Open Access
Lavai-Int. J. Educ.
Vol. 21, No. 32: p.41-52 June 2025
https://doi.org/10.5564/lavai.v21i32.4030
ISSN 2410-4507, eISSN 2959-9334

L A V A I
International Journal of Education

Original article

Using Microsoft Excel for Lesson Transcript Analysis

Gereltuya Tsereljav[1,*], Ganbaatar Tumurbaatar[2], Badrakh Jadamba[2], Battogtokh Tsiyen-Oidov[1]

[1]*Department of Informatics, School of Mathematics and Natural Sciences, Mongolian National University of Education, Mongolia*
[2]*School of Mathematics and Natural Sciences, Mongolian National University of Education, Mongolia*

A R T I C L E I N F O

History:
Received: 20 December, 2024
Revised: 12 April, 2025
Accepted: 02 May, 2025

Keywords:
Lesson analysis
Transcript based lesson analysis
Tools

A B S T R A C T

Transcript-Based Lesson Analysis (TBLA) involves transcribing entire lesson recordings into verbatim notes, followed by both quantitative and qualitative analyses to derive meaningful conclusions. This process demands substantial effort and time on the part of the researcher. For example, the researcher must listen to the lesson recording multiple times and transcribe every word spoken by both the teacher and students verbatim, ensuring complete accuracy without omissions or errors. Therefore, automating the TBLA process is essential not only for enhancing the reliability and accuracy of research findings but also for improving researcher productivity. This study introduces an automated digital tool for Transcript-Based Lesson Analysis (TBLA), developed using Microsoft Excel's Visual Basic for Applications (VBA) and graphical user interface (GUI) functionalities. We applied the tool to analyze an 11th-grade mathematics lesson, demonstrating that the use of this automated digital tool not only facilitates the researcher's workload but also enhances the validity and reliability of the results. In the future, the application of this digital tool can be expanded to enhance the outcomes of lesson analysis, facilitate lesson sensitivity analysis, and support the development of data-driven approaches in educational research.

1. Introduction

When analyzing a lesson using the Transcript-Based Lesson Analysis (TBLA) method, one of the established approaches to lesson analysis- a complete record of the entire lesson process is produced. The TBLA involves transcribing the entire lesson recording into verbatim notes, which are then subjected to both quantitative and qualitative analysis (Nozaki, 2022, p.306). Manual analysis of TBLA data is both time-consuming and susceptible to human error. Therefore, there is a need to use automated tools for TBLA. The study investigates methods for automating TBLA through Microsoft Excel and presents the outcomes of its application.

Corresponding author. Email: ts.gereltuya@msue.edu.mn
ORCID: https://orcid.org/0000-0003-3735-5246

To date, TBLA plays a crucial role in advancing lesson research by not only improving research outcomes but also enabling the inclusion of teachers who did not observe the lessons, thereby, enriching the interpretation and evaluation of instructional practices. Transcript-Based Lesson Analysis (TBLA) is a time-consuming and labor-intensive process that places a significant burden on teacher-researchers (Wang, Chen, 2024, p.3). In response to these challenges, researchers are actively seeking ways to simplify the TBLA process through the development and application of automated lesson analysis tools. Expanding and refining TBLA represents a critical step toward enhancing the depth, rigor, and impact of educational research.

