

Research Article

Enhancing pre-service mathematics teachers understanding of sampling distributions with conceptual change texts

Zeynep Medine Özmen¹ and Bülent Güven²

¹Trabzon University, Fatih Faculty of Education, Turkey (ORCID: 0000-0003-0232-9339)

²Trabzon University, Fatih Faculty of Education, Turkey (ORCID: 0000-0001-8767-6051)

The present study aimed to remediate pre-service teachers' misconceptions about sampling distributions and to develop their conceptual understanding through the use of conceptual change texts (CCTs). The participants consisted of 84 pre-service teachers. To determine the pre-service teachers' conceptual understanding of sampling distributions, an achievement test was utilized. Five conceptual change texts were prepared. In this study, the number of correct responses of pre-service teachers increased from pre-test to post-test and delayed test. This increase was statistically significant in favor of the post-test and delayed test. The results demonstrated that, due to the knowledge gained from the CCTs, the pre-service teachers improved their conceptual understanding about sampling distributions. Moreover, this study represents an important effort to integrate CCTs in mathematics and statistics education.

Keywords: Conceptual understanding; Sampling distributions; Conceptual change text; Pre-service teachers

Article History: Submitted 28 June 2021; Revised 5 February 2022; Published online 2 March 2022

1. Introduction

Learning has often been described as an active process in which students play a positive role, building on their previous knowledge, ideas and experiences to develop new knowledge (Driver, 1981, as cited in Çalık et al., 2007; Günay, 2005). According to this constructivist point of view, traditional teaching approaches are considered as inadequate for supporting conceptual knowledge, or the understanding of key concepts (Moseley & Brenner, 1997; Önder & Geban, 2006; Posner et al., 1982; Tsai, 1999). As such, there is a need for discovering alternative instructional techniques that can make learning more meaningful and permanent in diverse subject areas (Önder & Geban, 2006; Özmen & Naseriazar, 2018). This need applies to mathematics and statistics education, in particular, where students taught according to traditional methods often fail to develop conceptual-level understanding in a range of subject areas.

One such topic – sampling distributions – is viewed as a building block of statistics; yet this is one of the most difficult and abstract topics in terms of conceptual understanding (Özmen &

Address of Corresponding Author

Zeynep Medine Özmen, PhD, Trabzon University, Fatih Faculty of Education, Mathematics and Science Education Department, 61335, Trabzon, Turkey.

✉ zmozmen@trabzon.edu.tr

How to cite: Özmen, Z. M. & Güven, B. (2022). Enhancing pre-service mathematics teachers understanding of sampling distributions with conceptual change texts. *Journal of Pedagogical Research*, 6(1), 110-130. <https://dx.doi.org/10.33902/JPR.2022175561>

Guven, 2019; Vanhoof et al., 2007). A large body of research has been reported on this issue, pointing out that students often have difficulty comprehending sampling distributions on a conceptual level (Chance et al., 2004; Ozmen & Guven, 2019; Watkins et al., 2014). To address this problem, a number of researchers have focused on not only determining the misconceptions or challenges students encounter, but also on designing various learning environments to deal with these challenges. In this respect, attention has been drawn to the use of concrete examples or simulation activities (Ozmen & Guven, 2019; Vanhoof et al., 2007). However, even with such techniques, due to the abstract nature of the subject, students continue to exhibit difficulties (Chance et al., 2004; Watkins et al., 2014).

A key aspect of addressing misconceptions is the need to bring about a conceptual change with regard to students' misconceptions that arise during or as a consequence of instruction. In this respect, studies have suggested the use of conceptual change texts (CCTs) (Akpınar & Tan, 2011; Beerenwinkel et al., 2011; Çalık et al., 2007). Also, diSessa (2006) points out that "conceptual change" embodies a first approximation of the primary difficulty: students must build new ideas in the context of old ones, hence the emphasis on "change" rather than on simple accumulation or tabula rasa, or "blank slate" acquisition" (p. 265). Posner et al. (1982), in particular, stress that learners may be resistant to adjusting their misconceptions unless they encounter cognitive conflict. This kind of conflicts may be dealt with through the use of conceptual change texts (Castro, 1998). However, while studies have been carried out with respect to diverse learning paths such as the use of concrete examples and simulation activities, the impact of CCTs on conceptual understanding has not been directly examined in terms of sampling distributions.

1.2. Aim of the Study

This study aimed to remediate pre-service mathematics teachers' misconceptions about sampling distributions and to develop their conceptual understanding through the use of conceptual change texts (CCTs). Accordingly, the study addresses the following question:

What kind of contributions do CCTs have on the conceptual understanding of pre-service mathematics teachers about sampling distributions?

2. Theoretical Framework

2.1. Sampling Distribution

A sampling distribution has been defined as "the distribution of all values of statistics when all possible samples of a sample size n are taken from the same population" (Triola, 2010, p. 276). In other words, sampling distributions can be generated by determining all of the possible samples of size n . Afterward, related statistics (e.g. mean, standard deviation, or proportion) can be calculated and a new distribution built on the basis of these statistics. Moreover, the sample means distribution is generated by taking all n -sized samples from the population, calculating the means of all samples, and building a new distribution according to the values of the sample means.

Due to its abstract and difficult nature, the topic of sampling distribution is one that is particularly challenging for students (Chance et al., 2004; Ozmen & Guven, 2019). Also, it is stressed that the sampling distribution is fundamental to the understanding of statistical inference (Hancock & Rummerfield, 2020). Understanding sampling distributions requires students to integrate their knowledge about sample statistics, variability, and distributions to perform statistical inference on population parameters (Chance et al., 2004; Saldanha & Thompson 2007). However, given its role as a transition from descriptive statistics to inferential statistics, it is important that students develop an adequate conceptual understanding of the topic. Chance et al. (2004) underlined some of the common misconceptions about sampling distributions, which have been reported in other studies as well. In particular, misconceptions such as "sampling distribution should look like the population," "if the sample size increases, distribution of the samples is more likely to follow a normal distribution," and " $n > 30$ is a sufficient sample size to assume any sample means distributions would follow a normal distribution" are often cited (Chance et al., 2004; Ozmen & Guven,

2019). Hancock and Rummerfield (2020) stated that “there is still no consensus, however, on how best to correct these misconceptions” (p. 9). They also emphasized the importance of determining the most effective methods to clearly overcome misunderstandings about sampling distributions in an introductory course. Given the importance of the topic of sampling distributions and the difficulties students encounter in understanding it, researchers have worked to overcome the challenges by designing learning environments based on simulation or computer activities (delMas et al., 1999; Lane, 2015; Watkins et al., 2014; Zieffler et al., 2007). However, despite their efforts, these studies demonstrate that students continue to have difficulty understanding sampling distributions (Chance et al., 2004; Watkins et al., 2014). Therefore, there is an ongoing need to overcome this problem; and in this respect, conceptual change texts have been viewed as an effective tool (Köse et al., 2011; Södervik et al., 2015), as they include refutation and scientific explanations designed to overcome specific misconceptions.

2.2. Conceptual Change Texts (CCTs)

Conceptual change emerges through a process that provides conceptual understanding. Beerenwinkel et al. (2011) defined conceptual change as a learning path related to key concepts or alternative ideas developed through instruction on scientific concepts. Çalık et al. (2007) underscored that, during or as an outcome of the learning process, students may adopt alternative conceptions (i.e., misconceptions). It has also been suggested that alternative ideas may be developed by learners in the process of a teaching experiment (Çalık et al., 2007). Furthermore, Posner et al. (1982) point out that if learners’ pre-existing knowledge contradicts a new concept, the conceptual change process may become problematic. Thus, CCTs or refutation texts may be employed to overcome these problems; in instances where learners are unwilling to replace their alternative conceptions with the correct knowledge, CCTs may be an effective tool to achieve the necessary change (Posner et al., 1982).

In order to bring about meaningful understanding, CCTs must first acknowledge students’ misconceptions on a given topic, then directly refute them and introduce the appropriate scientific concepts (Hynd et al., 1991; Sinatra & Broughton, 2011). In this way, CCTs help learners to replace their misconceptions with the correct knowledge (Akpınar & Tan, 2011). In terms of their structure, CCTs may differ. However, in general, they consist of three main components: presentation of the misconception, refutation of the misconception, and remediation the misconception with the help of the currently accepted scientific explanation (Tippett, 2010).

