What is taught and learnt on confidence interval? A case study

Kyeong-Hwa Lee, Chang-Geun Song, Yeong Ok

What is taught and learnt on confidence interval? A case study

Kyeong-Hwa, Lee[1], Chang-Geun, Song[2] and Yeong Ok, Chong[3]

[1]Seoul National University, South Korea; khmath@snu.ac.kr

[2]Seoul National University Graduate School, South Korea; riquelmes@snu.ac.kr

[3]Gyeongin National University of Education, South Korea; yochong@gin.ac.kr

This study investigated what is taught and learnt with regard to relations of confidence interval and sample size when students were facilitated to reason about the graph of sampling distribution of sample means. For this aim, a teaching experiment was designed and implemented by researchers and teachers in the model of co-learning inquiry. Based on anthropological theory of the didactic, the praxeological analysis was conducted. Observed similarities and differences between the taught and learnt praxeology were discussed.

Keywords: Anthropological theory of the didactic, didactic transposition, confidence interval, sample size

Introduction

The concept of confidence interval is essential in inferential statistics. However, many students are in difficulty with understanding confidence interval (Hagtvedt et al., 2008). In particular, some research reported students' confusing on the relationship between confidence interval and other related concepts such as sample size. For example, Fidler and Cumming (2005) disclosed students' misconceptions that the width of confidence interval increases as sample size increases or is not affected by sample size. However, little research was carried on the teaching and learning of confidence interval (Pfannkuch et al., 2012).

To understand confidence interval, sample size, and their relationships, opportunities should be given for students to engage in mathematical and statistical reasoning (delMas, 2004). This study investigated what is taught and learnt regarding relations between confidence interval and sample size when students were facilitated to reason about the sampling distribution of sample means (Chance et al., 2004). For this aim, a teaching experiment was conducted in an upper secondary classroom. The framework of the Anthropological Theory of the Didactic (ATD) was used in the analysis (Bosch & Gascón, 2014). This paper reports on some preliminary results of the analysis.

Theoretical Background

Anthropological Theory of the Didactic

ATD is a theoretical approach characterized by its institutional perspective on mathematical knowledge. This view builds on the notion of didactic transposition (Chevallard & Bosch, 2014), which assumes that knowledge in classroom setting originates in the scholarly institution and is transformed as it adapts to different institutions. Bosch and Gascón (2014, p. 70) proposed four main institutions related to school didactic system (Figure 1). Our study focuses on the transposition between taught knowledge of teachers to learnt knowledge of students.

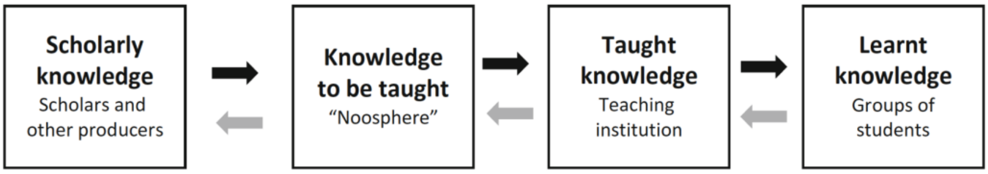


Figure 1: The didactic transposition process (Bosch & Gascón, 2014, p. 70)

In ATD, mathematical activities within each institution are modeled by the notion of *praxeology*. A praxeology consists of four intertwined elements, type of tasks (T), technique (τ), technology (θ), and theory (Θ). *Type of tasks* denotes similar problems which are solved with certain *technique*. The technique is produced, justified, explained, or reasoned by the discourse named *technology*, and the technology, in turn, is justified by *theory*.

According to Bosch and Gascón (2014), researchers should take their own epistemological stance on the knowledge they address to avoid bias in the analysis. The stance is formulated as a Reference Epistemological Model (REM) and is expressed by a praxeology. We briefly introduce our REM on the concept of confidence interval below.