Literature Review: Transcript-Based Analysis is a methodological approach that seeks to uncover the sequence and underlying meaning of classroom phenomena through detailed lesson transcripts (Arani et al., 2017, p.245). Lesson Study is recognized as an effective approach to supporting teachers' professional development, providing a structured framework for the continuous improvement of instructional practices (Yoshida et al., 2021, p.28; Tan et al., 2024, p.3). The TBLA methodology fosters collaboration between teachers and researchers while promoting the advancement of educational research grounded in scientific evidence (Fernandez et al., 2003, p.173). TBLA is considered a valuable method for analyzing the structure of classroom discourse, evaluating student participation, and assessing teacher strategies aimed at enhancing the quality of the learning process (Arani et al., 2017, p.259; Tan et al., 2021, p.9). Although qualitative approaches have traditionally been central to TBLA, recent years have seen the development of complementary quantitative methods (Nozaki, 2022, p.307). Since 2016, research and scholarly works in the field of Transcript-Based Lesson Analysis have been actively conducted in Mongolia. In his study "Methodological Study of Measuring Student Participation in Lessons," researcher T.Ganbaatar explored the characteristics of TBLA and presented a methodology for measuring and evaluating student participation in regular lessons (Ganbaatar, 2016, p.6). In their article "Results of Studying the Use of Modality Meaning of the Word through Transcript-Based Lesson Analysis," researchers D. Enkhtsetseg and E. Menguntulga investigated the characteristics of modality in language as examined through the lens of Transcript-Based Lesson Analysis (Jadamba, Enkhtsetseg, Munguntulga, 2021.p.4).

Using MS Excel in Educational Research: MS Excel is widely used for data analysis. It is estimated that over 1.1 to 1.4 billion people use Microsoft Excel worldwide (Narayana, 2023, p.3; Lee, 2025, p.110). MS Excel's data visualization capabilities, including the use of charts and macros, transform coded data into meaningful representations, thereby supporting informed pedagogical decision-making and providing insights into instructional practices.

While tools such as NVivo, MaxQDA, and ATLAS.ti offer powerful capabilities for TBLA, they also present financial and technical barriers for many researchers (Zamawe, 2015, p.14). The MS Excel-based tool developed for Transcript-Based Lesson Analysis is designed to automate the analysis of lesson transcripts by systematically processing data and generating quantitative results.

2. Materials and Methods

Data Collection: In this study, we used full audio and video recordings of an 11th-grade mathematics lesson on the topic "Conical Cones and Lateral Surface Areas," conducted in a public secondary school in Ulaanbaatar. We conducted our survey during the academic year of 2023-2024. The topic "Conical Cones and Lateral Surface Areas" was selected for its relevance to the 11th-grade mathematics curriculum. It enables students to apply geometric principles in a practical context, fosters dynamic teacher-student interactions, and aligns with key learning objectives. This makes it ideal for TBLA, as it allows exploration of how mathematical concepts are communicated and comprehended, while also assessing the effectiveness of question-based teaching strategies in enhancing student engagement and understanding. Since transcribing audio and video files manually is time-consuming, the recordings were converted into text-based notes using the Chimege API.

Full lesson transcripts: Lesson transcriptions should provide a comprehensive and accurate record of all teacher-student interactions. This approach enables the analysis of key indicators, such as the total number of interactions, word count, and the types of questions posed. Transcriptions of teacher-student communication during a regular lesson are presented in Table 1, spanning a total of 10 pages.

Table 1. Transcripts of Teacher-Student Dialogue from the Lesson Titled “Conical Cones and Lateral Surface Areas.”

Speaker	Conversation number	Conversation
T	31	We have studied in the previous lesson. The wall part is a cylinder. What else? Q1
S	32	Cone
S	33	Is it called a splitted cone? Is it called a truncated cone? It has the shape of a cone.
T	34	It says there is a cone. Where is it? Q1
S	35	It is above/on top.
T	36	On the roof part, right
S	37	But not completely. It's like it was cut and put.

Data analysis: During the lesson analysis, the following participation variables were identified based on the lesson transcript (Table 2). Transcript-based lesson notes enable the extraction of variables that can be analyzed using both quantitative and qualitative analytical methods. Within the MS Excel-based automated tool developed for this study, certain variables are computed directly through integrated formulas and macros, whereas others require manual recoding prior to analysis. Table 2 presents the participation variables that can be identified from the lesson transcripts.