Because science education involves abstract concepts, necessitating the development of conceptual understanding, CCTs are frequently employed. As such, a number of studies have focused on CCTs as applied in the context of science education (Çalık et al., 2007; Özmen & Naseriazar, 2018) to examine the effectiveness of conceptual or refutational texts in bringing about conceptual change. On the other hand, studies related to the use of CCTs in mathematics education are limited.

Moreover, according to research, students continue to encounter challenges, difficulties and misunderstandings about sampling distributions despite the use of simulation activities or concrete examples (Chance et al., 2004; delMas et al., 1999; Garfield & Ben-Zvi, 2008). Thus, there is an emerging need to overcome these challenges and remediate the resulting misconceptions. To address the gap in the literature, this study was designed to explore the impact of CCTs in mathematics education.

In the present study, the terms *population distribution*, *sample*, *sampling distributions*, *parameters* and *statistics* are frequently used. Brief descriptions of these terms are provided here.

- Population distribution: a distribution that has all of the elements of a research proposal.
- Parameter: a value that describes something about a population. In the present study, population mean, and standard deviation are used as parameters and references as μ and σ respectively.

- Sample distribution: a distribution that is generated by selecting n element from a population.
- Statistics: values that describe something about a sample. The mean and standard deviation of a sample are described with statistics and are referenced as \bar{X} and s respectively.

In more detail, a sampling distribution is a distribution of all values of statistics when all possible samples of a sample size n are taken from the same population (Triola, 2010, p. 276). Taking all possible samples means selecting all different n sized samples from a population. Sampling distribution of the mean is distribution of sample means with all samples having the same sample size (n) taken from the same population. Present study is limited with sampling distributions of mean (sample means distributions) and it is used as “sampling distribution” through the text.

3. Method

The aim of this study was to enhance pre-service mathematics teachers’ conceptual understanding of sampling distributions through the use of CCTs. With this in mind, a learning environment was designed in the context of a case study.

3.1. Participants

The participants in the study consisted of 84 pre-service mathematics teachers (18 male, 66 female) in their 3rd year of an elementary mathematics teacher education program at a state university in Turkey. This group was selected because sampling distribution is taught at this level in the context of a statistics course. Namely, in the Elementary Mathematics Teaching program, pre-service teachers (PTs) are required to take a one-semester statistics course during which they learn topics including organizing and classifying data; histograms; central tendency and dispersion measurements within classified data; normal and standard normal distribution; sampling; sampling distributions; confidence intervals; hypothesis testing; correlation and regression analysis. Accordingly, the statistic course ranged in scope from descriptive statistics to inferential statistics. In this context, sampling and sampling distributions play a significant role in the PTs conceptual understanding of inferential statistics.

The participants learned concepts such as organizing and classifying data; histograms; central tendency and dispersion measurements; normal and standard normal distribution; sampling; and sample means distributions. As such, subjects that involved sampling distributions were presented in detail. Prior to the experiment, the instructor discussed sampling and its importance in conducting research. After the detailed explanations of sampling, the lesson began with an example constituting samples ($n = 2$ sample-sized) from a population ($N = 5$). The PTs were asked to select all possible $n = 2$ sized samples from the population and to calculate the mean and standard deviation of the sample mean distributions. Afterward, they were asked to relate these statistics to the population parameters. This approach aimed to provide a basic demonstration of constituting sample means distributions and comparing them with population and sample distributions. The central limit theorem was taught according to this process, and the participants solved a series of practice problems. After sampling distribution is taught at a basic level, PTs are asked to answer pre-test questions. After this period inferential statistics subjects are taught as starting with the confidence intervals. Thus, there is any intervention about sampling distributions between pre-and post-test. Following the post-test period, delayed test is implied, and any intervention is applied regarding sampling distribution expect for the requirements of the lesson. A better understanding of sample means distributions leads to a better understanding of inferential statistics (e.g., confidence intervals, hypothesis testing, and so on). However, due to the difficult and abstract nature of sample mean distributions, students often have difficulties with conceptual understanding of this topic, leading to problems with comprehension of inferential statistics. With this in mind, we aimed to provide conceptual understanding about sampling distributions, rather than limiting the instruction to the procedural steps involved.

3.2. Research Design and Instruments

For the purposes of this study, the data were collected through an achievement test about sample means distributions. In addition, CCTs related to sample means distributions were used to enhance the PTs' conceptual understanding. The research process involved three main stages. In the first stage, the PTs completed a 30-minute pre-test. In the second stage, which took place in the following week, they read the CCTs. Two days later, they were asked to complete a 30-minute post-test. Finally, one week later, PTs completed a delayed test to measure the permanence of the sample means distributions knowledge acquired with the support of the CCTs (Özmen & Naseriazar, 2018; Södervik et al., 2015).

3.2.1. Sampling distribution achievement test

In order to determine PTs' conceptual understanding of sampling distributions, as well as the impact of the CCTs on their understanding, a sample means distributions achievement test was utilized. The test was developed by the researchers in view of the common misconceptions expressed in the existing literature (Ozmen & Guven, 2019; Chance et al., 2004). These misconceptions were considered not only in developing the test questions, but also in writing the CCTs. The test was designed in line with the theoretical framework and questions of Chance et al. (2004). This framework, also used in the study of Ozmen and Guven (2019), underscores the knowledge components required for understanding sampling distributions, including: (i) *knowing the meaning of "sampling distributions;"* (ii) *understanding the relationship between parameters and statistics;* (iii) *understanding the effect of sample size on the shape of distributions and statistics;* (iv) *understanding the relationship between the distributions of populations, samples and sample means;* and (v) *applying the theoretical knowledge of sampling distributions to probability problems about sample means* (p. 28). Because this study focused on conceptual understanding, concepts that did not apply to the theoretical knowledge of sampling distributions were not considered; accordingly, only three of the knowledge components (ii, iii, iv) were applied. The component of "knowing the meaning of sampling distributions" is mainly related to the process of generating sample means distributions based on a population; therefore, no items related to this concept were included on the test. Additionally, in developing the items for the achievement test, the study of Ozmen and Guven (2019) was also examined. The final version of the test comprised 16 true-false questions (see Appendix). In responding to the questions, the PTs were instructed to follow two steps: (i) determining whether the statement was true or false; and (ii) giving a suitable explanation for their answers. Sample questions are provided in Table 1 below in line with the related aspects.

Table 1

Sample Questions and Related Aspect of the Sample Means Distribution Achievement Test

Aspect	Questions	Sample Question
1. Relationship between parameters and statistics	Q3, Q9, Q11, Q14	Q11: The mean of any sample taken from a population, having the average μ , is equal to μ .
2. Effect of the sample size on the shape of the distribution and statistics	Q1, Q5, Q7, Q8, Q10, Q13, Q15, Q16	Q5: As the sample size (n) increases, the distribution of the sample becomes more like the distribution of the population.
3. Relationship between the distributions of populations, samples and sample means	Q2, Q4, Q6, Q12	Q6: Regardless of what the population distribution looks like, the sampling distributions of mean with all possible samples follows a normal distribution.

As Table 1 demonstrates, the test included four questions related to the aspect of *Relationship between parameters and statistics*, eight questions related to the aspect of *Effect of the sample size on the*

shape of the distribution and statistics, and four questions related to the aspect of *Effect of the sample size on the shape of the distribution and statistics*. More detailed information related test questions, correct answers and explanations are also given at Appendix.

3.2.2. Conceptual change texts

In order to promote the PTs' conceptual understanding of sample means distributions, five CCTs were prepared. These texts, which aimed to highlight facts about sample means distributions, were designed on the basis of two criteria outlined in the literature (Guzzetti, 2000; Tippett, 2010): (1) targeting a common misconception that creates challenges to understanding sample means distributions; and (2) refuting the current misconception by referring to appropriate and accepted scientific explanation via concrete elements such as visual illustrations or counterexamples. The following challenges, as reported in the literature, were considered in preparing the CCTs:

Whether;

- $n > 30$ is an important factor for any distribution to appear normal;
- the distributions of both population and the samples taken from this population are obligated to behave parallel to one another;
- the concepts of sample and sample means correspond to the same meaning;
- parameters and statistics must be equal.