REM on confidence interval

The idea of confidence interval is founded on Frequentism, as opposed to Bayesianism (VanderPlas, 2014). To the frequentist, a probability is defined as the relative frequency of occurrence of a repeatable event in the long run. In this sense, confidence intervals with C% confidence level mean that we would expect C % of the confidence intervals under repeated sampling to contain the fixed value of the parameter.

To illustrate the relationship between confidence interval and sample size, we further need another probability theory in our REM. The central limit theorem (Dekking et al., 2005, p. 197) implies that $\bar{X} \backsim N(\mu, \frac{\delta^2}{n})$ holds for the sample mean $\bar{X}$, the population mean μ, the (known) standard deviation of population δ, and sufficiently large sample size n. From this property, we obtain a formula, $P\left(-z\alpha/_2 \leq \frac{\bar{X}-\mu}{\delta/\sqrt{n}} \leq z\alpha/_2\right) = 1-\alpha$, where $z\alpha/_2$ is the $(1-{}^{\alpha}/_2)$ quantile of the standard normal distribution. Behar and Yáñez (2009) introduced two logically equivalent versions of this formula:

$$P(\bar{X}-d \leq \mu \leq \bar{X}+\mathrm{d}) = 1-\alpha \cdots (1) \qquad P(\mu-d \leq \bar{X} \leq \mu+d) = 1-\alpha \cdots (2)$$

where $d = z\alpha/_2 \frac{\delta}{\sqrt{n}}$ refers to the margin of error. Formula (1) can be interpreted as the probability of the parameter μ being in the interval $[\bar{X}-d, \bar{X}+\mathrm{d}]$, called *confidence interval*, equals to $1-\alpha$. Contrasting to formula (1), formula (2) represents that the probability of the sample mean $\bar{X}$ being in the interval $[\mu-d, \mu+\mathrm{d}]$ equals to $1-\alpha$. The probability theory shows that the probability of $\bar{X}$ lying in the interval $[\mu-d, \mu+\mathrm{d}]$ is equal to the integral of probability density function of $\bar{X}$ over the interval $[\mu-d, \mu+\mathrm{d}]$ (Dekking et al., 2005, pp. 57-58). The aforementioned meanings of confidence interval, confidence level, and probability equations (1), (2) constitute the theory Θ in our REM.

Building on the theory Θ, we can consider two techniques and their technology about the type of tasks of finding the relationship between the width of confidence interval and sample size. Based on the formula (1), the first technique $\tau_{algebra}$ refers to applying the equation $l = (\bar{X} + d) - (\bar{X} - d) = 2z_{\alpha/2}\frac{\delta}{\sqrt{n}}$, which is often found in curriculum materials (e.g. Lee et al., 2011, p. 181). By proportional reasoning (Post et al., 1988), which includes interpreting of the algebraic representation of proportionality ($\theta_{algebra}$), it is explained that the increasing sample size n decreases width of confidence interval. Another technique τ_{graph} can be found on the formula (2). τ_{graph} refers to comparing the area under graphs of sampling distribution within $[\mu - d, \mu + \mathrm{d}]$ for different sample sizes. This technique requires justification by reasoning about statistical or probabilistic idea including "describing what a sampling distribution would look like for different sample sizes based on shape" ($\theta_{graph,1}$) and "interpreting areas under the sampling distribution curve as probability statements about sample means". ($\theta_{graph,2}$) (Chance et al., 2004, p. 301). Also, the explanation about relations between $[\mu - d, \mu + \mathrm{d}]$ and confidence interval is included in the technology (Behar & Yáñez, 2009) ($\theta_{graph,3}$). These kinds of technology should be integrated to justify τ_{graph} as follows. Since the spread of sampling distribution decreases as sample size increases, the margin of error d satisfying the area under the sampling distribution over $[\mu - d, \mu + d]$ equals to $1 - \alpha$ become smaller as sample size increases and, accordingly, confidence interval $[\bar{X} - d, \bar{X} + \mathrm{d}]$ become narrower as sample size increases.