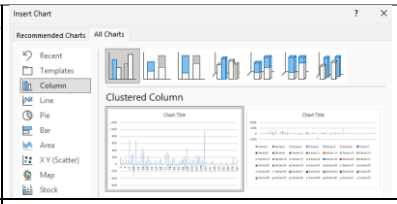
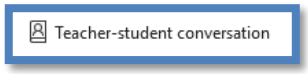
Table 2. Participation Variables Identified from Lesson Transcripts

Participation Variables Identified from Lesson Transcripts			Form
Teacher's activities	1	Length of teacher discourse	Automat
	2	Total teacher word count	Automat
	3	Teacher's questioning patterns	Automat
	4	Teacher's questioning pattern during the lesson	Automat
	5	Patterns of teacher's question types	Recode
	6	Methodology for segmentation of questions	Automat
	7	Number of questions in a lesson segment	Automat
	8	Patterns of question types across lesson segments	Automat
	9	Results of open and closed-ended questions	Recode
	10	Number and frequency of keywords mentioned by the teacher in relation to lesson content (Keyword 1, Keyword 2, ...)	Automat
	11	Number of most frequently repeated words	Automat
Student's activities	1	Total number of student conversations	Automat
	2	Total Word Count Spoken by the Student	Automat
	3	Max value of number of words used in responses to questions	Automat
	4	Number and frequency of keywords mentioned by the student in relation to lesson content (Keyword 1, Keyword 2, ...)	Automat
	5	Max value of number of words used in responses to questions	Automat

Table 2 presents the variables, indicating which are automatically generated and which require manual coding. An automated library of lesson analysis tools was developed using Excel VBA and applied to each participant-related variable.

Automation tools formulas (calculations): This section considers a formula and macro automation in Excel designed to calculate the length of teacher-student conversations from lesson transcripts.

Word count	Graphics
=LEN(E8) (Measuring the length or letter count of teacher and student speech. Since the formula counts the number of letters, it always returns a positive value. The line until the end of conversation is activated). =IF(C8="Б", I5,-I5)	Select the cells in column J, then choose the 'Clustered Column' chart type from the 'Chart' group under the 'Insert' tab.

Differentiating Teacher and Student Speech	
	
'Number of Words in Teacher-Student Conversation' command counts the words spoken by both the teacher and the student.	'Teacher-Student Conversation Graph' command visualizes the teacher's and student's speech in a graph.

Sub Tdialogi()
Dim ws As Worksheet
Dim lastRow As Long
Dim i As Long
Dim cValue As String
Dim gValue As Double
Set ws = ThisWorkbook.Sheets("TBLA")
lastRow = ws.Cells(ws.Rows.Count, "E").End(xlUp).Row
For i = 8 To lastRow
ws.Cells(i, "I").Value = Len(ws.Cells(i, "E").Value)
Next i
For i = 8 To lastRow
cValue = Trim(ws.Cells(i, "C").Value)
gValue = ws.Cells(i, "I").Value
If IsNumeric(gValue) Then
If cValue = "T" Then
ws.Cells(i, "J").Value = gValue
Else
ws.Cells(i, "J").Value = -gValue
End If
Else
ws.Cells(i, "J").Value = 0
End If
Next i
End Sub
	Sub graph1( )
	Dim newSheet As String
	WorksheetExists newSheet
	Dim ChartSheet1 As Chart
	Set ChartSheet1 = Charts.Add
	With ChartSheet1
	Dim lastRow As Long
	lastRow = Sheets("TBLA").Cells(Sheets("TBLA").Rows.Count, "H").End(xlUp).Row
	.SetSourceData Source:=Sheets("TBLA").Range("H1:H" & lastRow)
	End With
	End Sub

The research methodology is presented as follows (Figure 1):
(1) Selection of lesson videos
(2) Transcription of the lesson audio into the provided Excel template
(3) Transcription of the audio to text using the Chimege app
(4) Development of an automation tool using Visual Basic for Applications
(5) Recoding of the questions
(6) Presentation of the results in numerical or graphical form
(7) Results and improvements