The content and the challenges presented in the CCTs are outlined in Table 2.

Table 2

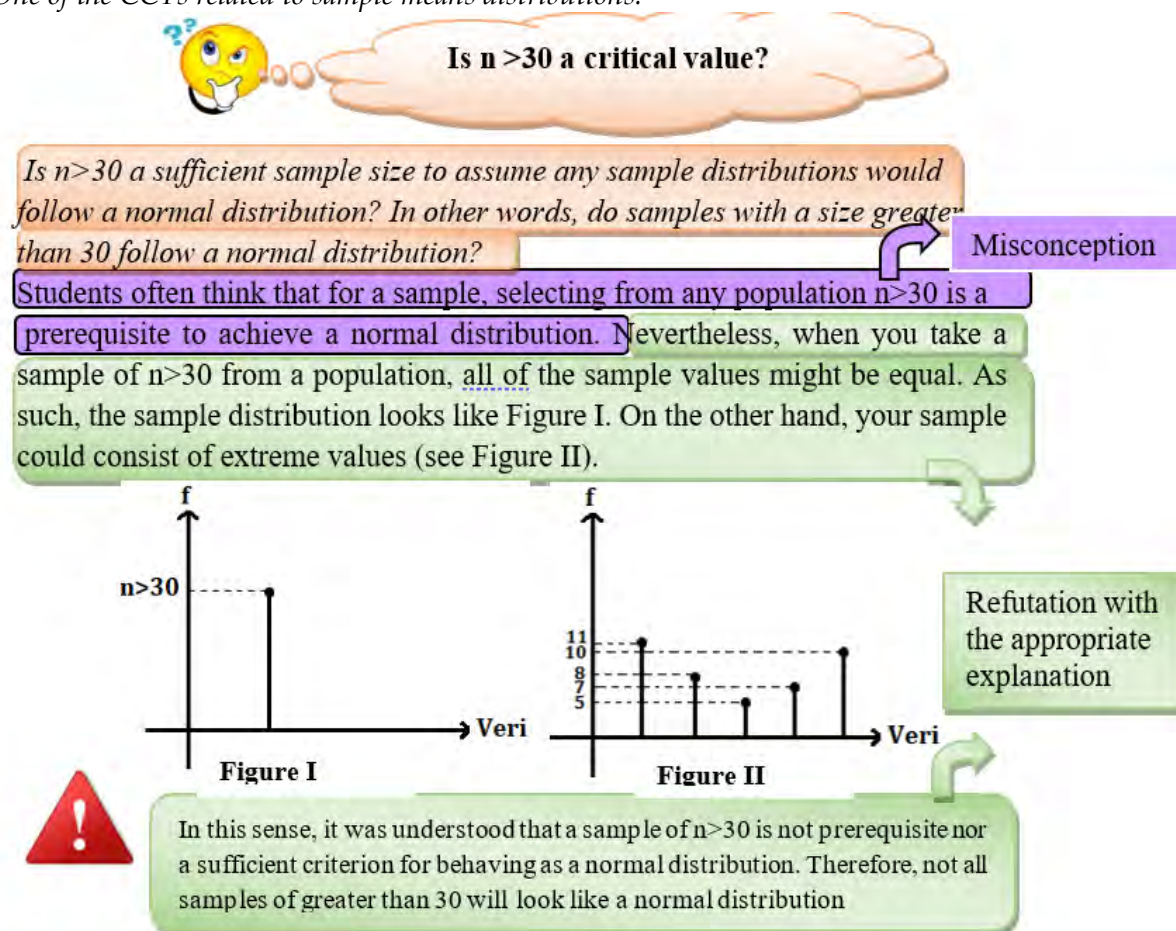
The Content and the Challenges Presented in the CCTs

CCTs	Content	Challenges
Is $n > 30$ a critical value?	The effect of a sample size of $n > 30$ on the distribution of the sample distribution	Samples of 30 or more follow a normal distribution.
Does the pear fall into its tree?	The relationship between distributions of populations and samples.	Populations and sample distributions must exhibit parallel behavior.
Are sampling and sample means the same?	The meaning of sample and sample means distributions	Sample and sample means distributions have the same meaning.
Must parameters and statistics be equal?	The relationship between population parameters and the statistics of samples and sample means.	Population parameters and sample statistics must be equal. The statistics of the samples and sample means can never exceed the population parameters, or they must be smaller than the population parameters. Sample means distributions have higher variability than population.
Where are the distributions going?	As the sample size increases, what does the shape of the sample and sample means distribution look like?	As the sample size increases, sample distributions move closer to a normal distribution. As the sample size increases, sample means distributions look more like the population.

The following is one of the CCTs prepared for the study:

Figure 1

One of the CCTs related to sample means distributions.



This CCT was grounded in the misconception in the existing literature that samples of $n > 30$ appear normal (Chance et al., 2004; Ozmen & Guven, 2019). In the first section of the CCT, PTs' attention was drawn to the related misconception. Then, two different examples of sample distributions with a sample size of greater than 30 were given in order to demonstrate that they could have a uniform or a skewed distribution. These examples served as a refutation of the misconception.

3.3. Data analysis

The data were analyzed through both qualitative and quantitative procedures. Before the data analysis, PTs who had completed all of the tests, including the pre-test, post-test, and delayed test, were identified. Any who had not completed all three of the tests were disqualified from the sample. The data induction process revealed that a total of 84 PTs had completed all three of the tests. In order to maintain their anonymity in analyzing and reporting the data, the participants were assigned codes PT1 (Pre-service Teacher 1) through PT84. Their responses to the test items were evaluated based on the categorical scoring table; all test items were scored as 2, 1, or 0 points. The scores were determined independently by two researchers. The coding procedure is outlined in Table 3.

Table 3
Coding Procedure for the Sampling Distribution Test

Questions	Scores		
	2	1	0
Q1-16	Choosing the correct option (True or false) and explaining the reason precisely.	Choosing the correct option without any explanation or with an inappropriate explanation.	Choosing the incorrect option, or no answer given.

Since the achievement test included 16 items, the maximum score a participant could achieve was 32. All of the responses to each of the questions were analyzed according to this scoring table, and descriptive statistics were calculated for all the three tests. In order to determine the development of the conceptual understanding of the PTs about sample means distributions, as well as the effectiveness of the CCTs on the PTs' achievement and understanding of sample means distributions, statistical analysis was performed. After the normality of the groups for the three tests was examined, ANOVA was performed in order to determine whether there was a statistically significant difference between the pre-test, post-test, and delayed test. In addition, the PTs' answers for all three tests were analyzed qualitatively to illustrate the development of their understanding or their persistence in their misconceptions.

4. Findings

After the answers of the PTs to all questions on each test (pre-test, post-test, and delayed test) were scored, PTs' level of success was determined based on frequencies. The distribution of the frequencies for each question and each test are presented in Table 4.

As the table demonstrates, there was an increase in the scores from the pre-test to the post-test and delayed test. Moreover, a greater number of the PTs' responses were evaluated with 0 points on the pre-test than on the post-test and delayed test. In addition, there was a notable difference in the number of responses evaluated with 2 points for items Q1, Q2, Q4, Q5 and Q11. However, there was no difference between the pre-test, post-test, and delayed test for the responses evaluated with 2 points on item Q12. On the other hand, while more of the post-test responses were evaluated with 2 points for items Q7, Q8, Q10, Q13 and Q16 than on the pre-test, a similar improvement was not seen on the delayed test. In addition, the percentage of 2-point responses to items (Q3, Q6, Q9, Q14 and Q15) were similar on all the tests, while the percentage of 0-point responses decreased from the pre-test to the post-test and delayed test for nearly all the items. This indicates that CCTs generally had a positive effect on the PTs' responses, as they were evaluated with 1 or 2 points. It was seen that PTs' justification and explanation success increased from pre-test to post and delayed test. In other words, the percentage of the answers getting 0 points were decreased. On the other hand, explanation success was 11.01% for pre-test, 29.46% for post-test, 21.73% for delayed test. Despite justification success was not dramatically changed. In other words, it can be said that 0-point answers changed with 2-point answers from pre-test to post and delayed test. After the distribution of the frequencies of the PTs' responses were determined, the average scores of the individual test items and the full tests were calculated and presented in Table 5.

