Journal of the Korea Society of Mathematical Education Series D:
Research in Mathematical Education
Vol. 16, No. 4, December 2012, 265–276

Preservice Teachers' Difficulties with Statistical Writing[1]

PARK, Min-Sun*
Department of Mathematics Education, Graduate School, Seoul National University,
Seoul 151-748, Korea; Email: dpdlxl03@snu.ac.kr

PARK, Mimi
Department of Mathematics Education, Graduate School, Seoul National University,
Seoul 151-748, Korea; E-mail: mimipark1127@gmail.com

LEE, Eun-Jung
Department of Mathematics Education, Graduate School, Seoul National University,
Seoul 151-748, Korea; E-mail: ejlee13@snu.ac.kr

LEE, Kyeong Hwa
Department of Mathematics Education, Seoul National University,
Seoul151-748, Korea; Email: khmath@snu.ac.kr

(Received October 24, 2012; Revised December 15, 2012; Accepted December 16, 2012)

These days, with the emphasis on statistical literacy, the importance of communication is the focus of attention. Communication about statistics is important since it is a way of describing the understanding of concepts and the interpretation of data. However, students usually have trouble with expressing what they understand, especially through writing. In this paper, we examined preservice teachers' difficulties when they wrote about statistical concepts. By comparing preservice teachers' written responses and interview transcripts of the variance concept task, we could find the missing information in their written language compared to their verbal language. From the results, we found that preservice teachers had difficulty in connecting terms contextually and conceptually, presenting various factors of the concepts that they considered, and presenting the problem solving strategies that they used.

Keywords: statistical writing, statistical literacy, written language, verbal language
MESC Classification: D79
MSC2010 Classification: 97D70

[1] A draft version of the article was presented at KSME 2012 Fall Conference on Mathematics Education at Korea National University of Education, Cheongju, Chungbuk 363-791, Korea; November 2–3, 2012 (*cf.* Park, Park, Lee & Lee, 2012)..

* Corresponding author

INTRODUCTION

Statistical literacy has been emphasized as a goal of statistics education. Statistical literacy is the ability to interpret, critically evaluate, and communicate statistical information, arguments, and data (Gal, 2004, p.70). In information-laden societies, statistical literacy is necessary for citizens since they have to choose and interpret useful information for them. With the emphasis on statistical literacy, the importance of communication is receiving a lot of attention. In addition to the generic reason that it improves statistical literacy, communication is important in statistics for many other reasons. Since statistics is concerned with information about the real world, people have to be able to take problems vaguely conceived in natural language terms through the statistical investigation and analysis cycle to arrive at conclusions that they can successfully communicate to others in natural language (Phillips, 2006).

Communication involves listening and speaking, reading and writing, and representing (Begg, 1997, p. 19). Among these, 'writing' is very important, since it is mainly used as a way of assessment (Weldon, 2007; Biehler, 2007; Truran, 1998), and it is also the last step of presenting what people conclude from data analysis. Nevertheless, students usually have trouble with expressing what they understand through writing (Pierce & Roberts, 1998, p. 1202). Many researchers have investigated ways to facilitate students' writing (Parke, 2008; Francis, 2005; Peck, 2005). However, researches usually suggest guidelines or particular writing formats, rather than discuss the parts that students cannot express through writing and suggest a solution. Thus, it is necessary to investigate which parts are being excluded from their writing and present a way to bridge the gap between their thinking and writing.

In this paper, by looking at both written and verbal language, we examined the missing information in preservice teachers' writing even though they understood and considered. For the investigation, a written task about the variance concept was carried out and interviews were conducted. We compared both written and verbal language in the preservice teachers' use of terms, presentation of factors of the concepts that they had considered and presentation of the solving strategies that they had used. From the results, we suggest instructional ideas for teaching writings and some consideration points for writing assessment.