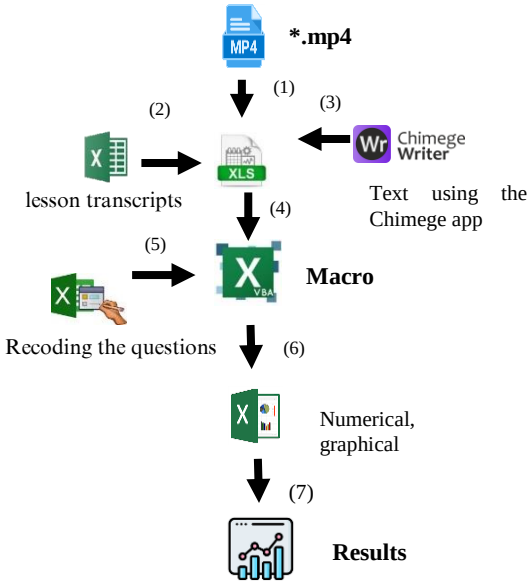


Figure 1. Transcript-Based Lesson Analysis Tool

3. Results

As part of this research, we conducted an experiment by inputting our own written and transcribed 11th-grade high school mathematics lesson transcripts into Microsoft Excel for lesson analysis.

Quantitative results: The lesson transcripts were entered into Excel, and numerical results along with corresponding graphs were automatically generated.

Table 3. Quantitative Results Derived From Lesson Transcripts

SP	CN	Conversation	Seg	WC	WC-TS	LC	LC-TS	QP	
T	31	We have studied in the previous lesson. The wall part is a cylinder. What else? Q1	Segment 1	14	14	84	84	1	11
S	32	Cone	Segment 1	1	-1	5	-5	0	11
S	33	Is it called a splitted cone? Is it called a truncated cone? It has the shape of a cone.	Segment 1	11	-11	68	-68	0	11
T	34	It says there is a cone. Where is it? Q1	Segment 1	7	7	41	41	1	12
S	35	It is above/on top.	Segment 1	3	-3	13	-13	0	12
T	36	On the roof part, right	Segment 1	4	4	26	26	0	12
S	37	But not completely. It's like it was cut and put.	Segment 1	7	-7	40	-40	0	12

SP = Speaker (Helps distinguish who is speaking: T=Teacher, S=Student); CN= Conversation Number (Line/order in the dialogue); Seg=Segment **(**Divides the

conversation into smaller sections for analysis); WC=Word Count; WS-TS=Word Count - Teacher, Student (To differentiate between the teacher's and student's speech, the length of the teacher's speech should remain as a positive value, while the length of the student's speech should be recorded as a negative value); LC=Letter count; LC-TS=Letter count - Teacher, Student; QP - Question pattern

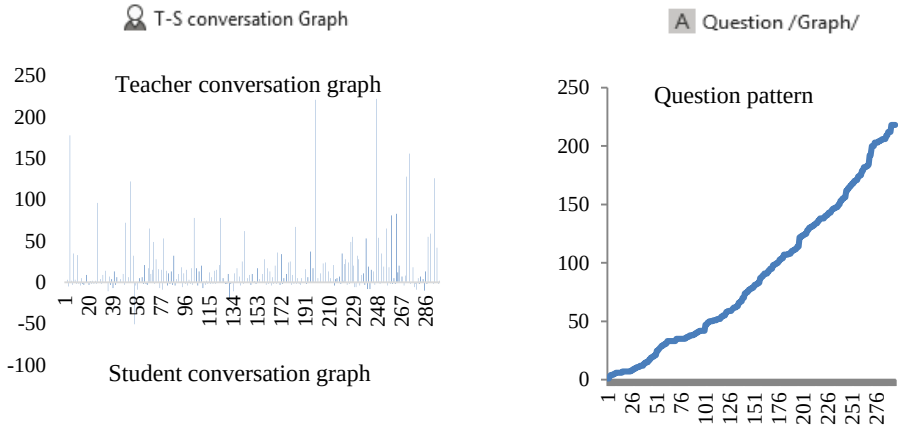


Figure 2. A: Length of Teacher-Student Conversation, B: Teachers' Question Pattern ("X-axis: dialogue sequence number / Y-axis: word count of teacher and student utterances")

Figure 2 provides an overview of teacher-student interactions during the lesson. The graph illustrates the length of these interactions, indicating which participant-the teacher or the student-dominates the conversation, when the teacher speaks less, and when student participation increases (Figure 2A). An overview of when the teacher's questions were asked during the lesson is shown in Figure 2B.