Table 4
The Distribution of Frequencies for Each Question and Each Test

	Pre-test						Post-test						Delayed test					
	1		0		2		1		0		2		1		0			
	f	%	f	%	f	%	f	%	f	%	f	%	f	%	f	%		
Q1	7	8.33	3	3.57	74	88.09	38	45.24	13	15.48	33	39.29	27	32.14	23	27.38	34	40.48
Q2	4	4.76	29	34.52	51	60.71	41	48.81	38	45.24	5	5.95	38	45.24	43	51.19	3	3.57
Q3	8	9.52	36	42.86	40	47.62	15	17.86	47	55.95	22	26.19	14	16.67	51	60.71	19	22.62
Q4	3	3.57	50	59.52	31	36.90	34	40.48	48	57.14	2	2.38	28	33.33	55	65.48	1	1.19
Q5	20	23.80	47	55.95	17	20.24	51	60.71	22	26.19	11	13.10	36	42.86	32	38.10	16	19.05
Q6	8	9.52	59	70.24	17	20.24	13	15.48	47	55.95	24	28.57	10	11.90	56	66.67	18	21.43
Q7	8	9.52	59	70.24	17	20.24	18	21.43	43	51.19	23	27.38	11	13.10	53	63.10	20	23.81
Q8	5	5.95	29	34.52	50	59.52	19	22.62	48	57.14	17	20.24	12	14.29	59	70.24	13	15.48
Q9	14	16.67	44	52.38	26	30.95	16	19.05	49	58.33	19	22.62	13	15.48	46	54.76	25	29.76
Q10	5	5.95	50	59.52	29	34.52	21	25	57	67.86	6	7.14	11	13.10	67	79.76	6	7.14
Q11	21	25	28	33.33	35	41.67	47	55.95	29	34.52	8	9.52	45	53.57	29	34.52	10	11.90
Q12	1	1.19	39	46.43	44	52.38	3	3.57	28	33.33	53	63.10	1	1.19	35	41.67	48	57.14
Q13	11	13.09	46	54.76	27	32.14	20	23.81	59	70.24	5	5.95	8	9.52	73	86.90	3	3.57
Q14	16	19.05	41	48.81	27	32.14	20	23.81	55	65.48	9	10.71	22	26.19	46	54.76	16	19.05
Q15	8	9.52	6	7.14	70	83.33	13	15.48	6	7.14	65	77.38	9	10.71	7	8.33	68	80.95
Q16	9	10.71	59	70.24	16	19.05	27	32.14	39	46.43	18	21.43	7	8.33	53	63.10	24	28.57
Total	148	11.01	625	46.50	571	42.49	396	29.46	628	46.73	320	23.81	292	21.73	728	54.17	324	24.11

Table 5
Average Raw Scores of the Test Items and Full Tests

Questions	Average Scores		
	Pre-Test	Post-Test	Delayed Test
Q1	0.20	1.06	0.92
Q2	0.44	1.43	1.42
Q3	0.62	0.92	0.94
Q4	0.67	1.38	1.32
Q5	1.04	1.48	1.24
Q6	0.89	0.87	0.90
Q7	0.89	0.94	0.89
Q8	0.46	1.02	0.99
Q9	0.86	0.96	0.86
Q10	0.71	1.18	1.06
Q11	0.83	1.46	1.42
Q12	0.49	0.40	0.44
Q13	0.81	1.18	1.06
Q14	0.87	1.13	1.07
Q15	0.26	0.38	0.30
Q16	0.92	1.11	0.80
Total	10.96	16.90	15.62

As Table 5 demonstrates, the averages on the post-test and delayed test were higher than the average on the pre-test. Moreover, nearly all the individual questions saw similar improvements. While the PTs were more successful in their responses to items Q5 and Q12, they were unsuccessful in answering Q1 and Q15 on the pre-test. In addition, they were more successful in responding to items Q2 and Q5, but unsuccessful in their answers to Q12 and Q15 on the post-test. Likewise, they performed better on items Q2 and Q11, while they were unsuccessful on items Q12 and Q15 on the delayed test.

After the PTs read the CCTs, they were more accurate in their responses in many questions on the post-test and delayed test, except for items Q6, Q7, Q9, Q12 and Q16. Their success rates were similar for all tests on these five questions, in particular; the pre-test scores were either higher than or equal to their scores on the post-test and delayed test for these questions. However, these differences did not have a significant effect. Once the responses of the PTs were coded as 0, 1, or 2, the raw scores were transformed into linear scores via Rasch Analysis. A summary of the statistics of the linear scores of the pre-test, post-test and delayed test are given in Table 6.

Table 6
Summary Statistics of the Linear Scores for the Pre-test, Post-test, and Delayed test

Linear scores	Pre-test	Post-test	Delayed-test
\bar{x}	-0.89	0.14	-0.09
SD	0.56	0.68	0.59

As shown in Table 6, the averages of the linear scores for the post-test and delayed test were higher than the pre-test scores, indicating that reading the CCTs allowed the PTs to begin developing a conceptual understanding of sampling distributions. Furthermore, they demonstrated greater success on the post-tests. As such, it can be inferred that administering the achievement tests after they had read the CCTs had the most significant effect on their understanding and achievement. On the other hand, because the delayed test was administered one week after the post-test and two weeks after they read the CCTs, the pre-service teachers' average scores on the delayed test were lower than on the post-test. However, the average scores

on the delayed test were higher than those of the pre-test. Thus, it can be stated that the CCTs had a positive effect on both the test scores and the PTs' understanding of sampling distributions.

In consideration of all the test scores, PTs were most successful on the post-test. Moreover, while there were differences in the success rates on the pre -test, post-test, and delayed test in favor of the post-test and delayed test, it is important to determine whether those differences were statistically significant. Therefore, ANOVA was carried out for the related measures; the test results are presented in Table 7.

Table 7
ANOVA Test Results for the Pre-test, Post-test, and Delayed test

Sum of the Squares	Mean Square	F*	df	p	Eta Squared Coefficient	Significant Difference**	p	CI
1641.056	882.201	110.953	1.86	.000	.572	Pre-Post	.000	4.811-7.07
						Pre-Delayed	.000	3.607-5.702
						Post-Delayed	.002	0.411-2.17

Note. * Due to the assumption that the sphericity is not seen, the Greenhouse-Geisser correction was considered and Bonferroni post-hoc test is implied. ; ** Bold printed items are the groups which significant difference favors

As Table 7 demonstrates, there was a statistically significant difference among the test scores ($F = 110.953, p < .000, \eta^2 = .572$), indicating that the differences between the pre-test, post-test and delayed test was also statistically significant. In order to determine which groups had a statistically significant difference, post-hoc tests were carried out. The results revealed that the differences between the pre-test and post-test, pre-test and delayed test, and post-test and delayed test were statistically significant.

Given that the PTs scored higher on the post-test and delayed test than the pre-test, it can be claimed that the CCTs helped the PTs to improve their understanding. Namely, the PTs' responses were generally scored as 0 on the pre-test, but this tendency shifted to 1 or 2 points on the post-test and delayed test items. Moreover, before reading the CCTs, the PTs could only respond that a given statement was true or false, without giving appropriate reasons for their responses.

In this sense, the CCTs did not have as great effect on Q12 and Q15 as on the other items. On the other hand, there was a considerably difference on their success rate on items Q1, Q2 and Q8 after reading the CCTs. For instance, Q1 involved the aspect of the *effect of the sample size on the shape of a distribution and statistics*. PTs were asked to imagine that, as the sample size (n) increases, the sampling distribution becomes more like a normal distribution. The related pre-test scores showed that 74 PTs (88.09%) responded that an increase of the sample size causes it to follow the normal distribution of a sample distribution, thus earning 0 points. On the other hand, after the treatment, the PTs realized that as the sample size increases, the sample distribution becomes more like the population, as reflected on the post-test (45.24% received a score of 2 points, and 15.48% received a score of 1 point) and the delayed test (32.14% received 2 points, and 27.38% received 1 point). For example, PT65 responded as follows for item Q1:

(True) Because the average of the sample means distribution and population are equal.

