secondary analysis based on TIMSS 2019 data to address the following research questions:

- Research Question 1: What is the correlation between classroom environment factors and teachers' job satisfaction?
- Research Question 2: How does the correlation between classroom environment factors and teachers' job satisfaction vary by country, Korea and Finland?

Research Background of Case 1

Teachers and students interact with each other in the unique context of the classroom. In the literature on mathematics education, studies focusing on students focused on mathematical achievement and affective domains such as attitudes of mathematics learners' (e.g. Kim et al., 2005; Park & Sang, 2011). However, most Korean studies on teachers emphasized their professionalism, such as meaningful teaching and learning methods that can be utilized in the classroom, and relatively little attention was paid to their personal satisfaction. However, in the wake of the 2023 Korean teacher protests, there has been a growing interest in teachers' job satisfaction.

When looking at studies on teacher job satisfaction, most of them examined on factors outside of the classroom, such as school personnel, organizational factors, and personal factors (Jo et al., 2015). However, considering that teachers are exposed to the classroom environment for a significant portion of their day, it is reasonable to assume that what happens in the classroom will have a significant impact on their job satisfaction. In a recent study of Ceylan and Ozbal (2020), the influence of the educational environment on teacher job satisfaction was investigated, utilizing data from the 2018 Teaching and Learning International Survey (TALIS) in Finland, Turkey, and Italy. Their findings revealed that Finland exhibited the most positive correlation. Exploring this further, it would be valuable to examine the correlation specifically among mathematics teachers.

In particular, math teachers might find the classroom environment more challenging than other subject areas, given that math classrooms are often characterized by social issues such as negative attitude toward mathematics. Nevertheless, few studies have considered factors within the classroom, such as the mathematics classroom environment or students' attitudes toward the classroom.

In order to respond to our research question, we selected TIMSS 2019 for our secondary data analysis, as it constituted the most recent collection of a substantial education dataset that aligned with our research objectives.

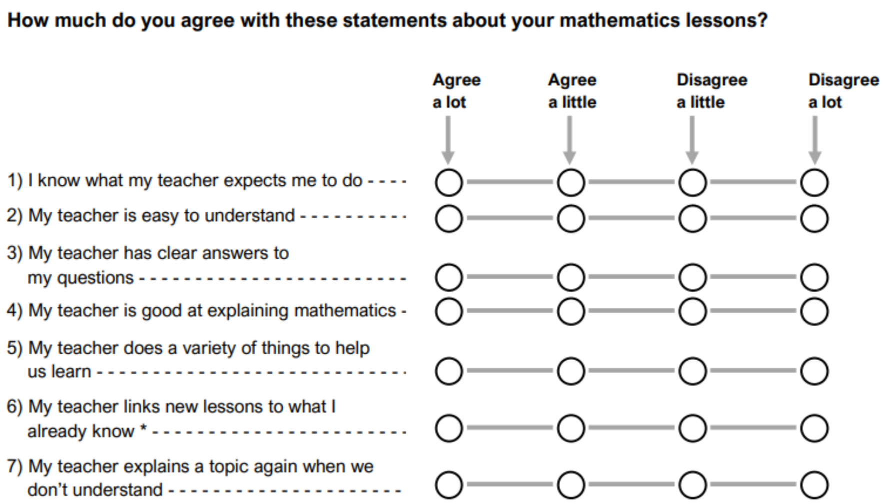
TIMSS 2019 asked questions about *classroom factors related to math learning* for grades 4 and 8. As sub-questions, the student survey of TIMSS 2019 had seven questions about the clarity of lessons (BSBM17A through BSBM17F) and six questions about the classroom environment (BSBM18A through BSBM18F). In addition, the teacher survey in TIMSS 2019 had five questions about teacher job satisfaction (BTBG08A through BTBG08E).

Meanwhile, TIMSS 2019 analyzed responses to individual questions on each topic into a single scale to define new variables. For example, the item response theory was applied to the seven questions about the clarity of lessons to create one scale called BSBGICM(Fishbein et al., 2021). In addition, TIMSS 2019 provided scores of BSBGICM for each country to enable international comparisons. The variables presented in TIMSS 2019 were defined by detailed questions with a clear purpose; therefore, the meaning of the examined variable "clarity of instruction" may be different from what people think "clarity of instruction" means, and this should be kept in mind when utilizing and interpret those variables in research. Our main interest, teachers' job satisfaction, was also provided as a single scale (BTBGTJS).

The average teacher job satisfaction score in South Korea was 9.1(Mullis et al., 2020), which is relatively low compared to other countries. Although both teachers and students shared the same classroom environment, students' academic achievement was among the highest in the world, while teachers' job satisfaction was at the bottom of the list.

