Retraction

Retraction: The Analysis of Data Distributions Presented in The Indonesian and English College Textbooks Based on Statistical Cognition and Meanings of Data (IOP Conf. Ser.: Mater. Sci. Eng. 536 012149)

Published 25 November 2021

It has come to the attention of IOP Publishing that this article should not have reached publication because of its substantial replication without citation of an earlier paper by the second author (Khairiani Idris 2017, PhD thesis, 1-28). The first author has confirmed they did not contribute to the paper, and the inclusion of the second and third authors and the publication of the paper occurred without their prior knowledge or consent. Consequently, this paper has been retracted by IOP Publishing. The authors have not confirmed whether they agree or disagree with this retraction.

Retraction published: 25 November 2021
The Analysis of Data Distributions Presented in The Indonesian and English College Textbooks Based on Statistical Cognition and Meanings of Data

Ummi Habibah¹, Khairiani Idris², Kai-Lin Yang³

¹Politeknik Negeri Lhokseumawe, Indonesia
²IAIN Lhokseumawe, Indonesia
³National Taiwan Normal University, Taiwan

E-mail: ummihabibah37@yahoo.co.id

Abstract. A statistics textbook should provide an opportunity for students to acquire skills in statistical cognition which involve statistical basic knowledge, reasoning and thinking. The meaning of data, on the other hand, is crucial for presenting the cognition. This study analyzed topic of data distributions presented in the English and Indonesian college statistics textbooks. While the Indonesian version heavily emphasizes on the basic knowledge related to procedures, the English version, variously elaborates each content with basic knowledge or reasoning. Data as numerical numbers are presented for showing procedures in both versions of textbooks. More diverse meaning of data, however, was found in the English version, including data as numbers with meaningful contexts to represent problems, and as information to be evaluated or criticized using reasoning and thinking.

1. Introduction

The concept of lifelong learning in wider context requires college students to interact with academic texts. Particularly, statistics as an evolving science implies the ability to learn through reading becomes prominent for college students. On the other hand, there was an increasingly strong call for statistics education to focus more on statistical literacy, reasoning and thinking in the late of 1990s, which arose from the concern that most statistics courses failed to produce statistically educated students [1]. In reference [2], the authors have proposed three components of statistical cognition to be assessed in statistics education. The statistical cognition is defined as the way students represent knowledge and develop competence in statistics, which composed of three categories: statistical literacy, reasoning and thinking.

Taking the above argumentation into consideration, we argued that a statistics textbook, particularly for college level, should provide an opportunity for students to acquire skills in statistical cognition. The meaning of data in [3], on the other hand, is crucial for presenting the cognition. This study aimed at analyzing the English and Indonesian college statistics textbooks with regards to the presentation of statistical cognitions and meanings of data.
The analyses were focused on the topics under data distribution, which was one of the big ideas in statistics class [2]. The topics include data displays, measures of center and spread (measure), and shape of the distribution (shape).

2. Methods
Two versions of introductory statistics textbooks for college students were selected for the analysis. The English textbook version was *Introduction to the Practice of Statistics* (Moore, McCabe, & Craig, 2009), which was among the best seller statistics textbooks in the popular bookstore website, such as www.amazon.com. The Indonesian version was *Metode Statistika* [4], which was the popular textbook in local bookstores in Indonesian and has been widely used in Indonesian universities.

2.1. Determining unit of analysis
The first step used in analyzing textbook was determining the unit of analysis within the text. Since the focus of analysis in this study was on how the topics under data distribution are addressed in the corpus part of textbooks, the types of text were initially categorized based on text goals usually used in college statistics textbooks. This may include introducing, describing, showing examples, etc. Hence, by considering the types of mathematics text proposed in previous studies (e.g., [5] and [6]) the different characteristics of the writing styles between college and school textbooks, we specified the types of text to suit the context of this study and the selected textbook analyzed. As the results, seven types of texts in college statistics textbooks were proposed as: Introductory text [I]; Definition and Formula [DF]; Explanatory text [E]; Clarification text [C]; Exemplary text [E]; Worked example [W]; and Practice [P].