Her response indicates that, although the aim of item Q1 was to get students to determine the effect of a sample size on the distribution of the sample, she confused the two terms and gave an incorrect response by considering the sampling distributions of mean. Thus, it could be said that PT65 did not have adequate knowledge about samples and sampling distributions. On the other hand, she received 2 points for item Q1 with her answer on the post-test, as follows:

(False) As the sample size increases, the sample distribution gets closer to the population distribution.

Her response in this case reveals that she realized that as the sample size increases, the sample distribution gets closer to the population distribution, rather than the normal distribution. As such, she could accurately explain the effect of the sample size on the sample distributions on the post-test. On the other hand, PT15 answered this item as follows:

(True) When the sample size (n) is greater than 30, the sample distribution follows a normal distribution. For example, if the sample size increases, sample distributions such as (she drew a uniform distribution) start to look like a shape such as (she drew a normal distribution).

When her answer was analyzed, it was observed she had the misconception that a sample size of 30 or more causes a sample distribution to look like a normal distribution. In her drawings, she tried to illustrate this idea. However, the drawings indicated the relationship between the shape of a population and the sample means distribution. However, on the delayed test, she responded to the same item (Q1) as follows and received 2-point for her response:

(False) As the sample size increases, the sample distribution comes closer to the population distribution. If the population follows a normal distribution, we can say that the sample distribution becomes more like normal. Otherwise, it would not display a normal distribution.

In this response, she explained that a sample distribution comes closer to the population distribution as the sample size increases. She also thinks that a population must follow a normal distribution in order for the sample distribution to also appear normal as the sample size increases. Otherwise, the sample would not follow a normal distribution as the sample size increases; rather, it would look like the population. As such, it can be claimed that that the PTs had successfully developed a conceptual understanding about the effect of the sample size on sample distributions because of the CCTs, since their responses generally referred to statements given in the texts.

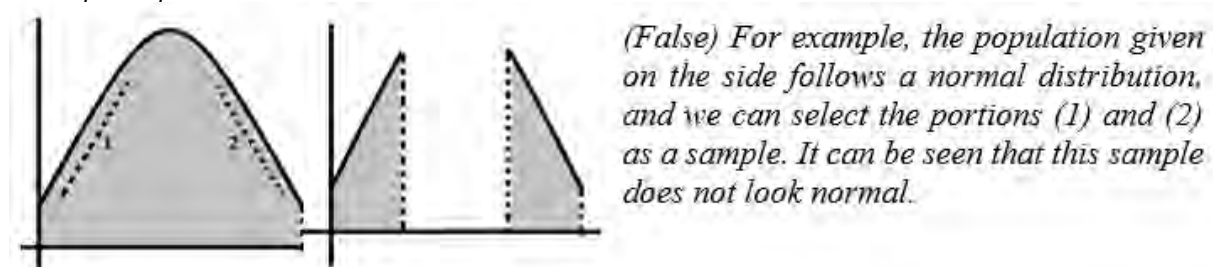
A similar effect could be seen for item Q2, where the pre-test scores were lower, and the post-test and delayed-test scores were considerably higher. Namely, while 60.71% of the PTs scored 0 on the pre-test, this ratio decreased dramatically on the post-test and delayed test, with percentages of 5.95% and 3.57% respectively. Item Q2 asked the PTs to think about whether selecting a sample from a population that looks like normal would follow a normal distribution. To illustrate the responses, PT38's answer for the item Q2 on the pre-test was as follows:

(True) If we select a small part (sample) from the big one (population), this sample shows the characteristics of the population.

This answer revealed that she had the misconception that the distribution of the population and sample must behave parallel, as the samples were taken from the population. However, on the post-test and delayed test, she received a score of 2 points on item Q2. Her response on the post-test was presented in Figure 2.

Figure 2

A sample response on item Q2



Here, she explains why the sample does not have to behave like the population, providing concrete example as given in the CCTs. As such, she understood that the shape of the sample and the population distribution can be different from each other. Thus, it is believed that the CCTs effectively overcame her misconception and developed her conceptual understanding about the relationship between sample and population distributions. Moreover, the student retained this understanding for the delayed test, where she also received 2 points for her explanation of why the statement was false.

With regard to the test scores, items Q11 (*The mean of any sample taken from a population, having the average μ , is equal to μ*), Q5 (*As the sample size (n) increases, the distribution of the sample becomes more like the distribution of the population*), and Q2 had the highest ratio of 2-point responses on the

post-test and delayed test. Moreover, the ratio of the answers scored as 2 points increased for nearly all the test items. This implies that the CCTs effectively supported the PTs not only with respect to determining the accuracy of the statements, but also to providing appropriate explanations. For example, on the pre-test, PT26 answered Q11 (concerning the relationship between population parameters and sample statistics) as follows:

(True) The average of the sample distribution is equal to the average of the normal distribution.

With this answer, she indicated a belief that the average of the sample and normal distribution (she refers to population) must be equal, revealing a misunderstanding about the relationship between the average of a sample and a population. In addition, she generalized the population distribution with the normal distribution. When normal distribution is taught, μ is referred to as the average of the distribution. However, μ is a parameter, rather than an absolute, as she appeared to believe with her response. On the other hand, her answers on the post-test and delayed test were scored as 2 points; her response was as follows:

(False) The average of the sample means distribution (for all possible n-sized samples) is equal to the population average (μ).

This indicates that she realized that the average of any sample taken from the population is not necessarily equivalent to the average of the population. Moreover, her understanding that the average of the sample means distributions, as generated by all n-sized samples taken from the population, is equal to the average of the population (μ), carried through to her response on the delayed test; therefore, her response was scored as 2 points.

As Tables 3 and 4 demonstrate, the PTs had difficulties in answering item Q15. In this regard, although the CCTs effectively enhanced their understanding about sampling distributions in general, they did not provide a similar result for this test item. Instead, the PTs persisted with their misconceptions and continued to give incorrect answers. This item, which concerned the effect of a sample size on the shape of the sample means distributions, asked the PTs to determine that the sample means distribution approaches the population distribution as the sample size increases. While they successfully responded to item Q5, which related to the same concept regarding the behavior of a sample distribution, they were not able to translate this knowledge to item Q15. As such, it is possible that they did consider the sample distribution, but they did not read the question carefully. For example, on item Q15, PT3 marked the statement as true and gave incorrect responses on all three tests, but she answered Q5 correctly on both the post-test and delayed test. Her responses to these items were as follows:

(True) For example, if we think that $n=2$ corresponds to 90%, $n=3$ or $n=4$ becomes 99%. In other words, we begin to consider all the values in a population as the sample size increases (Pre-test-Q15).

(True) As the sample size increases, we can represent the population more accurately, and the sample means distributions come closer to the population distribution (Post-test-Q15).

(True) Because we move from a portion to all of them gradually; therefore, we come closer to the population (Delayed-test-Q15).

(True) The probability is increasing. In other words, as n increases, it represents nearly all possible values, and the distribution comes closer to population (Post-test-Q5).

It is true (Delayed-test-Q5)

When her answers for all tests were examined, it was observed that she confused sample distribution with sample means distribution. It is possible that she still had the misconception that sampling distribution and samples are the same concept. On the other hand, the higher success rate of the PTs on the other items suggests that some of the PTs did not understand the basic idea in this statement.

In the present study, it was observed that reading the CCTs had a considerable effect on the PTs' success in responding to questions about sampling distributions. This impact was primarily seen in the ratio of answers scored with 2 points. Although they had higher scores for some of the

questions on the test (Q5, Q10, Q11, Q13 and Q14), their pre-test scores were similar for these items. For example, while the average of the PTs for Q10 on the pre-test was 0.71, the averages on the post-test and delayed test were 1.18 and 1.06, respectively. The averages of the post-test and delayed test were higher than the pre-test and greater than 1.00. However, the success rate in responding to this item was higher than on the other items on the pre-test. Item Q10 concerned the effect of a sample size on the shape of the distribution and statistics. PTs were expected to express that not all samples greater than 30 taken from a population will follow a normal distribution. In this respect, the success rate with respect to this item might depend not only on the CCTs, but also on the instruction about the sampling distributions that took place in the course.