Methods of Case 1

To answer Research Question 1, the researchers used *the Instructional Clarity in Mathematics Lessons scale* (BSBMICM) and *the Disorderly Behavior during Mathematics Lessons scale* (BSBGDML)" as independent variables and *Teacher Job Satisfaction* (BTBGTJS) as the dependent variable. However, the design of the TIMSS 2019 data, where one teacher was associated with multiple students, created a situation where one dependent variable had multiple independent variables for the analysis at the teacher level. In other words, BSBMICM and BSBGDML – students' variables – can be considered as distributions for each teacher, not values. Thus, we wanted to represent this distribution with one appropriate representative value.


How much do you agree with these statements about your mathematics lessons?

	Agree a lot	Agree a little	Disagree a little	Disagree a lot
1) I know what my teacher expects me to do	○	○	○	○
2) My teacher is easy to understand	○	○	○	○
3) My teacher has clear answers to my questions	○	○	○	○
4) My teacher is good at explaining mathematics	○	○	○	○
5) My teacher does a variety of things to help us learn	○	○	○	○
6) My teacher links new lessons to what I already know *	○	○	○	○
7) My teacher explains a topic again when we don't understand	○	○	○	○

Figure 1. The questionnaire of *the Instructional Clarity in Mathematics Lessons scale* (Mullis et al., 2020, p.459)

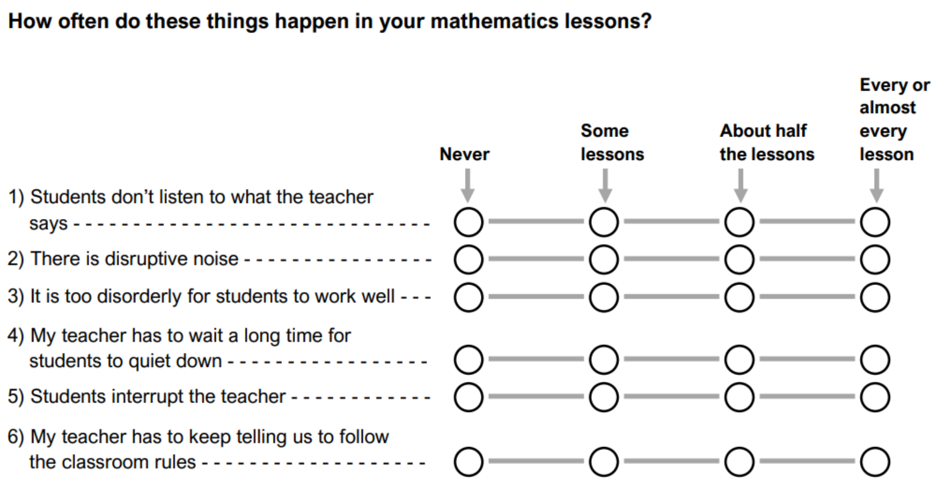


Figure 2. The questionnaire of *the Disorderly Behavior during Mathematics Lessons scale* (Mullis et al., 2020, p.463)

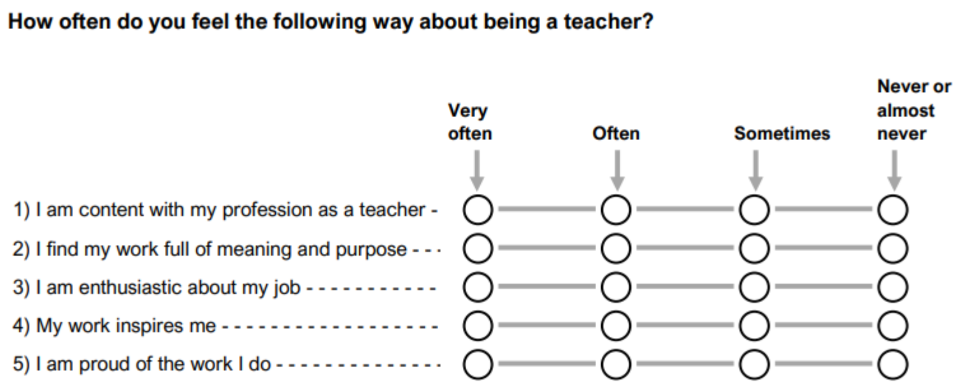


Figure 3. The questionnaire of *Teacher Job Satisfaction* (Mullis et al., 2020, p.405)

To address this issue, we first used the mean of the two independent variables as the representative value for the analysis. We then used the statistical program R 4.2.3. to conduct a linear regression analysis between the representative values of the independent variables and the dependent variable, teacher job satisfaction, at the teacher level. We also used the package *lme4* (Bates et al., 2015) to fit the model for mixed effects.