Table 1. Descriptions and operational classifications of the two components focused on analyzing statistics textbooks in the study

| Component                | Description                                                                                   | Operational classification               |
|--------------------------|-----------------------------------------------------------------------------------------------|------------------------------------------|
| Statistical cognitions   | Type of statistical cognitions used in presenting statistical terms                            | K: Statistical basic knowledge, which includes providing definitions and formulas, procedures, and reading data displays |
|                          |                                                                                              | R: Statistical reasoning, which includes elaborating properties of statistical terms and interpreting data |
|                          |                                                                                              | T: Statistical thinking, which includes thinking in producing data and evaluating or criticizing data |
| Meanings of data         | Type of meanings of data addressed in presenting statistical terms                            | D0: No data are addressed                |
|                          |                                                                                              | D1: Data as numerical numbers without context |
|                          |                                                                                              | D1-C: Data as numerical numbers with contexts |
|                          |                                                                                              | D2: Data as numbers for presenting problem contexts |
|                          |                                                                                              | D3: Data as numbers for investigation     |

2.2. Reliability of coding
In order to avoid the subjective analyses, a co-coder was selected to code several analysis units involved in analyzing textbooks in this study. Hence, there were two sets of coding for these analysis units. As a measure of reliability of the coding, the percentages of agreement between the coders were calculated by dividing the total number of agreements by the total number of agreement plus disagreements. The first round of double coding was conducted for 16% of analysis units in the topic of data display in English version. Before the coding process, the co-coder was trained to code different part of the textbook by using the guideline of coding to assure her understanding of the framework coding. This process resulted in the reliability percentages in 0.60 for statistical cognitions and 0.98 for meanings of data. Some revisions were subsequently done on the guidelines. Second round of double coding was then conducted for another 20% of analysis units in the topic of statistical measurements. The reliability percentages after the final coding process resulted in 0.85 and 1.00 for the statistical cognitions and meanings of data, respectively.
3. Results and Discussion

3.1. Statistical cognitions

Table 2 depicts the distributions of each statistical cognition component weighted to the number of analysis units in the three topics. Since there can be more than one component of statistical cognition presented in one analysis unit, the total statistical cognitions for each analysis unit was set to sum up as 1 point. For instance, when both statistical basic knowledge and reasoning were addressed in an analysis unit, each of the two components was given 0.5 points.

Table 2. The distribution of statistical cognitions weighted to the total analysis units in the three topics (English version; Indonesian version)

| Topic       | Knowledge (K) | Reasoning (R) | Thinking (T) |
|-------------|---------------|---------------|--------------|
| Data Display| (0.48; 0.84)  | (0.46; 0.14)  | (0.06; 0.03) |
| Measure     | (0.65; 0.85)  | (0.21; 0.12)  | (0.14; 0.03) |
| Shape       | (0.52; 0.61)  | (0.41; 0.17)  | (0.08; 0.22) |

Both versions of textbooks provide the highest percentages of analysis units presented basic knowledge in the three topics. In the English version, however, statistical reasoning is presented in almost similar percentages to basic knowledge, except for the topic of measures. As for Indonesian version, great differences between basic knowledge and the other two components of statistical cognitions were observed in the three topics. The different pattern can be noticed in the topic of measure in the English version due to the lowest percentages of reasoning presented (28%) compared to the other two topics (i.e., 69% for graph and 66% for shape). In Indonesian version, distinct pattern is shown in the topic of shape, in which the percentage of thinking is slightly higher (33%) than those in the other two topics. Unlike that in the English version, the topic of shape takes a small content (i.e., 2.89 pages with 9 analysis units) in Indonesian version, which is provided after the concept of the histogram. The whole discussions are presented in exploratory texts pointing out the different shapes of data distributions and mentioning some examples of contexts for each type of shape.

3.2. Meanings of data

Not all analysis units in the textbooks address the component of the meaning of data, which were assigned as D0. In the English version, there are only about half of analysis units addressing this component. That is, 56% in the topic of data display, 49% in the topics of measures and 53% in the topic of shape. In Indonesian version, the higher percentages of analysis units addressing this component are in the topic of data display, i.e., 73%. Topics of measure and shape have, respectively, 39% and 33% of analysis units addressing this component. Table 3 presents the distributions of the meanings of data for each analysis unit.

Table 3. The distribution of meanings of data weighted to the total analysis units in the three topics (English version; Indonesian version)

| Topic       | D0     | D1     | D1-C   | D2     | D3     |
|-------------|--------|--------|--------|--------|--------|
| Data Display| (0.44; 0.27) | (0.00; 0.19) | (0.13; 0.49) | (0.28; 0.05) | (0.16; 0.00) |
| Measure     | (0.51; 0.61) | (0.03; 0.23) | (0.31; 0.09) | (0.08; 0.07) | (0.08; 0.00) |
| Shape       | (0.47; 0.67) | (0.00; 0.11) | (0.28; 0.00) | (0.09; 0.11) | (0.16; 0.11) |

A significant difference between the two versions of textbooks regarding the meaning of data addressed is that almost no data as numerical numbers (D1) is presented in English version, while it is addressed in rather high percentages in the Indonesian version. In contrary, the Indonesian version
does not address the meaning of data as information for investigation (D3) in the topics of data display and measure.

To gain more insights into the ways the four types of meanings of data are presented, Tables 4 depicts the distributions of meanings of data with relate to the types of statistical cognitions. As described in discussions on the results for the component of statistical cognitions that there can be more than one statistical cognitions presented in one analysis unit, the types of statistical cognitions addressing meanings of data can be the combinations of two or more statistical cognitions. Hence, another category is added to the types of statistical cognitions addressing meanings of data: combinations of two or more statistical cognitions (C). The percentage values in each cell are obtained from the total number of analysis units presenting the corresponding statistical cognitions which address the corresponding meaning of data weighted to the total number of analysis units presenting the corresponding type of statistical cognition. For instance, 0.51 in the first row and first column in Table 5.2.5 is the percentage of analysis units presenting knowledge which are used to address D0 in English version. The value is obtained from the total number of analysis units presenting knowledge.

Table 4. The distributions of meanings of data addressed with regard to statistical cognitions (English version, Indonesian version)

| Statistical cognition | Meaning of data | D0      | D1      | D1-C    | D2      | D3      |
|-----------------------|----------------|---------|---------|---------|---------|---------|
| Knowledge (K)         |                | (0.51, 0.50) | (0.00, 0.22) | (0.40, 0.22) | (0.07, 0.06) | (0.02, 0.00) |
| Reasoning (R)         |                | (0.61, 0.91) | (0.00, 0.00) | (0.00, 0.09) | (0.33, 0.00) | (0.06, 0.00) |
| Thinking (T)          |                | (0.60, 0.50) | (0.20, 0.50) | (0.00, 0.00) | (0.00, 0.00) | (0.20, 0.00) |
| Combination (C)       |                | (0.31, 0.53) | (0.00, 0.21) | (0.25, 0.05) | (0.17, 0.16) | (0.29, 0.05) |

A significant difference between the two versions which can be noticed is that English version addresses D1 only with thinking, while Indonesian version addresses D3 only with combinations of statistical cognitions. Furthermore, analysis units presenting thinking in the English version are only used to address two types of meanings of data with similar percentages: D1 and D3, while in Indonesian version, only D1 is addressed using thinking. When analyzing these analysis units, it is found that D1 is addressed with thinking by encouraging students to think about the set of data having specific properties.

There are about half of the analysis units presenting knowledge address meanings of data in both versions. The meaning of data dominantly addressed using knowledge in English version is data as numerical numbers with context (D1-C), that is, 40%. Analysis units presenting reasoning are used to address D2, while those presenting the combinations of statistical cognitions are used to address the three meanings of data, in which D3 has a higher percentage. In Indonesian version, the dominant meanings of data addressed by using knowledge are D1 and D1-C. The analysis units presenting reasoning are only 9% used to address the meaning of data, which is D1-C. The combinations of statistical cognitions are used to address the four types of meanings of data.

4. Conclusion
The textbook analysis conducted in this study has indicated several characteristics of Indonesian textbook which are different from English textbook. These characteristics included the different emphasis on statistical cognitions used to present the topics under data distributions. Indonesian textbook more likely emphasizes on a knowledge-based view of statistics, in which statistics are presented as the procedures for dealing with problems. This was inferred by the fact that the textbook presents basic statistical knowledge in almost 90% of analysis units and more dominantly address
data as numerical numbers with or without contexts. The English textbook presents both knowledge based and problem-solving views of statistics in almost similar percentages of analysis units. This was inferred from the fact that it presents almost similar percentages of basic statistical knowledge and reasoning. Besides, in contrary to Indonesian textbooks, it more dominantly addresses data as numbers in problem contexts (D2) or data as information for investigation (D3).

Since how the different characteristics of statistics textbooks may influence students’ values on the learning and conceptions of statistics can be one of future research topics in this area. Additionally, further research in analyzing and comparing more statistics books used in Indonesia might reveal the textbook signature of the country. Future studies, then could explore whether the statistics textbook signature has an influence on student learning and views on statistics.

References
[1] Garfield, J. B., & Ben-Zvi, D. (2008). Developing students’ statistical reasoning: Connecting research and teaching practice. Springer Science & Business Media.
[2] Garfield, J. B., & Franklin, C. (2011). Assessment of learning, for learning, and as learning in statistics education. In Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education (pp. 133–145). Springer.
[3] Idris, K., & Yang, K.-L. (2017). Development and Validation of an Instrument to Measure Indonesian Pre-service Teachers’ Conceptions of Statistics. The Asia-Pacific Education Researcher, 26(5), 281–290. https://doi.org/https://doi.org/10.1007/s40299-017-0348-z.
[4] Shuard, H., & Rothery, A. (1984). Children reading mathematics. London: John Murray.
[5] Yang, K.-L. (2016). Analyzing Mathematics Textbooks through a Constructive-Empirical Perspective on Abstraction: The Case of Pythagoras’ Theorem. Eurasia Journal of Mathematics, Science and Technology Education, 12(4), 913–930.