Considering the frequency and average tables for the test items, the item-person maps, and the ANOVA test results together for all tests, it is clear that the CCTs were an effective tool for enhancing the conceptual understanding and correcting the erroneous ideas of the PTs about sampling distributions. This effect was particularly illustrated in their justification about the accuracy of the statements. On the pre-test, the PTs were only able to determine whether a statement was true (they mainly gave incorrect responses), without providing any explanation, or they supported their ideas with misunderstandings or misconceptions. However, on the post-test and delayed test, they began to provide correct answers, along with appropriate explanations for their responses. In addition, they began to support their ideas with concrete examples, rather than irrelevant statements. In this respect, although they generally chose to use the examples given in the CCTs, this was an important development in terms of being able to provide justifications with the help of appropriate examples. This effect was primarily apparent for items Q2, Q4, Q10, Q11, and Q14. For example, item Q11 concerned the relationship between the population parameters and the statistics. For this item, the PTs were expected to determine whether the average of any sample taken from a population with average μ is equal to μ . As the pre-test scores demonstrate, almost half of them (41.67%) generalized the average of the population and the sample, expressing that the average of all samples taken from a population with an average μ must be equal to μ . However, this tendency decreased on both the post-test (9.52%) and the delayed test (11.90%). Additionally, the percentage of 2-point responses increased from 25% to 55.95% and 53.57% respectively after the PTs read the CCTs. One instance of such a response, PT75's answer to item Q11, is as follows:

If we select individual samples from the population $\{1, 2, 3\}$ and the samples are $\{1\}, \{2\}, \{3\}$, the average of the samples and the population are equal (pre-test).

The Average of the sample means distributions and the samples are different. If a population consists of $\{1, 2, 3, 4\}$, μ is equal to $\frac{5}{2}$. Let's consider that our sample is $\{1, 2\}$. $\bar{X} = \frac{3}{2}$ (post-test).

If our population is $\{1, 2, 3, 4, 5\}$, $\mu = 3$. We can select the sample $\{1, 2\}$, $\bar{X} = \frac{3}{2}$ (delayed test).

As these responses demonstrate, PT75 answered incorrectly on the pre-test. Moreover, she confused the term of the sample and its average on the pre-test. Although she provided three different examples of the samples, she calculated the average of the samples collectively, rather than calculating the averages separately. Thus, her response was incorrect, and she scored 0 points. On the other hand, she responded correctly to this item and scored 2 points on both the post-test and the delayed test. In her responses, she imagined a population and calculated its average, as well as selecting a sample from this population demonstrating that the average of the sample could be different from μ .

Item Q4 concerned the relationship between the distributions of populations, samples, and sampling. In this instance, PTs were expected to be able to think and express that any sample taken from a non-normally distributed population could follow a normal distribution and to circle False. As the pre-test scores demonstrate, 36.90% of the PTs appeared to perceive a link between the distribution of the population and the sample, expressing that the distribution of a population and any samples taking from that population should behave in a parallel manner. However, this misconception decreased on the post-test (2.38%) and delayed test (1.19%); moreover, the

increased from the post-test ($f = 628$) to the delayed test ($f = 728$). Therefore, the decrease in the number of responses scored with 2 points appears to have impacted this statistically significant result. Namely, although there was a decrease in the success rate of the responses from the post-test to the delayed test, the statistically significant difference between the pre-test and the delayed test indicates that the PTs still retained the knowledge provided in the CCTs to the time of the delayed test. As such, further studies using a delayed test are recommended to examine how well students persist in maintaining their conceptual understanding, as proposed by Södervik et al. (2015).

Overall, both the current study and the existing literature agree that CCTs or refutation texts enhance PTs' conceptual understanding; moreover, some researchers have underscored that the acquisition of conceptual understanding is particularly noteworthy among students who had weak prior knowledge (Braach et al., 2013; Diakidoy et al., 2011; Södervik et al., 2013; Södervik et al., 2015). On the other hand, this study showed that the use of the CCTs promoted conceptual understanding not only among PTs with weak prior knowledge or poor performance on the pre-test, but also among those who had strong prior knowledge or performed well on the pre-test. It is possible that the quality of the CCTs and the accordance between the use of the CCTs and sampling distributions in this study may have differentiated the results of the present study from other research that involved different topics.

In this study, it was determined that the CCTs were effective for promoting the conceptual understanding of the PTs about sampling distributions, as there was a decrease in the number of responses that evidenced misunderstandings or misconceptions from the pre-test to the post-test and delayed test. On nearly all the questions, it could be seen that the PTs' conceptual understanding had improved through the number of appropriate explanations given for the statements. In other words, not only did they respond correctly to the questions, but they also gave meaningful explanations for their responses. Notably, the PTs demonstrated a change from their prior misconceptions such as: *the distributions of a population and sample must exhibit parallel behavior; as the sample size increases, the distribution of the sample approaches normal; associating a sample size ($n > 30$) with a normal distribution; the statistics of the distributions of sample and sample means cannot exceed the population parameters; and larger samples are better*. These erroneous beliefs expressed by the students are commonly reported in the literature (Chance et al., 2004; Ozmen & Guven, 2019; Watkins et al., 2014). However, by providing counterexamples, the CCTs served to exchange these beliefs with the correct scientific ideas. The structure followed in the CCTs may have influenced this outcome. The structure of the CCTs is built on three steps. When designing the CCTs, PTs are first encountered common misconceptions or alternative ideas through the aim of the CCT. Next, the texts refuted the misconceptions, and finally, they emphasized the appropriate and acceptable scientific explanations. With this result in mind, integrating similarly structured CCTs in statistics teaching, particularly for sampling distributions, may provide an effective solution for addressing the challenges related to this difficult topic, improving learning outcomes in inferential statistics.

On the other hand, although the CCTs were effective at promoting the conceptual understanding of the PTs about sampling distributions, the PTs still persisted in certain of their faulty ideas or misunderstandings for some questions, such as: *If a population follows a normal distribution, the sampling distribution of mean appears normal for all possible samples, without taking sample sizes into account (Q12), and the sampling distribution of mean approaches population distribution as the sample size increases (Q15)*. With respect to item Q15, PTs confused the two terms *sample* and *sampling distributions* and thus responded incorrectly. For instance, some responses stated: *That is true. As the sample size increases, the distribution of sample moves closer to the population distribution*. Similarly, Ozmen and Guven (2019) found that students tend to think about *sample*, rather than *sampling distribution*, in answering questions related to sampling distributions. Likewise, Chance et al. (2004) revealed that students in their study confused sampling distribution and the distribution of sample; while Chance et al. (2004) and Ko (2016) reported that students tend to think that sampling distributions move closer to the population. In this respect, although the meaning of

sample and sampling distributions and the relationship between them were explained in the CCTs, the PTs still demonstrated misunderstandings about these concepts, or they did not understand the statement correctly. To overcome this issue, simulations may be used to support the knowledge gained from the CCTs, as recommended by Ko (2016), Ozmen and Guven (2019), Wagaman (2013), and Zieffler et al. (2007). It should be stressed that both interventions - simulations and CCTs - should be used together, rather than individually, to overcome these persistent misconceptions.