To answer Research Question 2, we used the same independent variables, BSBMICM and BSBGDML, as important classroom factors. We also selected Finland, one of the countries with the best educational indicators in PISA and TIMSS, to compare with our results, but unlike research question 1, we performed linear regression analysis on the independent variables by country without considering proxy values. All linear models included the interactions terms of the independent variables.

Results of Case 1

The regression analysis results showed that the interaction between the classroom environment factors did not show statistical significance. In addition, all other terms in the linear model were not statistically significant at the alpha level of 0.05 (see Table 1). Since prior studies (e.g., Collie et al., 2012) had indicated a correlation between these classroom variables and teachers' job satisfaction, we had expected a significant correlation to emerge. However, the results were contrary to expectations. Thus, the research team began to consider reasons for these unexpected findings.

Table 1. Regression analysis results between classroom environment factors and teacher job satisfaction (BTBGTJS)

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	16.531	9.604	1.721	0.086
BSBGICM	-1.002	1.081	-0.927	0.355
BSBGDML	-0.787	0.883	-0.892	0.373
BSBGICM:BSBGDML	0.105	0.098	1.073	0.284
Multiple R-squared: 0.021, Adjusted R-squared: 0.008				

A primary reason identified was that the representative values from students' data to create the independent variables in the analysis did not properly reflect their distributional characteristics, and the mean differences between groups were not significant enough to impact the results as expected. Although the answer to Research Question 1 was inconclusive, the researchers still hypothesized that classroom environment factors would influence teachers' teaching activities and sought clues in the process of answering Research Question 2.

To answer Research Question 2, the international comparative study between South Korea and Finland, widely recognized as an educational leader, was conducted. Initially, like Research Question 1, the dependent variable BTBGTJS representing teachers' job satisfaction was used. However, Finish teachers also showed no significant correlation between classroom variables and teacher's job satisfaction, the dependent variable was changed to BTBGLSN, which relates to restrictions in teaching activities due to students being unprepared for lessons. The additional regression analysis results using the changed dependent variable are seen in Tables 2 and 3.

Table 2. Regression analysis between classroom environment factors and teacher job satisfaction (BTBGLSN) in Finland

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	9.275	0.584	15.857	<0.001
BSBGICM	-0.074	0.059	-1.249	0.211
BSBGDML	0.011	0.057	0.204	0.838
BSBGICM:BSBGDML	0.016	0.005	2.843	0.004
Multiple R-squared: 0.021, Adjusted R-squared: 0.008				

Table 3. Regression analysis between classroom environment factors and teacher job satisfaction (BTBGLSN) in Korea

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	8.349	0.720	11.590	<0.001
BSBGICM	0.011	0.079	0.151	0.880
BSBGDML	0.063	0.067	0.941	0.347
BSBGICM:BSBGDML	<0.001	0.007	0.014	0.989
Multiple R-squared: 0.021, Adjusted R-squared: 0.008				

According to the regression analysis results, statistical significance was observed in the Finland model, while no such significance was found in the Korea model. However, the researchers found it difficult to agree that statistical significance necessarily indicated meaningful results, for the following reasons:

The primary reason is that the new dependent variable BTBGLSN, which includes factors such as 'lack of prerequisite student learning', 'nutritional deficiencies', 'lack of sleep', 'absence from class', 'violent students', 'indifference', 'mental and psychological disorders', 'difficulties with the language of instruction', encompasses rather extreme situations of classroom limitations. These issues are somewhat distant from the main focus of the researchers, which is the classroom problems in Korea.

Discussion of Case 1

The key issue shown in the case was the alignment between the theoretical frameworks of the original studies that generated the data and the secondary analyses that apply the data to new research questions. This misalignment can arise because the original data collection was guided by specific research objectives, which may not necessarily match those of secondary studies. For example, variables designed to measure one concept in the primary study might be repurposed to measure a different concept in a secondary analysis, leading to potential mismatches in theoretical underpinnings and operational definitions. These issues underscore the importance of understanding the original context in which the data was collected, including the definitions of variables and the conditions under which data was gathered.