Pivot table results: Table 4 summarizes the pivottable analysis of classroom discourse, highlighting key differences between the teacher (T) and the students (S).

Table 4. Pivot Table Results

	Total number of words	Average number of words	Max of words	Min of words	Number of words	Number of question patterns	StdDev of Words
S	355	2.47	11	1	144	15	2.1
T	3975	27.23	222	1	146	198	37.7
Total	4330	14.9	222	1	290	213	

The results indicate that the teacher dominated the interaction in terms of both total word count and average utterance length. While students contributed a total of 355 words, averaging 2.47 words per speaking turn, the teacher produced 3.975 words, with an average of 27.23 words per turn. Although the number of speaking turns was nearly equal (S=144; T=146), the teacher's utterances were significantly longer and more varied, as demonstrated by the higher standard deviation. Additionally, the teacher posed 198 out of the 213 total questions recorded, compared to only 15 questions asked by students, reinforcing a teacher-centered communication pattern.

Timeline analysis of automated Excel-based analysis commands: All computations were performed using the Excel-based automated analysis tool. The experiments and data processing were conducted on an HP 15s-du2xxx laptop equipped with an Intel(R) Core (TM) i7-1065G7 CPU 1.30GHz (8 cores), 8 GB of RAM, and Microsoft Windows 11 Home Single Language 64-bit (Build 26100).

To evaluate the operational efficiency of the Excel-based lesson analysis tool, a time-based analysis was conducted. Specifically, the execution time of each macro command used to extract key participation variables from the lesson transcripts was measured and recorded. The results are presented in Table 5.

Table 5. Macro Working Time

Participation Variables Identified from Lesson Transcripts			Time/*Second*
Teacher's activities	1	Length of teacher discourse	0.002
	2	Total teacher word count	0.002
	3	Teacher's questioning patterns	0.004
	4	Teacher's questioning pattern during the lesson	0.006
	5	Patterns of teacher's question types	0.005
	6	Methodology for segmentation of questions	0.012
	7	Number of questions in a lesson segment	0.014
	8	Patterns of question types across lesson segments	0.002
	9	Results of open and closed-ended questions	0.002
	10	Number and frequency of keywords mentioned by the teacher in relation to lesson content (Keyword 1, Keyword 2, ...)	0.003
	11	Number of most frequently repeated words	0.015
Student's activities	1	Total number of student conversations	0.002
	2	Total Word Count Spoken by the Student	0.003
	3	Max value of number of words used in responses to questions	0.005
	4	Number and frequency of keywords mentioned by the student in relation to lesson content (Keyword 1, Keyword 2, ...)	0.014

The results indicate that the processing time for each macro remains exceptionally low, confirming the computational efficiency of the tool. "To evaluate the results of the two methods, the following key indicators were used for comparison:

Table 6. Comparison of Excel-Based Automation and Manual Analysis

Indicators	Excel-based automation	Manual analysis
Entering data	Fast, automated process	Slow due to manual entering
Data processing	Direct calculations and analysis performences	Formulas, graphs, and conclusions need to be made manually
Probability of error	Fewer errors (if formulated correctly)	High (possible human error)
Flexibility	Suitable for large amounts of data	More accurate with small datasets
Capacity for rational judgment	Automated analysis and visualization	Able to make logical data-based interpretations
Report and analysis	Automatic reporting capability	Long reporting period
Adding new data	Fast updates, instant calculations	Manual editing is required for each new data
Time efficiency	Fast due to multiple reuse	Time-consuming
Speed and efficiency	Very fast calculation	Time-consuming

The results indicate that Excel-based automation offers significant advantages in terms of speed, efficiency, and accuracy. Automated data entry, processing, and reporting facilitate faster operations and minimize the risk of human error, particularly when working with large datasets. In contrast, manual analysis is more time-consuming and prone to errors but may be preferable for smaller datasets that require detailed human judgment and interpretation. Overall, Excel-based automation demonstrates high effectiveness in enhancing time efficiency and ensuring accurate data analysis.