We chose to administer a delayed test in this study to determine whether the PTs' understanding of sampling distributions were permanent. As this understanding is an essential aspect of inferential statistics, we suggest that future studies should also include delayed tests. In this study, there was a statistically significant difference not only between the pre-test and the delayed test, but also between the post-test and the delayed test, indicating that the effect of the CCTs decreased in the time between the post-test and the delayed test. However, the average scores on the post-test ($\bar{X} = 16.90$) and the delayed test ($\bar{X} = 15.62$) were only slightly different, indicating that the greater number of 2-point responses on the post-test, as opposed to the delayed test, contributed to a statistically significant effect on the results. It is seen that PTs' accuracy success (2-point responses) changed as 11.01% (total 148), 29.46% (total 396), and 21.73% (total 292) from pre-test to post and delayed test. This is one of the important evidence about this effect. The percentage of the accuracy success increased dramatically. Of course, it would be more. It was seen that PTs' have necessary knowledge and justify correct answer. On the other hand, they could present partial answers when they explain their justification and these answers also get 1-point. Besides, the justification success (1-point responses) of the PTs' increased from pre-test to post and delayed test. Moreover 1 point is given both determining true option with any explanation or partial explanation. Thus, the percentage of the 1-point responses is higher than the 2-point responses. On the other hand, CCTs could be effective on the development of the explanation of the answers from the pre-test to the post and delayed test partial or exactly. In this sense, while the time that passed led to a decrease in the accuracy of the PTs regarding the statements, they were still able to determine correctly whether the statements were true. Therefore, to promote the permanence of the knowledge gained from the CCTs, more examples could be provided in the texts. Furthermore, in this study, the PTs were asked to read all five of the CCTs at the same time. This may have also led to a decrease in the permanence of the knowledge gained through the treatment.

Finally, the results demonstrated that, due to the knowledge gained from the CCTs, the PTs improved their conceptual understanding about sampling distributions, which is one of the more difficult topics in statistics. There are numerous studies in the literature concerning the use of CCTs to develop learners' conceptual understanding, but these are primarily related to science education; studies involving statistics and mathematics education on this topic are limited. As such, the current study addresses an important gap in the literature. Moreover, this study represents an important effort to integrate CCTs in mathematics and statistics education. To expand on the results, additional research should be conducted to investigate the effect of CCTs with respect to other topics related to inferential statistics, such as confidence intervals, hypothesis testing, and so on. In addition, because there are many other subjects in mathematics that present challenges for learners, CCTs may be applied in other contexts to promote conceptual understanding. Moreover, while the current study aimed to enhance conceptual understanding and to determine the effect of CCTs on PTs' understanding of sampling distributions during the process of pre-test to delayed test, researchers may also design a learning environment that is based on CCTs and supported with simulations. Similarly, in literature applying diverse activities prior to computer simulations appears to have positive effect on student exam performance (Hancock & Rummerfield, 2020). Also, Lane (2015) underlined that using a simulation is an effective way to teach about sampling distributions if it is used in a pedagogically sound manner. Such an environment could be even more effective in addressing persistent misconceptions in a

pedagogically sound manner. The fact that the CCT experiment was not compared to any alternative designs can be seen as a limitation; as such, comparative studies may also be carried out to assess the effectiveness of this technique in a more detailed experiment. Moreover, the present study showed that correcting misconceptions was a key factor in achieving learning gains, but the CCT experiment was not compared directly with non-CCT factors (e.g., traditional experiments). Therefore, comparison studies are suggested to reveal CCT's effects in detail.

References

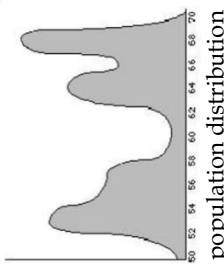
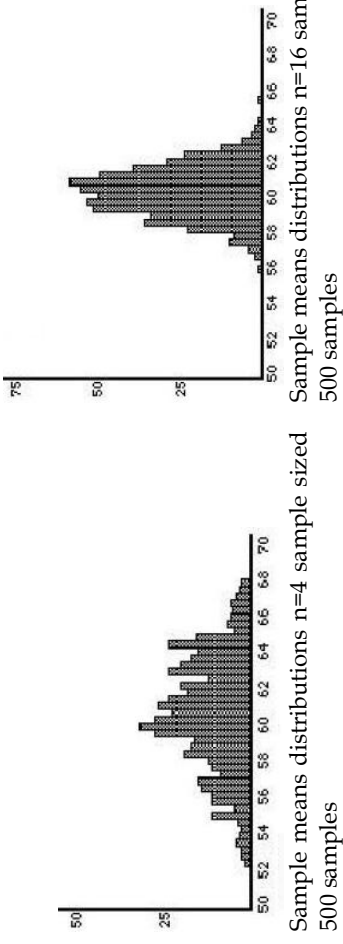
- Akpınar, M., & Tan, M. (2011). Developing, implementing, and testing a conceptual change text about relativity. *Western Anatolia Journal of Educational Sciences, Special Issue*, 138-144.
- Beerenwinkel, A., Parchmann, I., & Gräsel, C. (2011). Conceptual change texts in chemistry teaching: A study on the particle model of matter. *International Journal of Science and Mathematics Education*, 9(5), 1235-1259. <https://doi.org/10.1007/s10763-010-9257-9>
- Braach, J. L., Goldman, S. R., & Wiley, J. (2013). The influences of text and reader characteristics on learning from refutations in science texts. *Journal of Educational Psychology*, 105(3), 291-309. <https://doi.org/10.1037/a0032627>
- Çalık, M., Ayas, A., & Coll, R. K. (2007). Enhancing pre-service elementary teachers' conceptual understanding of solution chemistry with conceptual change text. *International Journal of Science and Mathematics Education*, 5(1), 1-28. <https://doi.org/10.1007/s10763-005-9016-5>
- Castro, C. S. (1998). Teaching probability for conceptual change. *Educational Studies in Mathematics*, 35, 233-254. <https://doi.org/10.1023/A:1003182219483>
- Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 275-293). Kluwer.
- delMas, R. C., Garfield, J., & Chance, B. (1999). A model of classroom research in action: developing simulation activities to improve students' statistical reasoning. *Journal of Statistics Education*, 7(3), 1-16.
- Diakidoy, I. A., Mouskounti, T., & Ioannides, C. (2011). Comprehension and learning from refutation and expository texts. *Reading Research Quarterly*, 46(1), 22-38. <https://doi.org/10.1598/RRQ.46.1.2>
- diSessa, A. A. (2006). A history of conceptual change research: Threads and fault lines. In R.K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences* (pp. 265-281). Cambridge University Press.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: connecting research and teaching practice*. Springer.
- Guzzetti, B. (2000). Learning counter-intuitive science concepts: What have we learned from over a decade of research? *Reading and Writing Quarterly*, 16(2), 89-98. <https://doi.org/10.1080/105735600277971>
- Günay, B. (2005). *Conceptual change text oriented instruction to facilitate conceptual change in atoms and molecules* [Unpublished masters' thesis]. Middle East Technical University, Ankara.
- Hancock, S. A., & Rummerfield, W. (2020). Simulation methods for teaching sampling distributions: Should hands-on activities precede the computer? *Journal of Statistics Education*, 28(1), 9-17. <https://doi.org/10.1080/10691898.2020.1720551>
- Hynd, C., Qian, G., Ridgeway, V., & Pickle, M. (1991). Promoting conceptual change with science texts and discussion. *Journal of Reading*, 34(8), 596-601.
- Ko, E. S. (2016). Development of an understanding of a sampling distribution. In D. Ben-Zvi & K. Makar (Eds.), *The Teaching and Learning of Statistics* (pp. 63-70). Springer.
- Köse, S., Kaya, F., Gezer, K., & Kara, İ. (2011). Computer assisted conceptual change texts: An example on course implementation. *Pamukkale University Journal of Education*, 29(1), 73-88.
- Lane, D. M. (2015). Simulations of the sampling distribution of the mean do not necessarily mislead and can facilitate learning. *Journal of Statistics Education*, 23(2). <https://doi.org/10.1080/10691898.2015.11889738>
- Moseley, B., & Brenner, M. A. (1997). *Using multiple representations for conceptual change in pre-algebra: A comparison of variable usage with graphic and text based problems*. ERIC Document Reproduction Service: ED413184.
- Önder, İ., & Geban, Ö. (2006). The effect of conceptual change texts oriented instruction on students' understanding of the solubility equilibrium concept. *H.U. Journal of Education*, 30, 166-173.
- Ozmen, Z. M. & Guven, B. (2019). Evaluating students' conceptual and procedural understanding of sampling distributions. *International Journal of Mathematical Education in Science and Technology*, 50(1), 25-45. <https://doi.org/10.1080/0020739X.2018.1467507>