One of the intricate challenges in the case of secondary analysis is the creation and interpretation of newly generated variables, particularly when the original dataset lacks specific measures needed for current research questions. When researchers derive new variables, such as aggregating scores or behaviors across multiple teachers and students, they often encounter difficulties in ensuring that these variables validly represent the underlying constructs they aim to measure. For instance, calculating a mean or median to represent a classroom dynamic based on diverse student responses can be misleading if the aggregation dilutes significant behaviors or outliers that are crucial for understanding the educational environment. This problem is exacerbated when the connection between teachers and students is not consistent across the dataset, making it challenging to track or interpret influences or outcomes accurately. Generating these new variables without a strong theoretical rationale can lead to data that are difficult to interpret and potentially irrelevant to the research objectives, thus compromising the validity and applicability of the findings. Hence, careful consideration must be given to how these variables are constructed and used within the broader framework of the study's goals and the theoretical context of the existing data.

Our analysis results highlight significant challenges in international comparison studies, particularly the difficulty in understanding and accurately interpreting educational practices and outcomes across different cultural contexts. In our binational comparison study between Korea and Finland, we faced substantial difficulties. Finland was selected for its renowned best practices in education. However, our research team lacked a comprehensive understanding of the Finnish education system and the nuances of teachers' job satisfaction in Finland.

This lack of expertise made it almost impossible for us to accurately interpret the Finnish results and evaluate the validity of our interpretations. Additionally, we struggled with interpreting newly generated variables specific to the Finnish context. These challenges underscore the importance of conducting and interpreting international comparison studies with a keen awareness of cultural and contextual differences. Our experience demonstrates that without this understanding, the interpretation of data can be misguided, and the conclusions drawn can be severely compromised. It highlights the necessity of deep cultural and contextual knowledge to ensure that the findings of such studies are valid and meaningful.

III. CASE 2: SECONDARY ANALYSIS WITH SELS

Meanwhile, the research team sought to address the following Research Questions 3 and 4 using data from SELS:

- Research Question 3: How do secondary mathematics teachers' teaching methods change longitudinally?
- Research Question 4: Does a teacher's career influence changes in teaching methods?

Literature Review of Case 2

Teachers' teaching methods and instructional activities can achieve genuine professionalism when they positively affect students both cognitively and affectively (Chi et al., 2011). Teachers play a crucial role in influencing student learning and are actively involved in various aspects (like teaching-learning communities, training, consulting, subject study groups) to evolve their teaching methods and enhance their professionalism. Previous studies related to teaching methods have shown impacts on cognitive aspects like students' academic achievement, as well as affective aspects like interest, motivation, and engagement (Cheong et al., 2015; Kim, 2015; Lee, 2021).

Factors influencing mathematics achievement are constantly changing and evolving, thus necessitating longitudinal studies to predict and analyze learners' development. The need for ongoing research is highlighted by the fact that changes in teaching methods and academic achievement do not happen overnight but evolve over time. Unlike cross-sectional studies, longitudinal research can monitor changes over a certain period and explain individual differences. For instance, although the seventh curriculum emphasizing student-centered teaching methods has been in place for some time, it cannot be conclusively said that teachers in the field are implementing these methods effectively. Additionally, most research on teaching methods has been cross-sectional.

Teaching methods can actively engage students in learning and impact affective domains such as interest and beliefs about mathematics (Jang, 2019; Yu & Kim, 2020). For example, Kim (2015) found that the more teacher-centered the instruction, the higher the students' confidence in mathematics. Lee (2021) showed that instructional activities providing challenging problems and explanatory teaching methods influence students' perception of the value of mathematics.

They are also crucial in affecting students' mathematical thinking and problem-solving abilities. Teaching methods reflect the teacher's intentions and instructional design aimed at achieving educational and instructional objectives. It's crucial to see if teachers understand and reflect the intents of revised curriculums and educational policies within their teaching.

The 2015 revised curriculum likely caused changes in teachers' instructional activities and student assessments, ultimately impacting students' competencies. Research is needed to identify features observed in teachers' instructional activities and students' core competencies following the introduction of the 2015 revised curriculum policies (Kim et al., 2020). Since the seventh curriculum, the emphasis has been on communication, shifting from teacher-centered to student-centered and from lecturing to discussion-focused instruction, enhancing students as the subjects of learning.

The direction and educational objectives emphasized in the revised curriculum alter the goals that students within the classroom need to achieve, necessitating that teachers adapt their teaching methods accordingly. With the global trend towards shifting from knowledge-based to competency-based curricula, the 2015 revised curriculum in South Korea introduced core competencies for students at the national level for the first time. This revision presupposes that student core competencies are cultivated across all subjects based on instruction, emphasizing competency-based instruction tailored to subject characteristics, along with improved teaching-learning and assessment methods. Methods such as discussion, cooperative, and inquiry-based learning are necessary, with student assessments progressing process-oriented (Ministry of Education, 2015).

Teaching experience is also a critical factor in determining actual teaching methods and beliefs about pedagogy (Pang, 2002). Many teachers continuously develop their mathematics teaching methods through their teaching experiences gained throughout their teaching careers. For example, teachers develop their mathematical knowledge for teaching while conducting lessons and can change their perspectives on mathematics learning by observing students' mathematical thinking processes. Additionally, they realize the joy of learning and teaching mathematics through interactions with students and assess the success of their teaching methods based on student responses. However, it's important to carefully analyze the commonly accepted premise that 'a teacher's teaching career influences the development of teaching methods'.

Methods of Case 2

We utilized the SELS dataset due to our interest in the Seoul metropolitan area, which is the largest metropolitan region in Korea and significantly influences other areas in the country. Additionally, some of the researchers in this study are involved in this region. Specifically to address Research Question 3, it was necessary to capture the dynamic nature of teaching methods by using a subset of variables that reflect recent trends in teaching skills. Among the available mixture models, such as latent profile analysis, we chose K-means clustering for longitudinal data because it is more appropriate for separating participants into groups with distinct characteristics based on observed variables and analyzing changes in teaching methods used by teachers over time.

Out of necessity due to the availability and consistency of the data, the analysis was confined to the second, fifth, and sixth years of the study, despite initial plans to cover a more extended period from the second to the sixth year. While this limitation in the data range posed challenges, the method still successfully uncovered distinct patterns and trends in teaching methods, which were categorized into three to five discernible clusters.

During the period from 2012 to 2016, the study utilized longitudinal data to trace and assess changes in the teaching strategies employed by these educators. The goal was to understand not just the evolution of these methods but also how the experiences and career lengths of the teachers influenced these changes. This longitudinal approach provided a comprehensive view of the educational landscape over time, allowing for a nuanced analysis of pedagogical shifts. In this study, we will share the clustering results about the teaching method #5 (GT185002), utilization of supplementary materials in a lesson.

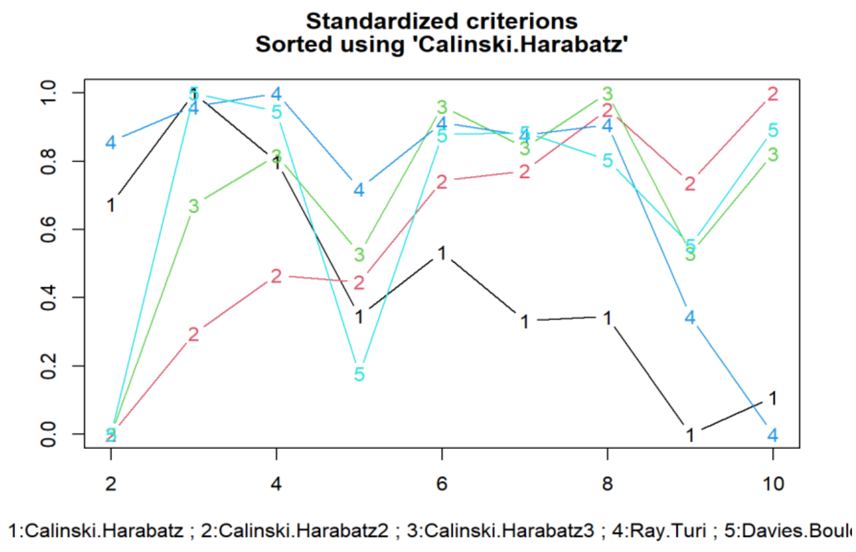


Figure 4. Model comparison

In applying the K-means clustering method, the R 4.2.3 package KmL was used, which facilitates the clustering of longitudinal data. This technique, known as Longitudinal Cluster Analysis (LCA), is particularly adept at handling data where the aim is to understand changes over time and not just static snapshots. LCA works by grouping data points based on similarities across time, thus forming clusters that represent different growth profiles or patterns. To determine the most statistically robust model, various Cluster Validity Indices (CVIs) were employed. These indices, including the Calinski-Harabatz and Davies-Bouldin, helped in evaluating the effectiveness of the clustering by assessing the separation and compactness of clusters. After careful consideration of these indices, a model with four clusters was deemed optimal, striking a balance between complexity and clarity in illustrating the shifts in teaching practices over the years.

Results of Case 2

In Figure 5, Group A began slightly above the midpoint on the scale, showing consistent and slight decreasing over the first four years, approaching the highest level of agreement by the end. Group B started around the midpoint and this group showed a steady rise, reaching near maximum agreement by the final year. Group C started out with scores just around the midpoint, showing little change for the first three years. However, Group C reported markedly increasing to nearly the level of agreement by the sixth year, indicating a slow but significant shift in stance over time. Lastly, Group D began near neutral, dips closer to agreement in the fifth year, and then back to the midpoint in the final year. Each group demonstrated a unique trajectory in terms of changing attitudes or acceptance levels regarding the subject matter assessed over the period.

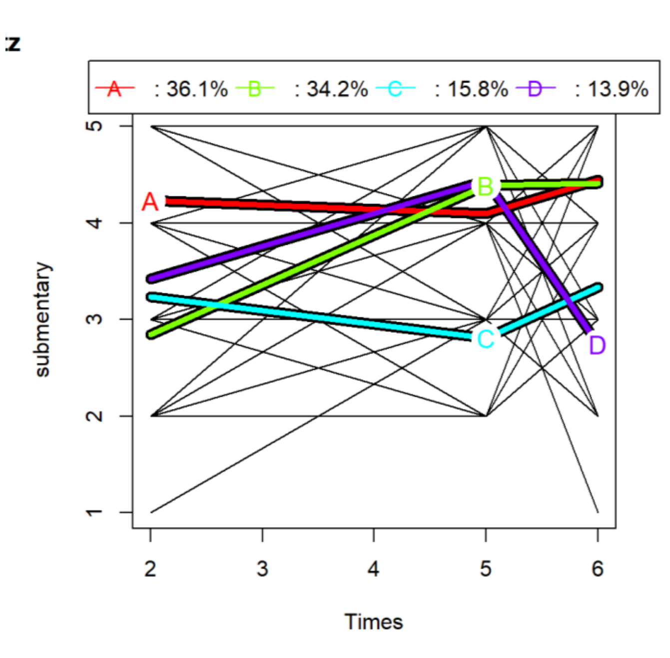


Figure 5. 4-Cluster model results

It was interesting to see that Groups A and C showed parallel patterns, with a slight decrease in the utilization of supplementary materials, followed by an increase at the end of the observed years. Additionally, Groups B and D showed similar increases in the utilization of supplementary materials from Year 2 to Year 5, but only Group D regressed afterward. This raised an intriguing question about what causes the differences between Groups B and D. It is important to note that the x-axis represents stages of the SELS, not teaching experience. Therefore, the research team planned to examine the factors that differentiate Groups A and C, as well as Groups B and D, in the next phase of their study.

Discussion of Case 2

The research team initially questioned why they were unable to locate teachers' data for years 3 and 4. After completing their analysis, they returned to examine the data structure and discovered inconsistencies in ID structures over time. This issue came to light following an inquiry with the Seoul Metropolitan Office of Education's Educational Research and Information Service. The investigation revealed that the same teacher ID might represent different individuals in different years, thereby compromising the longitudinal integrity of the dataset. Unlike student data, which generally allows for more consistent tracking, the teacher data in SELS does not maintain this continuity, posing significant challenges for longitudinal studies. Thus, we concluded that the same teacher ID across different years does not necessarily represent the same individual.

One of the significant issues encountered in the use of longitudinal data relates to the misunderstanding of the data structure, particularly the non-uniformity of ID structures over the years. This was highlighted during a detailed inquiry with the Seoul Metropolitan Office of Education's Educational Research and Information Service, which revealed a critical caveat in the SELS dataset. Despite the appearance of continuity in teacher IDs across different years, it turned out that the same ID might correspond to different teachers from year to year. This inconsistency undermined the longitudinal integrity of the dataset, as the teacher data did not track the same individuals over time. Unlike student data, where individual tracking might be more consistent, the teacher data in SELS lacked this longitudinal continuity. This revelation was crucial for the researchers relying on these IDs for longitudinal analysis, as it posed significant challenges in interpreting changes and trends over time within the educational workforce. Such challenges necessitated rigorous checks and balances in data handling and analysis to avoid erroneous conclusions based on faulty assumptions about data continuity.

IV. CONCLUSION

The purpose of this study was to illustrate the challenges and complexities encountered when conducting secondary analysis and international comparsion using large-scale public data sets, particularly those related to mathematics education. Large-scale data sets such as the Trends in International Mathematics and Science Study and the Seoul Educational Longitudinal Study provided invaluable resources for the researchers looking to derive educational implications and develop theories. However, these opportunities also came with significant methodological hurdles that can affect the validity and applicability of research findings.

As found in the literature review, secondary analyses often struggle with the issue of overly narrow operational definitions of variables. When variables are defined too narrowly, they may not capture the breadth of the concept as intended in the new research framework, limiting the scope of the findings and potentially leading to incorrect interpretations. This becomes particularly problematic in educational research, where complex constructs such as teaching quality or student engagement are reduced to simple, quantifiable measures that may not fully represent the underlying phenomena.

The main issue highlighted in Case 1 is the alignment between the theoretical frameworks of the original studies that generated the data and the secondary analyses that apply the data to new research questions. Misalignment can occur because the original data collection was guided by specific research objectives, which may not align with those of secondary studies. For instance, variables designed to measure one concept in the primary study might be repurposed to measure a different concept in a secondary analysis, leading to potential mismatches in theoretical underpinnings and operational definitions. This emphasizes the importance of understanding the original context in which the data was collected, including the definitions of variables and the conditions under which the data was gathered.

Another complex challenge in secondary analysis is creating and interpreting newly generated variables, especially when the original dataset lacks specific measures needed for current research questions. Researchers often struggle to ensure that these new variables validly represent the underlying constructs they aim to measure. For example, aggregating scores or behaviors across multiple teachers and students can be misleading if it dilutes significant behaviors or outliers that are crucial for understanding the educational environment. This problem is further complicated when the connection between teachers and students is inconsistent across the dataset, making it difficult to track or interpret influences accurately. Generating these new variables without a strong theoretical rationale can result in data that are hard to interpret and potentially irrelevant to the research objectives, compromising the validity and applicability of the findings. Therefore, careful consideration must be given to how these variables are constructed and used within the broader framework of the study's goals and the theoretical context of the existing data.

The research team encountered significant challenges with the longitudinal integrity of teacher data in the SELS dataset. Initially puzzled by missing data for years 3 and 4, they discovered inconsistencies in ID structures upon further investigation. An inquiry with the Seoul Metropolitan Office of Education's Educational Research and Information Service revealed that the same teacher ID could represent different individuals across different years, unlike the more consistent tracking available in student data. This misalignment compromised the ability to conduct accurate longitudinal studies, as it became clear that the same teacher ID did not necessarily refer to the same person over time. This issue highlighted the need for rigorous checks and understanding of data structures to ensure valid longitudinal analysis.

Our analysis results underscored significant challenges in international comparison studies, particularly in understanding and interpreting educational practices across cultural contexts. In a binational comparison study between Korea and Finland, we encountered substantial difficulties due to our limited understanding of the Finnish education system and teachers' job satisfaction. This lack of expertise made it nearly impossible to accurately interpret the Finnish results and evaluate the validity of our interpretations. Additionally, interpreting newly generated variables specific to the Finnish context was challenging. These issues highlight the necessity of deep cultural and contextual knowledge to ensure the validity and meaningfulness of such studies. Without this understanding, data interpretation can be misguided, leading to compromised conclusions and ineffective educational changes.

In conclusion, while secondary analysis of large-scale public datasets offers a rich vein of potential insights for educational research, it requires meticulous attention to the theoretical alignment, operational definitions, and statistical interpretations. Researchers must rigorously assess the compatibility of the original data with their theoretical frameworks and research objectives. They should be cautious of overinterpretation of statistically significant results that may not hold practical significance. By addressing these issues, researchers can better leverage the power of secondary analysis to contribute valuable findings to the field of mathematics education.

REFERENCES

Altman, N., & Krzywinski, M. (2015). Association, correlation and causation. *Nature Methods, 12*(10), 899-900. https://doi.org/10.1038/nmeth.3587

Banks, G., Kepes, S., & Banks, K. (2012). Publication bias: The antagonist of meta-analytic reviews and effective policymaking. *Educational Evaluation and Policy Analysis, 34*(3), 259–277. https://doi.org/10.3102/0162373712446144

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. http://dx.doi.org/10.18637/jss.v067.i01.

Castle J. (2003). Maximizing research opportunities: Secondary data analysis. *Journal of Neuroscience Nursing, 35*(5), 287-290.

Ceylan, E., & Özbal, E. Ö. (2020). The effects of extrinsic and intrinsic factors on teacher' job satisfaction in TALIS 2018. *International Online Journal of Primary Education, 9*(2), 244-259.

Cheong, M., Kim, H., & Moon, Y. (2015). The relationship between teaching methods accepted by learners and academic achievement factors on academic achievement. *Korean Journal of Youth Studies*, *22*(7), 129-150.

Chi, E., Yang, M., & Cheong, Y. (2011). Effects of teachers' activities for instruction and evaluation on students' self-regulated Learning and academic achievement. *The Journal of Elementary Educatio, 24*(4), 165-184.

Collie, R., Shapka, J., & Perry, N. (2012). School climate and social–emotional learning:

Predicting teacher stress, job satisfaction, and teaching efficacy. *Journal of Educational Psychology, 104*(4), 1189–1204.