LITERATURE REVIEW

In mathematics education, there are some studies that argue that writing can support

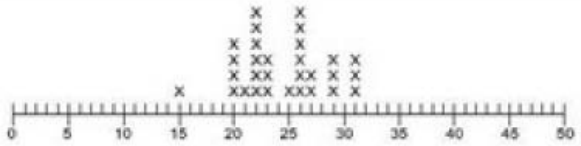
the problem solving by improving meta-cognitive ability. Pugalee (2004) compared ninth grade students' written and verbal descriptions of their algebraic problem solving processes. Through the comparison, he tried to find the connection between problem solving and writing. As a result, students who wrote descriptions of their thinking were significantly more successful in the problem solving tasks than the students who verbalized their thinking. Differences in students' use of both types of language showed that writing can be an effective tool in supporting meta-cognitive behaviors.

In statistics education, writing is emphasized as well. Lipson & Kokonis (2005) showed that a writing task may be classified as a meta-cognitive activity, and in it of itself provides a means of facilitating the development of conceptual understanding in students. Parke (2008) also investigated the influence of writing. Individual writing assignments, small group activities, and a student-led scoring activity enhanced students' writing as well as their reasoning, understanding, and confidence. Writing tasks encouraged students to take a holistic view of the statistical process (Lipson & Kokonis, 2005, p.8).

Accordingly, there are many studies presenting ways to facilitate students' writing. Francis (2005) presented an approach that involves giving students a process to follow, clear instructions on the sort of language which is appropriate and some model reports to use as a guide. Peck (2005) also suggested some ways of facilitating students' writing: being explicit about what is needed for good communication in different settings, emphasizing the importance of context, asking questions that require explanation and interpretation throughout the course, not accepting "mechanics only" answers as correct on homework or exams, encouraging students to read as well as to write, and asking students to write about statistical processes.

Even though both the importance of statistical writing and the way of facilitating students' writing are issued continuously, students still have much trouble with writing. Francis (2005) pointed out some of the reasons, which are as follows: considering statistics as divorced from the real world rather than a source of information about the real world, not knowing what the statistical analysis is about, experiencing difficulty in writing a cohesive report even when understanding a particular analysis, not understanding some of the subtleties of the language, having difficulty with understanding and using statistical terms correctly, and being unsure of what should be included in writing and what should not. All of these difficulties can be connected to the main ideas of statistical literacy. Especially, these are relevant to statistical knowledge and context knowledge, which are knowledge elements in a model of statistical literacy given by Gal (2004, p. 51).

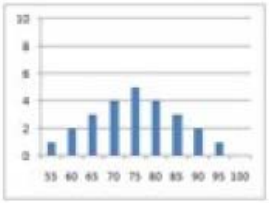
2. There is a spinning disk that is half-white and half-black. A class did 30 sets of 50 spins and the results for the number of times it landed on the black part are recorded below.

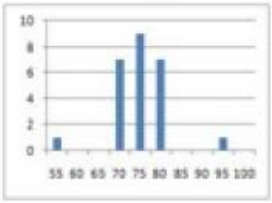


2-1) If you got a variance for the graph above, what does it mean?

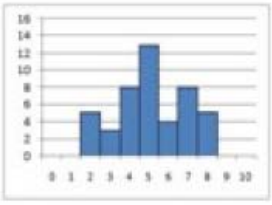
5. For each pair of graphs, determine which graph has the bigger variance. Explain your answer.

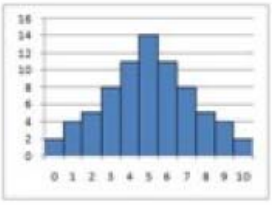
5-1) ① ②



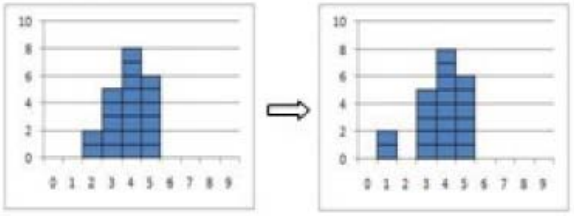


5-2) ① ②





8. In the following graph, if the two lowest data points had a value of 1 instead of 2, how would the variance be changed? Explain your answer.