Method

This study is qualitative research performed through a case study of a designed teaching experiment (Yin, 2014). We choose our design-based research methodology based on the model of *co-learning inquiry* (Jaworski, 2004). In this model, teachers are treated as *partners* of researchers rather than passive participants of research, and teachers and researchers collaboratively design, implement, and evaluate lessons in the iterative process. The main power of co-learning inquiry model is that the result of the research can contribute to "sustainable new practices" of teachers, since "teachers engage in their own purposeful activity" in this model (ibid., p. 19).

Participants in this study were a high school teacher from a public upper secondary school in Korea and his 35 students. Around the topic of statistical inference, five lessons were designed and implemented. Audio and video recordings of the lessons and students' worksheets were collected.

In this study, we analyzed the last lesson of which learning objective was to understand the relations among confidence interval, confidence level, and sample size. We emphasize that students were encouraged to use graphs of sampling distribution to explain these relations. The collected data were analyzed by praxeological analysis (Bosch & Gascón, 2014). The focus of the analysis was on the didactic transposition between taught knowledge to learnt knowledge about relations between confidence interval and sample size. Based on our elaborated REM on confidence interval (Table 1), we analyzed the four elements of the taught and learnt praxeology and revealed their similarities and differences. The designed activities were interpreted as *type of tasks* in ATD sense. The teacher's utterance and the teaching materials were mainly analyzed to figure out the other components of the taught praxeology. On the other hand, worksheets of students and their audio recordings were analyzed to find out the learnt praxeology. In particular, students' answers on the following task were analyzed to describe technique, technology, and theory parts of the learnt praxeology: *Consider three* *confidence intervals with 90% of confidence level. Size of each sample is 4, 25, and 100. Array these intervals according to their width. Explain your answer using sampling distributions.*

Type of tasks	T : To find the relations between the width of confidence interval and sample size
Technique	$\tau_{algebra}$: To apply $l = 2z_{\alpha/2}\frac{\delta}{\sqrt{n}}$ (Lee et al., 2011, p. 181) τ_{graph}: To compare the area under graphs of sampling distribution function within $[\mu - d, \mu + d]$ for different sample sizes
Technology	$\theta_{algebra}$ (for $\tau_{algebra}$): Proportional reasoning (Post et al., 1988) $\theta_{graph,1}$ (for τ_{graph}): Reasoning about the impact of sample size on graphs of sampling distribution (Chance et al., 2004) $\theta_{graph,2}$ (for τ_{graph}): Graphical interpretation of probability statements about sample means (Chance et al., 2004) $\theta_{graph,3}$ (for τ_{graph}): Explanation about relations between $[\mu - d, \mu + d]$ and confidence interval (Behar & Yáñez, 2009)
Theory	Θ: The meaning of confidence interval and confidence level (VanderPlas, 2014); the meaning of $P(\bar{X} - d \leq \mu \leq \bar{X} + \mathrm{d}) = 1 - \alpha$ and $P(\mu - d \leq \bar{X} \leq \mu + d) = 1 - \alpha$ (Behar & Yáñez, 2009)

Table 1: REM on confidence interval

Result

What is taught?

Two taught praxeology $\wp_1$, $\wp_2$ were revealed in the lesson. The components of $\wp_1$ are summarized in Table 2.

Type of tasks	Empty
Technique	Empty
Technology	Graphical interpretation of probability statements about sample means ($\theta_{graph,2}$) Explanation about relations between $[\mu - d, \mu + d]$ and confidence interval ($\theta_{graph,3}$)
Theory	The meaning of confidence interval with confidence level; The meaning of $P(\bar{X} - d \leq \mu \leq \bar{X} + \mathrm{d}) = 1 - \alpha$ and $P(\mu - d \leq \bar{X} \leq \mu + d) = 1 - \alpha$ (Θ)