4. Discussion

Automating Transcript-Based Lesson Analysis (TBLA) presents clear benefits, including faster processing times and reduced risk of data errors. Traditional TBLA methods, often reliant on manual transcription, require significant effort from teachers or researchers, thus limiting accessibility and scalability, particularly in educational settings with limited resources. The development of an Excel-based automated tool addresses these challenges by offering a user-friendly and efficient platform for in-depth analysis. This tool facilitates a more structured and precise approach to evaluating classroom interactions, ultimately enhancing the scalability of TBLA in educational research.

While advanced AI-powered systems, such as TeachFX and TalkMoves, offer enhanced capabilities like speech recognition and real-time feedback, the Excel-

based tool offers its own set of advantages (Suresh, A et al., 2022, p.4654). One notable benefit is its transparency. Unlike AI systems, which often operate as "black boxes", the Excel tool allows educators to see and understand how data is processed clearly. Teachers can easily review the raw data and calculations, fostering greater trust in the results.

The simplicity of the Excel tool also makes it highly adaptable, as teachers can customize the analysis to focus on specific metrics that align with their instructional goals (Winarti et al., 2021, p.977). For instance, educators can modify the tool to track the frequency of certain question types or categorize students' utterances, providing valuable insights into the dynamics of classroom dialogue. In comparison to AI-based solutions like TeachFX and TalkMoves, which automate data collection and provide real-time feedback, the Excel tool's flexibility is its key strength. TeachFX, for example, uses AI algorithms to transcribe and analyze classroom discourse, providing real-time feedback on teacher-student talk ratios, question frequency, and other metrics (Petrilli, M. J., 2024. p.49). AI tools also require high levels of technical proficiency, which may limit their accessibility for teachers who are not familiar with advanced technology (Li et al., 2024. p.2). In contrast, the Excel-based tool is intuitive and easily integrated into existing classroom practices, making it an ideal choice for teachers who may not have access to or experience with more complex systems. Furthermore, although advanced systems like TeachFX can generate detailed, real-time reports, they are typically expensive and may not be available to educators in low-resource settings. The scalability of automated transcription systems further enhances the appeal of the Excel-based tool. Automated transcription technologies, such as OpenAI's Whisper model, have significantly reduced transcription time, allowing researchers and educators to shift from the time-consuming transcription task to more complex analysis (Andreyev, 2025, p. 14). The scalability of these systems also makes them ideal for longitudinal studies, large classrooms, or research that involves frequent lesson observations. These advancements in transcription technology enhance the overall effectiveness of TBLA, enabling teachers and researchers to conduct large-scale analyses with greater efficiency. Despite these advantages, the Excel tool does have limitations. While the Excel tool provides essential metrics, it cannot offer the same level of qualitative analysis that AI-based systems can, particularly in complex conversational contexts. However, the simplicity of the Excel tool, which focuses on quantifiable data like word counts and speaking time, may be sufficient for many teachers who are seeking immediate, actionable insights into classroom dynamics.

Conclusion: The results of the study indicate that using Microsoft Excel to automate the transcript-based lesson analysis method holds practical

significance. In addition to teaching users how to analyze lesson transcripts, this online resource offers the opportunity to examine and draw conclusions on various factors, including the structure and format of the lesson, teacher-student question-and-answer sessions, and strategies for improving the learning process. The study offers the advantage of enabling a comprehensive examination of the subject based on its results, while also identifying areas for improvement in the teacher's own teaching. By analyzing and refining the level and keywords of questions, teachers can enhance the quality of classroom instruction by fostering a multifaceted communication style through higher-level questioning. In the future, additional variables will be incorporated to more accurately assess psychological and behavioral involvement.

5. References

Andreyev, A. (2025). Quantization for OpenAI's Whisper models: A comparative analysis [arXiv preprint]. arXiv. https://arxiv.org/abs/2503.09905

Arani, M.R., Shibata, Y., Sakamoto, M., Iksan, Z., Amirullah, A.H. and Lander, B. (2017), "How teachers respond to students' mistakes in lessons: A cross-cultural analysis of a mathematics lesson", International Journal for Lesson and Learning Studies, 6 (3): pp. 249-267. https://doi.org/10.1108/IJLLS-12-2016-0058

Jadamba, B., Enkhtsetseg, D., Munguntulga, E. (2021). Results of studying the use of modality meaning of the word through transcript-based lesson analysis. *Quality Assurance in Higher Education, International Conference. Mongolian National University of Education*.

Fernandez, C., Cannon, J., Chokshi, S. (2003). A US-Japan lesson study collaboration reveals critical lenses for examining practice. In *Teaching and Teacher Education.* 19(2): pp. 171-185. https://doi.org/10.1016/S0742-051X(02)00102-6

Ganbaatar, T. (2016). *Methodological study on measuring student participation in the classroom*. Doctoral dissertation. Ulaanbaatar.

Lee, H. N. (2025). ChatGPT, Python, and Microsoft Excel. ChatGPT. *Evidence-Based Medicine*, 8(2): pp.110-112. https://doi.org/10.5195/jmla.2025.2065

Li, H., Xiao, R., Tseng, Y. J., Liao, G. (2024). From unseen needs to classroom solutions: Exploring AI literacy challenges & opportunities with project-based learning toolkit in K-12 education. *Proceedings of the AAAI Conference on Artificial Intelligence*. 29145-29152. https://doi.org/10.1609/aaai.v39i28.35187

Narayana, D., Ranjan, S., Tyagi, N. (2023). *Basic Computational Techniques*

for Data Analysis: Taylor & Francis Groyp: London. https://doi.org/10.4324/9781003398127

Nozaki, S. (2022). The study of students' discussion process in a teacher-developed problem-solving learning design: Transcript-based lesson analysis of a social studies lesson. *International Journal for Lesson and Learning Studies*, 11(4): pp.305-317. https://doi.org/10.1108/IJLLS-05-2022-0072

Petrilli, M. J. (2024). Next-Gen Classroom Observations, Powered by AI. *Education Next*, 24(4): pp.47-55.

Tan, S., Goei, S. L., Willemse, T. M. (2024). Global insights on lesson study in initial teacher education: A systematic literature review encompassing English, Japanese, and Chinese language sources. *Teaching and Teacher Education*, 152: pp.1-18. https://doi.org/10.1016/j.tate.2024.104791

Tan, S., Nozaki, S., Fu, H., Shibata, Y. (2021). The principles of teacher ' s decision-making in Japanese board writing (bansho) process. *Asia Pacific Journal of Education*, 43(1): pp.236-251. https://doi.org/10.1080/02188791.2021.1924119

Suresh, A., Jacobs, J., Harty, C., Perkoff, M., Martin, J. H., & Sumner, T. (2022). *The TalkMoves dataset: K-12 mathematics lesson transcripts annotated for teacher and student discursive moves*. arXiv. https://arxiv.org/abs/2204.09652

Zamawe, F. C. (2015). The implication of using NVivo software in qualitative data analysis: Evidence-based reflections. *Malawi Medical Journal*, *27*(1): pp.13-15. https://doi.org/10.4314/mmj.v27i1.4

Wang, D., Chen, G. (2024). Are perfect transcripts necessary when we analyze classroom dialogue using AIoT?. *Internet of Things*, 25: 101105. https://doi.org/10.1016/j.iot.2024.101105

Winarti, A., Saadi, P., & Rajiani, I. (2021). Applying transcript based lesson analysis in enhancing communication pattern between teacher and students in chemistry classroom. *European Journal of Educational Research*, 10(2): pp.975-987. https://doi.org/10.12973/eu-jer.10.2.975

Yoshida, N., Matsuo, N., Matsuda, M., Sato, Y. (2018). *Analysis and interpretation of lessons through the collaboration between university and school: Historical approach to the Lesson Study in Japan and a case study for the integrated perspectives*. Bulletin of the Graduate School of Education, Hiroshima University. Part III, Education and Human Science, 67: pp.27–36. https://doi.org/10.15027/46804