11. Of the followings, determine which sentence has the bigger variance. Explain your answer.

① Age of trees in a national forest ② Diameter of new tires coming off one production line
③ Scores on an aptitude test given to a large number of job applicants ④ Daily rainfall
⑤ Weight of a box of cereal

Figure 1. Examples of several items in written assessment

METHODOLOGY

In this research, 44 preservice teachers took a writing assessment and 12 of them, who were selected by the method of stratified sampling, were interviewed. Since preservice teachers should evaluate their students in the future and should be able to respond correctly in the writing assessment and interview, they are appropriate participants.

The task used in the writing assessment was about a variance concept. We were focused on presenting items which require explanation and interpretation (Peck, 2005) to facilitate preservice teachers' writing rather than presenting items which require writing about a simple definition of variance. As a result, there were items asking about the meaning of variance in a particular context, on comparing the degree of variances with reasons, and on estimating the change of variance when data sets were changed with reasons. The items were taken from previous studies on variability (Watson, Kelly, Callingham & Shaughnessy, 2003; Canada, 2004; Lee & Meletiou-Mavrotheris, 2003; delMas & Liu, 2005; and CAOS test[2]), and after the pilot study, the items were modified for preservice teachers' understanding. Several items used in the writing assessment are given in Figure 1.

The writing assessment took 40 to 70 minutes and the interview took 30 to 60 minutes per person. In the interview, preservice teachers were asked about their way of thinking when they took the writing assessment. By analyzing both the written responses and interview results, we tried to find the missing information in their written language compared to the verbal language which revealed what the preservice teachers knew or had considered. There is some possibility that the researcher's reactions during the interview could affect the interviewee's response; also, preservice teachers were able to change their answers after they thought about the question again meta-cognitively. Because of these limitations, we excluded the parts for which preservice teachers changed their answers and spoke retrospectively. The interview was in a semi-structured format, and every interview was recorded and transcribed for analysis. From the written response and interview transcripts, we compared the following: the subjects' usage of terms, presentation of factors of the concepts that they considered and presentation of the problem solving strategies that they used.

RESULTS AND DISCUSSION

After comparing the preservice teachers' written response and interview results, there

[2] cf. https://apps3.cehd.umn.edu/artist/caos.html

were the following differences between written language and verbal language: connecting of the terms contextually and conceptually, presentation of various factors of the concepts that they considered, and presentation of the problem solving strategies that they used. These were not apparently exposed in written language, which means that these are the points with which preservice teachers have difficulty.

Difficulty in connecting terms contextually

In the writing assessment, almost all of the items were based on some context. Thus, the preservice teachers were required to interpret the meaning of the variance of the given data sets rather than merely provide the formal definition of the variance. This requirement ensured that the preservice teachers could connect their usage of terms to the context. Several preservice teachers, in their written responses, described the formal definition of variance instead of reflecting the context of the problem. However, in the interview, they showed much understanding of the context.

Table 1. Examples of difficulty in connecting terms contextually

Student (Items in Figure 1)	Written Language	Verbal Language
S7 (Q2-1)	The degree of the distance of the data values from the mean	Maybe I was thinking about the mean value that I calculated. I used some process of elimination. I can eliminate the same number of each data which are at the same distance from 25. It was not 25 exactly. Maybe about 24.5 or 25.5, if I remember it correctly… So I think that the variance would be the degree of distance from that point.
S10 (Q2-1)	The degree of the spread in the result	In the experiment, somewhat regular values should be given. So point 15 seems awkward. If this graph was not given and only the variance value was given, we can calculate the probability without looking at this graph. If the variance seems unusually big, then we can expect that in the experiment, there were some extreme points, like 15. That's what we can know from the variance.

Both S7 and S10 described the formal definition of variance in written language. So their terms that were used to show the meaning of the concept did not reflect the context well. In contrast, the preservice teachers considered the mean of the given data sets and focused on the outlier in the interview, which means that they used appropriate words connected to the context. As Cobb & Moore (1997) said, data in statistics are not just numbers, but numbers with a context. Therefore, considering the context is essential. Peck (2005) mentioned that the importance of context is a primary reason that communi-

cation is such an important aspect of statistics problems. From the result, we examined the preservice teachers' difficulty in connecting terms contextually in written language.

Difficulty in connecting terms conceptually

When communicating about statistics, the usage of terms showing the understanding of the concept which is included in the problem was required. The task was about a variance concept. To show conceptual understanding of the variance, the preservice teachers should properly connect explanatory terms like mean, deviance, and frequency to common words. If they use their own informal terms, then their response would be considered to be the opposite of a conceptual response. Some preservice teachers used their own terms in the writing assessment; even though they connected terms conceptually well in the interview.

Table 2. Examples of difficulty in connecting terms conceptually

Students (Items in Figure 1)	Written Language	Verbal Language
S9 (Q5-1)	Graph ① Less pointed	... In graph ②, the data values are very crowded around 75 and in graph ①, the data values are evenly spread from 75. So I thought ②'s variance is smaller because it is much gathered. I mentioned "pointed" in my reason because that is the term that I usually use.
S9 (Q5-2)	Graph ① Thick tails	Suppose that the mean is 5. If the values of both ends are big, then the deviance would be big accordingly. Since the values show big deviance, meaning that the values on the very end are big, or have high values, the variance is bigger. That's why I wrote "thick tails."
S1 (Q11)	Item ④ Wide spectrum from no rain to very much rain	There are days when there is no rain and there are days when there is very much rain. We have to find the mean of rainfall from those days. So if the mean lies between these days, then the deviance would be big.

We can see that S9 used his own terms like "pointed" and "thick tails." He said that he used those words because they were the words that he usually used. However, in the interview, he tried to approach the variance conceptually by connecting terms like mean, data values, and deviance. Likewise, S1 referred to a big range as a "wide spectrum." However, in the interview, she showed her understanding of the variance by using terms like mean and deviance. Using statistical terms is important since it is an element of the statistical knowledge base in a model of statistical literacy. Moreover, preservice teachers should connect those terms to the common words conceptually to show their understand-

ing of the concept. From Table 2, we could see that preservice teachers feel difficulty in connecting terms conceptually.

Difficulty in presenting various factors of a concept

When dealing with the variance concept, people should consider various factors of variance: Mean data values, frequency, and variability, which can be seen from the distribution. Especially, when trying to compare the degree of variance, they should consider more than two factors. The preservice teachers presented one or two factors related to the meaning of variance in the written assessment; however, in the interview, they explained variance by including various factors that they had actually considered.

Table 3. Examples of difficulty in presenting various factors of a concept

Students (Items in Figure 1)	Written Language	Verbal Language
S5 (Q5-1)	① The number of data values and degree from the mean are bigger in ① than ②.	… In graph ①, since the size of the data sets are the same, the degree of spread would affect the variance. Obviously, the mean of both graphs is 75, and they are symmetric. If in graph ②, the values of 55 and 95 were 45 and 105, then there would be many things to consider. However, they are 55 and 95 and both the number and data are the same as those of ①. Thus, they do not affect the variance very much. …
S11 (Q8)	The variance gets bigger since the gathered data values are scattered.	S11: ... If we look at the shape of the graph, only these data values are moved in this way. Then the mean would be moved in the same way and the distance between these values and the mean would be bigger. If we square those numbers, that is, if we consider the formula of variance, then the value of $(x-m)^2$ gets bigger. So the variance gets bigger. R: The explanation you just gave and the writing… When you solved this problem, did you consider what you just said? S11: I thought about the movement of the mean. I don't think I considered all of the things specifically. I just thought that the mean would be changed and the deviance would be bigger.

In the writing assessment, S5 only considered the mean and data values. However, from the interview, we could find that he was considering the same value, symmetric graph, and the same range, which are rather various and specialized. Also, S11, in his written response, wrote only about his consideration of the movement of data values. That

is a different result from that of the interview in which he mentioned the movement of the mean and the change of the deviance of other data values. He said that he had somewhat considered those factors when he solved the problem. Distribution itself is a multifaceted concept (Bakker & Gravemeijer, 2004). When considering the variability of the distribution, we should integrate various factors of the distribution. Regarding this point, we could find that the preservice teachers had difficulty in writing about integrating the various factors of the concept.

Difficulty in presenting problem solving strategies

Table 4. Examples of difficulty in presenting problem solving strategies

Students (Items in Figure 1)	Written Language	Verbal Language
S11 (Q5-1)	① Values are rather spread to the outer side.	S11: The end points, which are 55 and 99, are 1 each, and the middle area is empty. In graph ②, all of the values are stacked in the middle. So if we consider the possibility of changing ② to ①, that is, keeping in mind the changing situation, then we can imagine cutting these points in the middle and sending them to the side. We can think like that. R: So you made a transformation? S11: Yes, I changed the data… the shape of the graph. Then now we can think that if the mean is 75, we sent data values around the mean to the side, and the variance got bigger.
S6 (Q5-1)	① The frequencies of the values which are far away from the mean are relatively bigger.	S6: ... If we draw the shape of the graph, ② is much thinner and ① has wide bell-shaped. In my mind, the variance and standard deviation of the wide bell-shaped graph are always big. So I could make a guess. R: How did you make those images? S6: In high school, I learned a lot about normal distribution and I usually draw the graph in a bell-shape. Also, I saw that thinner graphs have small standard deviation and wide graphs have big standard deviation. I applied those images to this problem.
S7 (Q5-2)	① The number of values that are far away are bigger than those that are near.	If we change the order of the data values like a step function, … If we change these two values, then it would be in the shape of steps. Also, if we send this value to the end, then it would be in the shape of steps. Oh, not steps, but mountains. No, maybe a pyramid? Anyway, it would be in the shape of a mountain. Then both graphs are shaped like a mountain, which is the shape that has the smallest variance. So, intuitively, the first one was much more twisted than the mountain shape, which means that it has a bigger variance.

When comparing variance, people usually use some strategies. delMas & Liu (2005)

presented various strategies that students used in comparing standard deviation. The preservice teachers did not mention the strategies that they used in their written responses. However, after the interview, we could find that they indeed used some strategies like changing one graph to the other graph or imagining a bell-shaped graph with which they had already dealt.

S11, in the writing task, compared the variance of the graphs using the degree of the spread. However, in the interview, he mentioned the strategy that he had used, which was sending some of the data values in one graph to apply them to the other graph. By checking the movement of the data values, he could compare the variance of the graphs. S6 also presented some factors of variance in writing, which differed from the interview result where she mentioned her image of bell-shaped graphs. In the case of S7, he mentioned the strategy of transforming the graph like S11 did. Also, by saying that the "mountain shape is the one that has the smallest variance," he also used some images in his mind like S6 did.

All three preservice teachers presented strategies and they were focusing on the informal aspect of the distribution that Bakker & Gravemeijer (2004) emphasized. However, these considerations were not communicated well in the writing assessment, which means that the preservice teachers have difficulty in presenting problem solving strategies.

CONCLUSION

All four aspects that are presented in the results section are important aspects that students should have for statistical literacy. Students should be able to connect terms contextually and conceptually, and present various factors of a concept and the problem solving strategies that they considered or used. From the interviews, we could find that the preservice teachers considered all of the four aspects; however, they could not explain them in writing very well though they were able to discuss the aspects during the interviews. If we compare those difficulties with the result of Francis (2005), difficulty in connecting terms contextually is relevant to "considering statistics as divorced from the real world rather than a source of information about the real world," and difficulty in presenting various factors of a concept and the problem solving strategies that they considered or used is relevant to "being unaware of what belongs in writing and what does not." Francis also mentioned the difficulty of understanding and using statistical terms correctly; however, in this paper, we found the additional difficulty of connecting statistical terms and common words conceptually rather than the direct use of statistical terms.

This study included tasks which require explanation and interpretation to facilitate preservice teachers' writing (Peck, 2005). However, the gap existing between the

preservice teachers' written and verbal language indicates that they still have difficulty in writing. Further research should consider other ways of facilitating students' writing. The difficulties of presenting various factors of a concept and the problem solving strategies that they considered or used could be resolved if we include some items that ask students to write about statistical processes, as Peck mentioned.

REFERENCES

Bakker, A., & Gravemeijer, K. (2004). Learning to reason about distribution. In: D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 147–168). Dordrecht: Kluwer Academic Publishers.

Begg, A. (1997). Some emerging influences underpinning assessment in statistics. In: I. Gal & J. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 17–25). Amsterdam: IOS Press.

Biehler, R. (2007). Assessing students' statistical competence by means of written reports and project work. In: B. Chance & B. Philipps (Eds.), *Proceedings of the IASE Satellite Conference on Assessing Student Learning in Statistics*. Guimaraes, Portugal, August 19–21, 2007.

Canada, D. (2004). *Elementary Preservice Teachers' Conceptions of Variation*. Unpublished doctoral dissertation. Portland, USA: Portland State University.

Cobb, G. W. & Moore, D. S. (1997). Mathematics, statistics, and teaching. *Am. Math. Mon.* **104(9)**, 801–823.

delMas, R. & Liu, Y. (2005). Exploring students' conceptions of the standard deviation. *Statistics Education Research Journal* **4(1)**, 55–82.

Francis, G. (2005). An approach to report writing in statistics courses. In: *Proceedings of the IASE Satellite Conference on Statistics Education and the Communication of Statistics. Sydney, Australia;April 4–5, 2005*.

Gal, I. (2004). Statistical Literacy. In: D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 47–78). Dordrecht: Kluwer Academic Publishers.

Lee, C. & Meletiou-Mavrotheris, M. (2003). Some difficulties of learning histograms in introductory statistics. In: *2003 Proceedings of the American Statistical Association, Statistics Education Section* [CD-ROM], (pp. 2326–2333). Alexandria, VA: American Statistical Association.

Lipson, K. & Kokonis, S. (2005). The implications of introducing report writing into an introductory statistics subject. In: *Proceedings of the IASE Satellite Conference on Statistics Education and the Communication of Statistics. Sydney, Australia, April 2005*.

Park, M.-S.; Park, Mimi; Lee, E.-J. & Lee, K. (2012). Preservice Teachers' Difficulties with Statistical Writing. In: J. Cho, S. Lee & Y. Choe (Eds.), *Proceedings of KSME 2012 Fall Confer-*

ence on Mathematics Education at Korea National Universityof Education, Cheongju, Chungbuk 363-791, Korea; November 2–3, 2012 (pp. 399–411).

Parke, C. (2008). Reasoning and communicating in the language of statistics. *Journal of Statistics Education* **16(1).**

Peck, R. (2005). There's more to statistics than computation - teaching students how to communicate statistical results. In: *Proceedings of the IASE Satellite Conference on Statistics Education and the Communication of Statistics. Sydney, Australia, April 4–5, 2005.*

Phillips, B. (2006). Statistics education and the communication of statistics: A report on the IASE/ISI satellite meeting, 2005. In: *Proceedings of the 7th International Conference on Teaching Statistics. Salvador, Bahai, Brazil, July 2–7, 2006.*

Pierce, R., & Roberts, L. (1998). Introductory statistics: Critical evaluation and clear communication. In: *Proceedings of the 5th International Conference on Teaching Statistics. Singapore, June 1998.*

Pugalee, D. (2004). A comparison of verbal and written descriptions of students' problem solving processes. *Educ. Stud. Math.* **55(1–3)**, 27–47. ME **2004c.**02263

Truran, J. (1998). Using students' writings to assess their cognitive and affective development in an elementary economic statistics course. In: L. Pereira-Mendoza, et al., (Ed.), *Proceedings of the 5th International Conference on Teaching Statistics. Singapore, June 21–28, 1998.*

Watson, J. M.; Kelly, B. A.; Callingham, R. A. & Shaughnessy, J. M. (2003). The measurement of school students' understanding of statistical variation. *Int. J. Math. Educ. Sci. Technol.* **34(1)**, 1–29. ME **2003b.**01627

Weldon, L. (2007). Assessment of a writing course in statistics. In: B. Chance & B. Philipps (Eds.), *Proceedings of the IASE Satellite Conference on Assessing Student Learning in Statistics, Guimaraes, Portugal; August 19–21, 2007.*