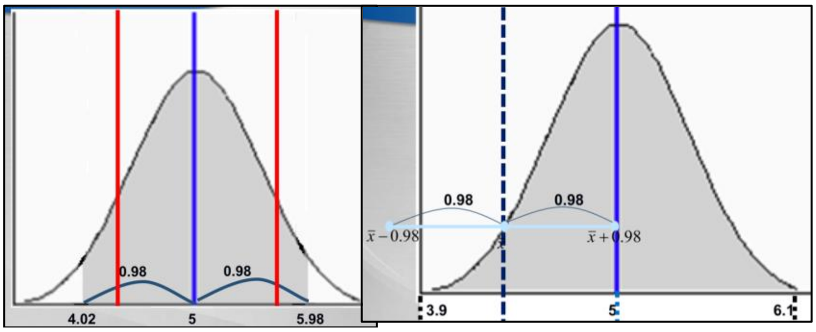
Table 2: Components of the taught praxeology $\wp_1$

$\wp_1$ was a local know-how which consisted of only technology and theory parts. The theory part of $\wp_1$ included the meaning of probability statements of sample means, $P(\bar{X} - d \leq \mu \leq \bar{X} + \mathrm{d}) = 0.95$ and $P(\mu - d \leq \bar{X} \leq \mu + \mathrm{d}) = 0.95$. The meaning of each statement together with confidence interval and confidence level was explained by the teacher in the sense of Frequentism.

Teacher: As you can see, among 100 confidence intervals, only 5 do not contain and the other 95 contain (the population mean). … It means that the level of confidence is 95%.

Grounded on the meaning of probability, the relations between $[\mu - d, \mu + d]$ and confidence interval $[\bar{X} - d, \bar{X} + \mathrm{d}]$ were explained. Graphic representation of the probability statements $P(\bar{X} - d \leq \mu \leq \bar{X} + \mathrm{d}) = 0.95$ and $P(\mu - d \leq \bar{X} \leq \mu + \mathrm{d}) = 0.95$ was used with the explanation.

Teacher: Here's the summary. The probability of $\bar{X}$ being contained between $\mu - d$ and $\mu + \mathrm{d}$ is equivalent to the probability of population mean contained in this one ($[\bar{X} - d, \bar{X} + \mathrm{d}]$). [PowerPoint slides that teacher used with this explanation. Numbers in slides were chosen from an exercise without contextual meanings.]



Although the technology and theory part of $\wp_1$ ground the robust foundation for its potential practice, this praxeology was 'in a vacuum' in that "what kind of problems it can help to solve" is not known (Barbé et al., 2005, p. 237).

Table 3 describes components of another taught praxeology $\wp_2$.

Type of tasks	To compare the probability of sample mean being in a given range when graphs of sampling distribution with different (hidden) sample sizes are given
Technique	To compare the shape of graphs and area under graphs of sampling distribution within a given range (τ'_{graph})
Technology	Reasoning about the impact of sample size on graphs of sampling distribution ($\theta_{graph,1}$) Graphical interpretation of probability statements about sample means ($\theta_{graph,2}$)
Theory	Empty

Table 3: Components of the taught praxeology $\wp_2$

$\wp_2$ was organized around a type of tasks, to compare the probability of sample mean being in a given range when graphs of sampling distribution with hidden sample size (4, 25, 100) are given. The technique τ'_{graph} of $\wp_2$ consisted of two subsequent sub-techniques. The first sub-technique was to compare the shape of sampling distribution function to figure out hidden sample sizes. This sub-technique was justified by the technology $\theta_{graph,1}$, which was the discourse about the impact of sample size on graphs of sampling distribution. The following sub-technique was to compare the area under the graphs of sampling distribution. This sub-technique required interpreting areas under the sampling distribution curve as probability statements about sample means ($\theta_{graph,2}$).

One may note that the technique τ'_{graph} is slightly different from τ_{graph} in that τ'_{graph} did not require the explanation about confidence interval. Hence, the technology and theory related to the concept of confidence interval were missing in $\wp_2$. The theory about probability sentences was also void in $\wp_2$, which made the theory part of $\wp_2$ empty.

What is learnt?