- Özmen, H., & Naseriazar, A. (2018). Effect of simulations enhanced with conceptual change texts on university students' understanding of chemical equilibrium. *Journal of the Serbian Chemical Society*, 83(1), 121-137. <https://doi.org/10.2298/JSC1612220650>
- Posner, G., Strike, K., Hewson, P., & Gertzog, W. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science & Education*, 66(2), 211-227. <https://doi.org/10.1002/sce.3730660207>
- Saldanha L., & Thompson P. (2007). Exploring connections between sampling distributions and statistical inference: an analysis of students' engagement and thinking in the context of instruction involving repeated sampling. *International Electronic Journal of Mathematics Education*, 2(3), 270-297. <https://doi.org/10.29333/iejme/213>
- Sinatra, G. M., & Broughton, S. W. (2011). Bridging reading comprehension and conceptual change in science education: The promise of refutation text. *Reading Research Quarterly*, 46(4), 374-393. <https://doi.org/10.1002/RRQ.005>
- Södervik, I., Mikkilä-Erdmann, M., & Vilppu, H. (2013). Promoting the understanding of photosynthesis among elementary school student teachers through text design. *Journal of Science Teacher Education*, 25(5), 581-600. <https://doi.org/10.1007/s10972-013-9373-9>
- Södervik, I., Virtanen, V., & Mikkilä-Erdmann, M. (2015). Challenges in understanding photosynthesis in a university introductory biosciences class. *International Journal of Science and Mathematics Education*, 13(4), 733-750. <https://doi.org/10.1007/s10763-014-9571-8>
- Tippett, C. D. (2010). Refutation text in science education: A review of two decades of research. *International Journal of Science and Mathematics Education*, 8(6), 951-970. <https://doi.org/10.1007/s10763-010-9203-x>
- Triola, M. F. (2010). *Elementary statistics* (11th ed.). Addison Wesley.
- Tsai, C. C. (1999). Overcoming junior high school students' misconceptions about microscopic views of phase change: A study of an analogy activity. *Journal of Science Education and Technology*, 8(1), 83-91. <https://doi.org/10.1023/A:1009485722628>
- Vanhoof, S., Castro-Sotos, A. P., Onghena, E., & Verschaffel, E. (2007). Students' reasoning about sampling distributions before and after the Sampling distribution activity. In International Statistical Institute (Ed.), *Proceedings of the 56th Session of the International Statistical Institute* (pp.1-4). Lisbon, Portugal.
- Wagaman, J. C. (2013). Sampling distributions for introductory statistics students using internet polling software. *International Journal of Mathematical Education in Science and Technology*, 44(1), 468-473. <https://doi.org/10.1080/0020739X.2012.729616>
- Watkins, A. E., Bargagliotti, A., & Franklin C. (2014). Simulation of the sampling distribution of the mean can mislead. *Journal of Statistics Education*, 22(3), 1-21. <https://doi.org/10.1080/10691898.2014.11889716>
- Zieffler, S. A., Garfield, J., delMas, R., & Gould, R. (2007, August). Studying the development of college students' reasoning about statistical inference [Paper presentation]. *Fifth International Research Forum on Statistical Reasoning, Thinking and Literacy (SRTL-5)*, University of Warwick, UK.

Appendix Test Questions, Answers and Ideal Explanation

Test Questions	Answer	Ideal Explanation
Q1) As the sample size (n) increases, the distribution of sample becomes more like population. It is not sample becomes more like normal	False	As the sample size (n) increases, the distribution of sample becomes more like population. It is not obligatory to show normal distribution.
Q2) Any sample selected from a population which behaves normal also looks normal	False	You could select your sample among extreme values from the population distribution. Thus, your sample could show uniform or skewed distribution and could not behave normal.
Q3) Regardless of the distribution of the population, the mean of the sampling distribution of the mean with all possible n-unit samples selected from this population is equal to μ	True	When the all-possible n-unit samples are considered, each of data would be used parallel with the frequencies again and the average values will be equal. Also, CLT tell us that if the original population has mean μ , the mean of the sample means distributions will also be equal to μ .
Q4) Samples selected from a population that is not normally distributed do not show normal distribution.	False	It is false. For example, you can select a part of the distribution which seems to look like normal.
Q5) As the sample size (n) increases, the distribution of the sample becomes more like the distribution of the population.	True	As you increase the sample size, it is started to assemble to the population distribution and close to its characteristics
Q6) Regardless of what the population distribution looks like, the sampling distributions of mean with all possible samples follows a normal distribution	False	According to CLT, if the population distribution does not look like normal, the sampling distributions with all possible n unit samples which taken from the population should behave normal as $n > 30$. Thus, the statement is false.
Q7) As the number of the samples of size (n) increases, the sampling distributions of the means would gradually behave like the normal distribution.	True	The CLT tells us that for a population with any distribution, the distribution of the sample means approaches a normal distribution as the sample size increases.
Q8) In order for the mean of any sample selected from a normally distributed population to be equal to μ , $n > 30$ should be chosen	False	False. For example, we have a population looks like normal and μ (population mean) is equal to 7. We can select 8-unit sample as below: 1, 6, 7, 11, 3, 4, 3, 6, $\bar{X} = 7$. Although sample size is small than 30, \bar{X} could be equal to μ .
Q9) The standard deviation of the sampling distributions of the mean with all possible n-sized samples selected from a population will be equal to the standard deviation of the population.	False	Standard deviation of the sample means distributions is called as standard error and calculated with the formula of $\frac{\sigma}{\sqrt{n}}$. Thus, standard deviation of the sample means distribution have to be smaller than the standard deviation of the population.
Q10) Any sample selected from a population of $n > 30$ may not be normally distributed.	True	True. We can select the extreme values and we can obtain skewed distributions.
Q11) The mean of any sample taken from a population, having the average μ , is equal to μ	False	Suppose that the population is 7, 3, 13, 12, 15, 12, 16, 10 and we can select the sample as 3, 12, 15, 10. $\mu = 11$, $\bar{X} = 10$. Thus, the mean of sample and population could be different.
Q12) If a population follows a normal distribution, the distribution of the sample mean appears normal for all possible samples, without taking sample sizes into account.	True	The CLT tells us that for a population with any distribution, the distribution of the sample means approaches a normal distribution as the sample size increases. Also, CLT principles point out that. *For a population with any distribution, if $n > 30$ then the sample means have a distribution that can be approximated by a normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ *If $n < 30$ and the original population has a normal distribution, then the sample means have a normal distribution with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$

Appendix continued

Test Questions	Answer	Ideal Explanation
Q13) For a distribution to look normal, it is sufficient and necessary for it to be $n > 30$.	False	False. Suppose that your distribution is about the birth height of the babies and its size is 40. If you chose 40 new-birth babies among the extreme values (babies which have very high or low birth-heights). Therefore, $n > 30$ and it does not follow the normal distribution.
Q14) The standard deviation of a population may be smaller than the standard deviation of any sample selected from this population.	True	Suppose that the population elements are 6, 6, 7, 8, 8, 8, 9, 10, 10, 10, 11, and the sample which taken from this population is 6, 6, 10, 11. $\sigma = 1,554$ (population mean) and $s = 2,63$. Therefore σ could be smaller than s .
Q15) The distribution of the sample means approaches the population distribution as the sample size increases.	False	The CLT tells us that for a population with any distribution, the distribution of the sample means approaches a normal distribution rather than population distribution as the sample size increases. In other words, if the sample size is large enough, the distribution of sample means can be approximated by a normal distribution, even if the original population is not normally distributed.
		 <p>population distribution</p> <p>Suppose that we take $n=4$ and $n=16$ sample sized 500 samples. The distribution of the sample means would be as below:</p>  <p>Sample means distributions $n=4$ sample sized 500 samples</p> <p>Sample means distributions $n=16$ sample sized 500 samples</p>
Q16) As the sample size increases, the variation of the sampling distributions of mean would gradually decrease.	True	It could be seen that as the sample size increases sample means distributions behave normal distributions rather than look like the population distribution. When we select all possible n sample sized samples from a population standard error would be equal to $\frac{\sigma}{\sqrt{n}}$ and therefore it would gradually decrease as the sample size. If n is bigger $\frac{\sigma}{\sqrt{n}}$ would be smaller too.