Fishbein, B., Foy, P., & Yin, L. (2021). TIMSS 2019 user guide for the international database (2nd ed.). Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-database/

Hatano, G., & Inagaki, K. (1998). Cultural contexts of schooling revisited: A review of the learning gap from a cultural psychology perspective. In S. G. Paris & H. M. Wellman (Eds.), *Global prospects for education: Development, culture and schooling* (pp. 79-104). American Psychological Association.

Henson, R., & Roberts, J. (2006). Use of exploratory factor analysis in published research: Common errors and some comment on improved practice. *Educational and Psychological Measurement, 66*(3), 393–416. https://doi.org/10.1177/0013164405282485

Holliday, W. G., & Holliday, B. W. (2003). Why using international comparative math and science achievement data from TIMSS is not helpful. *The Educational Forum, 67*(3), 250–257. https://doi.org/10.1080/00131720309335038

Hwang, J., & Shin, D. (2021). Effects on equity in mathematics education: Multilevel analysis via the PISA 2015. *The Mathematics Education, 60*(4), 451-466. https://doi.org/10.7468/mathedu.2021.60.4.451

Jang, J. (2019). *Analysis of teachers' instructional ability and method type recognized by students and influence factor* [Unpublished doctoral dissertation, Korea University].

Jeong, S., Hwang, J., & Ahn, J. (2023). Investigating students' profiles of mathematical modeling: A latent profile analysis in PISA 2012. *Research in Mathematical Education, 26*(3), 235–252. https://doi.org/10.7468/JKSMED.2023.26.3.235

Jo, D., Ka, S., Yu, J., & Kim. J. (2015). A study of factors influencing teachers' job satisfaction. *The Journal of Korean Education, 42(*1), 51-80.

Kim, E., Kim, G., Lee, S., Kim, K., & Yoon, Y. (2020). The influence of 2015 curriculum policy on the teachers' teaching activities and students' key competencies. *The Journal of Politics of Education, 27*(2), 169-188. https://doi.org/10.52183/KSPE.2020.27.2.169

Kim, S. H., Kim, K., & Lee, J. (2005). Comparison of features of mathematically gifted, scientifically gifted and common students in cognitive, affective and emotional aspects. *The Mathematical Education, 44*(1), 113-124.

Kim, S. (2015). An analysis of teacher effects on fourth-grade students' attitudes toward mathematics based on TIMSS 2011 results. *The Mathematical Education, 54*(2), 195-206. https://doi.org/10.7468/mathedu.2015.54.2.195

Lee, S. (2021). *A study on teachers' instructional strategies affecting students' mathematical value* [Unpublished master's thesis, Hanyang University].

Lindström, M., Johansson, S., & Borger, L. (2024). Does formal teacher competence matter for students' mathematics achievement? Results from Swedish TIMSS 2019. *Educational Research and Evaluation*, 1–30. https://doi.org/10.1080/13803611.2024.2367486

Ministry of Education (2015). *2015 revised mathematics education curriculum*. Ministry of Education. http://ncic.re.kr/mobile.dwn.ogf.inventoryList.do

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: https://timssandpirls.bc.edu/timss2019/international-results/

Pang, J. (2002). A study on the factors influencing mathematics teachers' instruction. *The Mathematical Education*, *41*(3), 257-271.

Park. S., & Sang. K. (2011). Characteristics of and factors affecting on students' attitude toward mathematics. *School Mathematics*, *13*(4). 697-716.

Pearl, J. (2009). Casual inference in statistics: An overview. *Statistics Surveys*, *3*, 96-146. https://doi.org/10.1214/09-SS057

Pflugler, C., Malzer, T., Jäschke, T., Wiesche, M., & Krcmar, H. (2018). "Do I want to have losers in my team?"-A quantitative study of learning from IT project failure. In R. Torres, N. Gerhart, & A. Negahban (Eds.), *Proceedings of the 51st Hawaii International Conference on System Sciences* (pp. 4348-4537). Hilton Waikoloa Village, Hawaii, USA.

Shepherd, D., Patzelt, H., Williams, T., & Warnecke, D. (2014). How does project termination impact project team members? Rapid termination, 'Creepingdeath', and learning from failure. *Journal of Management Studies, 51*(4), 513–546. https://doi.org/10.1111/joms.12068

Tripathy J. (2013). Secondary data analysis: Ethical issues and challenges. *Iranian Journal of Public Health, 42*(12), 1478-1479.

Yu, S., & Kim, N. (2020). The structural relationship of learner-centered instruction, feedback experience, affective achievement, and mathematical competencies by middle schoolers. *Journal of Educational Evaluation*, *33*(4), 791-814. http://dx.doi.org/10.31158/JEEV.2020.33.4.791