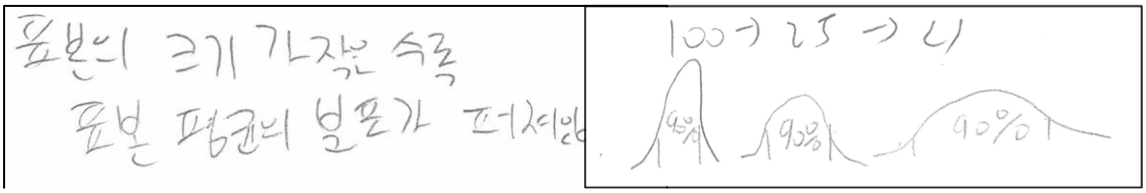
The praxeology of the taught and learnt praxeology are summarized in Table 4.

	Taught praxeology ($\wp_1$, $\wp_2$)	Learnt praxeology
Type of tasks	- To compare the probability of sample mean being in a given range ($\wp_2$)	- To find the relations between the width of confidence interval and sample size
Technique	- To compare the shape of graphs and area under graphs of sampling distribution within a given range ($\tau'_{graph}, \wp_2$)	- To apply $l = 2z{\alpha}/_{2}\frac{\delta}{\sqrt{n}}$ ($\tau_{algebra}$, 8%) - To compare the area under graphs of sampling distribution within $[\mu - d, \mu + d]$ for different sample sizes (τ_{graph}, 85%)
Technology	- Reasoning about the impact of sample size on graphs of sampling distribution ($\theta_{graph,1}, \wp_2$) - Graphical interpretation of probability statements about sample means ($\theta_{graph,2}$, $\wp_1$, $\wp_2$) - Explanation about relations between $[\mu - d, \mu + d]$ and confidence interval ($\theta_{graph,3}, \wp_1$)	- Proportional Reasoning ($\theta_{algebra}$, 8%) - Reasoning about the impact of sample size on graphs of sampling distribution ($\theta_{graph,1}$, 85%) - Graphical interpretation of probability statements about sample means ($\theta_{graph,2}$, 48%)
Theory	- The meaning of confidence interval with confidence level; The meaning of $P(\bar{X} - d \le \mu \le \bar{X} + \text{d}) = 1 - \alpha$ and $P(\mu - d \le \bar{X} \le \mu + d) = 1 - \alpha$ ($\Theta, \wp_1$)	Empty

Table 4: Components of the taught and learnt praxeology

Although the taught praxeology $\wp_1$ and $\wp_2$ did not contain $\tau_{algebra}$ as a technique, some students (8%) used $\tau_{algebra}$ to explain the relationship between the width of confidence interval and sample size. For example, a student explained that his technique $\tau_{algebra}$ was justified "since the width of confidence interval equals to $2 \times a \times \frac{\delta}{\sqrt{n}}$, the width becomes smaller as n increases". This kind of justification corresponds to proportional reasoning $\theta_{algebra}$.

Excluding non-replied answers, remaining 30 students (85%) applied τ_{graph}. However, their technology differed in degree (Figure 2). 13 students (37%) only relied on the reasoning about the impact of sample size on the shape of sampling distribution ($\theta_{graph,1}$). The other 17 students (48%) could integrate $\theta_{graph,1}$ with $\theta_{graph,2}$ when they justified their technique. These two technological elements were also found in the taught praxeology $\wp_2$. However, $\theta_{graph,3}$, which was another technological element of the taught praxeology $\wp_1$, was missing in the learnt praxeology even though $\theta_{graph,3}$ was essential to fully justify technique τ_{graph} of the learnt praxeology.



[Sampling distribution of the sample means is more spread out with the smaller sample size (left)]

Figure 2: Examples of τ_{graph} with $\theta_{graph,1}$(left) and τ_{graph} with $\theta_{graph,1}$ and $\theta_{graph,2}$ (right)

The meaning of confidence interval and confidence level (in the sense of Frequentism) was implicit in students' worksheets or in their dialogues. Moreover, the expression $P(\bar{X} - d \leq \mu \leq \bar{X} + \mathrm{d}) = 1 - \alpha$ or $P(\mu - d \leq \bar{X} \leq \mu + d) = 1 - \alpha$ was hardly dealt with in students works, which leads to the theory part of learnt praxeology lacking the meaning of those expressions.

Discussion

In this study, we analyzed what is taught and learnt in a didactic setting which encouraged reasoning about sampling distribution. Some notable similarities and differences between them were observed.

Firstly, technology $\theta_{graph,1}$ and $\theta_{graph,2}$ were both taught and learnt in the classroom setting. Agreeing with Chance et al. (2004, p. 300) that $\theta_{graph,1}$ and $\theta_{graph,2}$ are kinds of reasoning "students should be able to do with their knowledge of sampling distributions", we argue that essential reasoning about sampling distribution could be integrated in the teaching of relationships between confidence interval and sample size. Furthermore, this kind of reasoning would contribute to the understanding of the relationship between confidence interval and sample size since the notion of sampling distribution and its representation are closely related with each of those concepts.

Secondly, $\theta_{graph,3}$ as well as theory underneath it were found in the taught praxeology but were entirely implicit in the learnt praxeology. These technology and theory were only contained in the local know-how $\wp_1$. Similar phenomenon was also appeared at Barbé et al. (2005) in the case of teaching of limits of functions. We infer that the reason of the incongruity between the taught and learnt praxeology was partly because the related practice of the taught praxeology $\wp_1$ was unknown. Hence, we expect that another taught praxeology with a concrete task whose technique is justified by $\theta_{graph,3}$ and its theory would derive more strict discourse in learnt praxeology.

Finally, $\tau_{algebra}$ with $\theta_{algebra}$ was not taught but learnt in this lesson. This result shows the potential that diverse techniques, including $\tau_{algebra}$ and τ_{graph}, could be employed in the activity of searching relations between confidence interval and sample size. Discussing different techniques and their technology, different local praxeology of students could be connected and integrated into regional praxeology (Bosch & Gascón, 2014).

The research was funded by the Korean Ministry of Education and the National Research Foundation of Korea (NRF-2016S1A5A2A01024273)

References

Barbé, J., Bosch, M., Espinoza, L., & Gascón, J. (2005). Didactic Restrictions on the Teacher's Practice: The Case of Limits of Functions in Spanish High Schools. *Educational Studies in Mathematics*, *59*(1), 235–268.

Behar, R., & Yáñez, G. (2009). Experts and students' conceptions regarding confidence intervals. *Heurística*, *16*, 5–12.

Bosch, M., & Gascón, J. (2014). Introduction to the Anthropological Theory of the Didactic (ATD). In A. Bikner-Ahshags & S. Predger (Eds.), *Networking of theories as a research practice in mathematics education* (pp. 67–83). Springer.

Chance, B., del Mas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 295–323). Springer.

Chevallard, Y., & Bosch, M. (2014). Didactic transposition in mathematics education. In S. Lerman (Eds.), *Encyclopedia of mathematics education* (pp. 170–174). Springer.

Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics: Understanding why and how.* Springer Science & Business Media.

delMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 79–95). Springer.

Fidler, F., & Cumming, G. (2005). Teaching confidence intervals: Problems and potential solutions. *Proceedings of the 55th International Statistics Institute Session.*

Hagtvedt, R., Jones, G. T., & Jones, K. (2008). Teaching confidence intervals using simulation. *Teaching Statistics*, *30*(2), 53–56.

Jaworski, B. (2004). Insiders and outsiders in mathematics teaching development: The design and study of classroom Activity. *Research in Mathematics Education*, *6*(1), 3–22.

Lee, K. S., et al. (2011). *Hwag-lyul-gwa tong-gye [Probability and statistics]*. Seoul: MiraeN

Pfannkuch, M., Wild, C. J., & Parsonage, R. (2012). A conceptual pathway to confidence intervals. *ZDM*, *44*(7), 899–911.

Post, T., Behr, M., & Lesh, R. (1988). Proportionality and the development of prealgebra understandings. *The Ideas of Algebra, K-12, 1988*, 78–90.

VanderPlas, J. (2014). Frequentism and bayesianism: a python-driven primer. *ArXiv Preprint ArXiv:1411.5018.*

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage.