MIDDLE SCHOOL STUDENTS' STATISTICAL LITERACY: ROLE OF GRADE LEVEL AND GENDER

AYŞE YOLCU
Hacettepe University
ayseyolcu@hacettepe.edu.tr

ABSTRACT

This study examined the role of gender and grade level on middle school students' statistical literacy. The study was conducted in the spring semester of the 2012-2013 academic year with 598 middle-school students (grades 6–8) from three public schools in Turkey. The data were collected using the Statistical Literacy Test, developed based on Watson's (1997) statistical literacy framework. Two-way ANOVA results revealed no significant grade level differences although female students performed significantly better than male students. The spiral curriculum in middle school mathematics may explain the lack of differences between grades. The higher performance of female students may be related to the linguistic aspects of statistical literacy, in contrast to the situation in school mathematics.

Keywords: Statistics education research; Mathematics curriculum; Turkey

1. INTRODUCTION

Statistical literacy has become one of the important life skills for individuals (Gal, 2004). In order to contribute in an informed way to debates in areas from health care to politics, certain statistical knowledge is required (Carmichael, 2010). Such knowledge comprises understanding, interpreting and evaluating statistical claims, which can be referred to as statistical literacy (Watson, 1997). Although the importance of statistical literacy has been appreciated in statistics education research (e.g. Gal, 2004; Watson, 1997), the extent to which Turkish mathematics education could develop statistical literacy in students from each grade has not been investigated much. One of the few studies examining statistical literacy of Turkish students is limited to the 8th-grade and gives no indication of whether statistical literacy develops with respect to grade level (Yolcu, 2012). The current study investigates the role of grade level on the statistical literacy of Turkish middle-school students. Further, although gender has been considered an important issue in mathematics education (Leder, 1992), studies investigating the effect of gender on statistical literacy are rather scarce. The present study investigates the role of gender on statistical literacy of Turkish middle-school students from grades six to eight, where ages range between 13 and 15.

There are several definitions and conceptualizations of statistical literacy found in the literature. Basically, statistical literacy addresses the interpretation of statistical messages and communication with such messages through written or spoken word from a critical eye (Carmichael, 2010). In line with this description, Wallman (1993) indicated that statistical literacy includes both understanding and critically evaluating statistical messages that are necessary for daily lives of individuals and for their decision making, as she highlighted both for the personal and societal needs. Similarly, Gal's (2004) conceptualization also included the need for statistical literacy for active citizenship and pointed out that communication with statistics, and interpretation and judging of statistical claims are the key skills of statistically literate individuals.

This present study employs Watson's (1997) three-tiered statistical literacy framework, which contains common features of several definitions. Watson (1997) defined statistical

literacy as understanding, interpreting and evaluating statistical information that is confronted by individuals during their daily lives. She presents statistical literacy in a three-tiered framework:

Tier 1: The main feature of this tier is the familiarity with terminology used in statistical messages; for example, understanding the 'sample' as terminology is a typical characteristic of Tier 1.

Tier 2: The second tier is the interpretation of these statistical messages. To illustrate, Tier 2 contains application of ideas related to sample and sampling. Interpretation of statistical information that includes the concept of sample also lies in the second tier of statistical literacy.

Tier 3: The third tier consists of critical evaluation of statistical claims which can be biased or misleading. For instance, critical evaluation of the given sample and its generalizability to a particular population constitutes a Tier 3 aspect of statistical literacy within the context of this study.

2. MIDDLE-SCHOOL STUDENTS' STATISTICAL LITERACY

There are several studies that have examined statistical literacy of middle-school students. Watson and Callingham (2003) investigated the hierarchical nature of statistical literacy with school students from grades three to nine. They identified six hierarchical levels in line with Watson's (1997) three-tiered framework. Understanding terminology without engagement with context appeared in the first two levels, which were labeled 'idiosyncratic' and 'informal'. Calculations with little engagement with context appeared in the next two levels, labelled 'inconsistent' and 'non-critical'. Appropriate calculations with contextual engagement were the skills seen in the last two levels, 'critical' and 'critical-mathematical'. In this study, the majority of middle school students' responses placed them in 'informal', 'inconsistent' and consistent 'non-critical' levels a few students were in the 'critical' or 'critical-mathematical' levels.

Middle-school mathematics curriculum in Turkey aims to develop informed citizens who possess knowledge of statistics with an appreciation of the importance of the position of statistics in society (MoNE, 2005, 2013), but there are few research studies concerning Turkish middle-school students' statistical literacy. Yolcu (2012) investigated 8th-grade Turkish students' statistical literacy levels based on Watson's (1997) three-tiered framework. The sample consisted of 1074 8th-grade students randomly selected from urban schools in Turkey. The results revealed that students performed differently at each tier of statistical literacy. They performed lowest in the third tier of statistical literacy, where they were required to evaluate inappropriate statistical claims. Their performance was slightly higher in the first tier, which included understanding statistical terminology. However, their performance was the highest in the second tier, which was focused on interpreting statistical claims in context.

Since statistical literacy comprises different statistical content, such as average, sample, graph or chance, researchers have examined these specific aspects of statistical literacy in their studies. For example, Watson and Moritz (2000) investigated the understanding of the concept of sampling related to statistical literacy with middle-school students in grades 3, 6 and 9. The research instrument consisted of 11 items relating to the three-tiered framework for statistical literacy (Watson, 1997) and students' answers were evaluated as prestructural, unistructural, multistructural, or relational, depending on whether individuals showed non-statistical, single, multiple or interrelated statistical ideas. The results revealed that a developmental sequence existed in conceptualizing sampling as in the first and second tier of statistical literacy. That is, the performances in questions related to the first and second tier showed a correspondence with prestructural, unistructural, multistructural, and relational levels in terms of grade level of participants. However, performances of students in the third-tier tasks showed that questioning claims in the context of sampling was suitable for students who were not yet at the relational level. In addition to these results, to examine the longitudinal development of the sampling concept related to statistical literacy, they repeated the same study after two years and after four years with the same students. The two-year results indicated that 15% of the students performed at a lower level, 48% at a higher level and 37% at the same level as in their previous

assessment. However, within four years, only 7% of the students performed at a lower level, 24% of them did not change their level and 60% performed at a higher level. Performing at a lower level might be related to students' motivation, whereas performing at a higher level within four years indicated that understanding of the concept of sample might be developing gradually.

Aspects of statistical literacy related to interpretation of graphs was investigated to document differences regarding grade levels by Aoyama and Stephens (2003). Their study revealed that there is an identification of levels of graph interpretation from the lowest level A (basic reading of tables and graphs) to the highest level F (creating new dimensional information). The participants in this study were 55 students from grades 5 and 8. The results revealed that 95% of eighth-grade students and 82% of fifth-grade students could read beyond the data in the lower-level tasks. However, there was no appropriate response from both of the groups for the level-F task, which is consistent with the sixth level, 'critical-mathematical', of Watson and Callingham's (2003) statistical literacy levels. The reason for this result was that students lacked sufficient experience with evaluating statistical information in graphs, both in and out of school settings.

Another study conducted by Watson and Kelly (2008) investigating the literacy aspect of statistical literacy obtained similar findings. In this study, the researchers explored the vocabulary of statistical literacy, asking students about the meaning of terms such as 'sample', 'variation' and 'random' across grades. More explicitly, they asked students in grades 3 and 5 ($N = 359$) about the meaning of 'sample', and students in grades 7 and 9 ($N = 379$) about the meaning of 'sample', 'random' and 'variation'. For questions about 'sample', they found a significant difference between grades 3 and 5, with a medium effect size, and between grades 5 and 7, with a small effect size. However, there were no significant differences between grades 7 and 9 on any of the terms.

Beyond Turkey, various studies have considered middle-school students' statistical literacy and the role of grade level, but the results are inconsistent. Although students' understanding of sampling in the context of statistical literacy has been shown to develop gradually over two to four years (Watson & Moritz, 2000), there were no appreciable differences in the linguistic aspect of statistical literacy between adjacent grades (Watson & Kelly, 2008). Therefore, the current study contributes to existing literature regarding the question of whether development of statistical literacy corresponds to increase in grade level.

Considering teaching and learning mathematics, investigating the role of gender has become an international trend (Fennema, 2002). However, studies investigating the possible effect of gender on statistical literacy in middle school contexts are rather scarce and provide inconsistent results. Watson and Moritz (2000) examined the role of gender on the understanding of the concept of sampling as part of their study. In the lower grades, female students gave significantly higher-level responses than male students in Tier 2 questions. Male students tended to answer according to their contextual knowledge rather than by considering sample size. Similarly, in Tier 3 questions, significant gender differences were found in favor of female students. Despite these results, it is not possible to conclude whether females or males produce generally higher statistical literacy performances. The current study examining the role of gender on statistical literacy makes a significant contribution to statistical literacy research with pre-college students.

The purpose of this study is to examine the role of grade level and gender on middle-school students' statistical literacy. In line with these aims, it proposes the following research question: Is there a significant mean difference in middle-school students' statistical literacy scores with respect to grade level and gender?

3. METHODOLOGY

3.1. PARTICIPANTS AND CONTEXT

The participants of the study were 598 middle-school students from sixth to eighth grade from three public schools in a district in Ankara, the capital of Turkey. Table 1 shows the distribution of participants by grade and gender.

Table 1. Number of participants and percentages in terms of grade level and gender

Grade	Female	Male	Total
6th	110 (18.4%)	102 (17.1%)	212 (35.5%)
7th	106 (17.7%)	104 (17.4%)	210 (35.1%)
8th	88 (14.7%)	88 (14.7%)	176 (29.4%)
Overall	304 (50.8%)	294 (49.2%)	598 (100.0%)

Data were collected in the spring semester of the 2012–2013 academic year in participating students' classrooms. At the time of data collection, the Turkish education system had just gone through a major change, with elementary schools divided into elementary (grades 1 to 4) and middle (grades 5 to 8) schools. However, the new mathematics curriculum for grades 5 to 8 was implemented only in schools starting with the 2013–2014 academic year. Therefore, participants were 6th, 7th and 8th-grade students who were taught according to the previous curriculum and they are referred to as middle-school students in this study.

3.2. MIDDLE-SCHOOL MATHEMATICS CURRICULUM IN TURKEY

Turkish pre-college education is a centralized system with a national curriculum for all content areas at all grade levels, and with national examinations. Students who participated in this study have been taught using the same mathematics curriculum throughout elementary and middle school. Although the new middle-school curriculum (MoNE, 2013) had been prepared and published , it was being revised at the time of data collection, and so students were studying from the previous mathematics curriculum, MoNE (2005). This middle-school mathematics curriculum addressed sampling, average, probability, variation, and tables and graphs in grades 6 to 8 at the time of the study (MoNE, 2005). Therefore, based on the centralized nature of the national curriculum, participating students in this study had been taught these statistical topics, as well as basic informal inferential topics such as comparing data sets. The objectives and topics for each grade are presented in Table 2.

Table 2 shows that the main content for the statistics and probability curriculum was the same across grades, though each grade level had different objectives. The distribution of content indicates the spiral nature of the elementary mathematics curriculum. For example, students learn about tables and graphs in each grade; however, the topics differ. Students study bar graphs in 6th grade, line graphs and pie charts in 7th grade, and histograms in 8th grade. In the change beginning from the 2013–2014 academic year, the time spent on probability concepts was reduced, leaving only the objective concerning 'probability of an event'. In the statistics and data analysis section, 'measures of spread' have been kept, though the concept of standard deviation has been removed, since it is considered that it includes overly-complicated calculations.

3.3. STATISTICAL LITERACY TEST

The Statistical Literacy Test (SLT, see Appendix) was initially designed to investigate 8th-grade students' statistical literacy based on Watson's (1997) three-tiered statistical literacy framework. It consists of 17 questions and sub-questions, developed according to a table of specifications based on three tiers of statistical literacy. This test contained both multiple-choice and open-ended questions. In the selection of questions the researcher consulted

textbooks, curriculum documents, a variety of statistics books, and the existing literature; for detailed explanation of the test development see Yolcu (2012). These questions were based on concepts of sample, average, graphs and tables, probability, inference, and variation, in parallel with the contents of the Turkish mathematics curriculum.

Validity of the SLT was assessed by gathering expert opinions as to whether the items were consistent and appropriate in terms of the three-tiered statistical literacy framework, content, curriculum, and context. Then a pilot study was conducted to measure reliability of the scale and the difficulty level of items. Students' responses for open-ended items were evaluated using a holistic rubric prepared according to the pilot study and related literature. The coding scheme for the open-ended items is given in Table 3.

Table 2. Statistics and probability content area in Turkish school mathematics

Grade	Topic	Content
6th	Identification of likelihood of situations	Basic principles of counting
	Basic concepts of probability	The concepts of experiment, outcome, sample space, random Probability of an event
	Types of events	Impossible and certain events Complementary events
	Types of probability	-N/A
	Formulation of problems for research and data collection	Research problem formulation Choosing appropriate sample Data collection
	Table and graphs	Bar graphs and their misinterpretation
	Central tendency and spread	Arithmetic mean, range
7th	Identification of likelihood of situations	Permutations
	Basic concepts of probability	N/A
	Types of events	Discrete and non-discrete events
	Types of probability	Calculation of probability with geometric knowledge
	Formulation of problems for research and data collection	N/A
	Table and graphs	Bar graphs, line graphs, pie charts
	Central tendency and spread	Median, mode, interquartile range
8th	Identification of likelihood of situations	Combinations and permutations
	Basic concepts of probability	N/A
	Types of events	Dependent and independent events
	Types of probability	Experimental, theoretical, and subjective probability
	Formulation of problems for research and data collection	Research problem formulation for comparing two data sets and for the sample given
	Table and graphs	Histogram
	Central tendency and spread	Standard deviation

Table 3. Coding scheme for open-ended items

Response Types for open-ended items	Classification of Responses	Score
Incorrect or context-based responses	Non-statistical/incorrect	0
Correct, without appropriate statistical explanations	Pre-statistical	1
Correct, with appropriate statistical explanations	Statistical	2

This holistic rubric provides a framework for evaluation of open-ended items. The codes in the rubric are explained below in more detail.

Non-statistical/incorrect An incorrect response includes both errors or mistakes and also inaccurate responses based on particular context that the item has. For example, responses such as "Smoking is detrimental to our health" (for SLT question 15) or "Children should not watch TV for three hours" (for SLT question 10) are counted as context-based incorrect responses.

Pre-statistical This category includes a correct answer to the particular item though with an absence of appropriate statistical explanation. An example would be responses that recognize that there is an outlier in the data set, though students cannot explain why this outlier should not be used in the arithmetic average, or that in this case the median should have been used instead of the mean.

Statistical A correct, statistical response is an indicator of statistical literacy. That is, students understand the statistical information and are able to critically evaluate it. In addition, they are able to provide statistical justifications or explanations of their understanding, interpretation, or evaluation. For example, responses that state whether or not a sample is generalizable to a particular population, giving statistical reasons such as size or representativeness of the sample.

In the pilot study of the SLT, Cronbach's alpha coefficient was 0.72; further, the interclass correlation coefficient (ICC) for two raters was calculated as 0.87 for open-ended items (Yolcu, 2012).These values indicate that the SLT is a valid and reliable test.

For this current study, the items in the SLT remained the same except for item 11. In this item, the term 'standard deviation' was used in the original test for 8th-grade students. Here, the term 'standard deviation' was replaced by 'measure of spread', which was also familiar to 6th and 7th-grade students. The final English version of the SLT (translated by the author) which was used in this study can be found in the Appendix. In this study, the reliability coefficient Cronbach's alpha was calculated as 0.74. Approximately 10% of randomly-selected cases were scored by another mathematics education researcher and the ICC was found as 0.83, which indicated quite high consistency between scorers (Shrout & Fleiss, 1979).

4. RESULTS

Since there are different points assigned to each item of the SLT depending on the format (multiple choice or open-ended), students' total scores were obtained by dividing their scores by the maximum possible score, resulting in scores in the range between 0 and 1. The means (M) and standard deviations (SD) of statistical literacy scores by grade level and gender are given in Table 4. This shows that statistical literacy scores of middle-school students were generally low (with an overall mean of 0.32), with very little difference between grades 6, 7 and 8, but indicating that female students performed better than males.

Table 4. Descriptive statistics of SLT scores with respect to grade level

| Grade | Statistical Literacy Scores: Mean (SD) | | | Sample Size (female; male) |
	Female	Male	Total	
6th	0.35 (0.13)	0.30 (0.12)	0.33 (0.13)	212 (110; 102)
7th	0.36 (0.15)	0.29 (0.15)	0.32 (0.15)	210 (106; 104)
8th	0.33 (0.15)	0.31 (0.14)	0.32 (0.15)	176 (88; 88)
Total	0.35 (0.14)	0.30 (0.14)	0.32 (0.14)	598 (304; 294)

A two-way analysis of variance was conducted to explore the impact of grade level and gender on total statistical literacy scores. Preliminary analyses were performed to ensure that assumptions including independence of observations, normal distribution for each group, and homogeneity of variance were satisfied. The interaction between gender and grade level was not statistically significant, $F(2, 592) = 2.03$, $p = 0.13$, indicating that there was no significant difference in the effect of grade on statistical literacy scores for female and male students (see Figure 1).

Since the interaction effect was not statistically significant, the main effects of grade level and gender were examined separately. There was a statistically significant main effect for gender, $F(1, 592) = 15.20$, $p < 0.01$, in favor of female students (mean 0.35) compared to male students (mean 0.30). However, the effect size was interpreted as small using Cohen's (1988) guidelines, as the partial eta-squared was 0.03, indicating small practical significance. The main effect for grade level, $F(2, 592) = 0.04$, $p = 0.97$, did not reach statistical significance. In other words, no significant difference was observed in terms of grade level of middle-school students' statistical literacy scores.
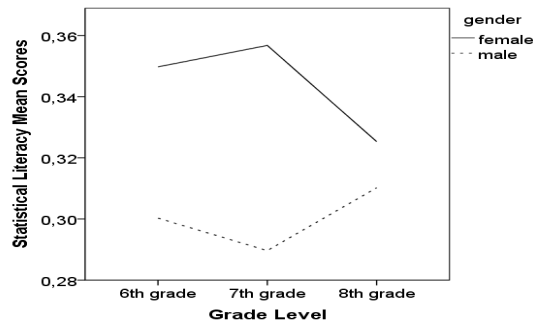

Figure 1. Interaction between Grade level and Gender on Statistical Literacy Mean Scores.

5. DISCUSSION

The results of this study showed that Turkish middle-school students' statistical literacy performances, in which they were required to understand, interpret and evaluate statistical information, could be interpreted as low in each grade level. This is despite the stated aims of national curriculum, and is consistent with a previous study of eighth-grade Turkish students (Yolcu, 2012). There were non-significant mean differences between sixth-, seventh-, and eighth-grade students' statistical literacy scores. Female students had significantly higher scores than male students, yet the effect size was small.

Several studies in the literature indicated that there is a difference in statistical literacy between grades and that higher-grade students performed better in a various aspects of statistical literacy (e.g., Aoyama & Stephens, 2003; Watson & Kelly, 2008). However, in Watson and Kelly's (2008) study, where they examined the definitions of statistical concepts across grades, either there were no significant differences or differences with small effect size between adjacent grades. Similarly, Watson and Moritz (2003) could not find any grade-level differences between adjacent grades in their investigation of linguistic aspects of the interpretation and evaluation of chance. This indicates that students in close grade levels will show similar outcomes in terms of statistical literacy, in line with the results of this study. Middle-school students involved in the current study were from grades 6 to 8, and the closeness of these grades might be the reason for the similar results in their statistical literacy scores.

The spiral nature of the Turkish middle-school mathematics curriculum, in which the same content domains are presented at each grade level, might explain the similar statistical literacy scores for grades 6 to 8. Since all middle-school students become familiar with graphs and other statistical concepts, similar results for statistical literacy at each grade are likely to be observed.

Furthermore, students in Turkey have to take a national placement examination for high schools in the 8th grade. As they approach this time, students prepare for this examination and they become more focused on application of mathematical or statistical ideas rather than evaluation of them. This may help to explain why students' overall statistical literacy results vary little between grades 6 and 8.

The results of this study in terms of gender indicated that female students performed better in statistical literacy. Gender has been an important issue in mathematics education and

research reports have indicated that boys have higher mathematics achievement than girls (Leder, 1992); the higher performance of female students in this study is considered noteworthy. The reason for this result might be found in the distinction between the domains of mathematics and statistical literacy; the latter involves a greater linguistic aspect (Gal, 2004; Watson, 2006) that may explain previously reported lack of gender differences in predicting interest in and self-efficacy for statistical literacy (Carmichael et al., 2010). Moreover, a study with Australian children which investigated gender differences on mathematics achievement and teacher ratings on various content domains of mathematics found no significant difference between female and male students in terms of mathematics achievement; however, teachers rated girls at higher levels in tasks involving data, while they rated boys at higher level in tasks involving place value and computation (Carmichael, 2013). This suggests that girls may outperform boys in the topics of data and chance.

Higher performance of female students in the area of statistical literacy can also be found in the literature. For example, Watson and Moritz (2000) examined the possible gender role on understanding of the concept of sample. The results revealed that at lower grade levels, female students had higher level responses than male students, who gave more wrong responses on the basis of their contextual knowledge. This is in agreement with the findings of the current study. Other empirical studies conducted with Turkish students have indicated that there is no difference between female and male students in terms of mathematics achievement (e.g., Bulut, Gür, & Sriraman, 2010). Furthermore, a recent meta-analysis indicated that there are no differences between the mathematical achievement of male and female students (Lindberg, Hyde, Petersen & Linn, 2010). Therefore, the role of gender in statistical literacy in favor of girls is considered to be consistent with the statistics education literature, more recent research in mathematics education and with Turkish findings.

6.   CONCLUSIONS AND IMPLICATIONS

The results of this study could contribute to curriculum development in the Turkish middle-school curriculum, as objectives might be modified in relation to support for statistical literacy, as previously noted in Yolcu's earlier (2012) study. For example, objectives regarding evaluation of statistical claims might be added to the middle-school mathematics curriculum so that students would have an opportunity to examine biased statistical claims in a critical way. However, further research on statistical literacy is needed and qualitative research on how middle-school students critically evaluate statistical claims would be an important contribution to the statistics education literature.

The lack of grade level difference in statistical literacy scores of middle-school students in this study did not indicate there was no developmental sequence. The students in this study were from adjacent grades and of similar ages. More research should be carried out using a wider range of grade levels to document whether or not there is a developmental hierarchy in statistical literacy scores.

The results of this study regarding similar statistical literacy scores in grades 6 to 8 could be attributed to statistics instruction in middle schools and maybe to the pressures of national examinations. It seems that the majority of class time in higher grades was devoted to the application of statistical ideas, leaving little time for opportunities to develop conceptual understanding and critical evaluation. Instead, a teaching approach that included project work together with real life data and media reports could be employed in order to increase the statistical literacy of middle-school students (Merriman, 2006). Teachers might incorporate daily news, including statistical reports appearing in the media to the statistics lessons. In addition, there needs to be more integration of context into statistics teaching. Essentially, statistical literacy deals with data in context and plays the role of a bridge between statistics and everyday life.

Technological tools should also be integrated into statistics instruction in schools so that students might handle the procedural aspects of application of statistical ideas easily and focus more on conceptual understanding. Regarding critical evaluation of statistical claims, a relevant study where critical thinking skills were emphasized during statistics instruction reported

enhanced statistical conceptions of participants (Doyle, 2008). A similar critical-thinking approach might be employed in classrooms. Teachers should be aware that statistical claims appearing in the media may be one-sided, biased, or misleading. A pedagogical approach where students reflect, discuss, and evaluate statistical claims rather than accepting them without question may improve outcomes.

The results of this study in terms of gender indicated that more research is needed to understand why and how this gap occurs in favor of female students. However, in classroom settings while teaching statistical concepts, a diverse range of contexts and examples should be considered in order to engage both female and male students with the statistical content.

In this study, which aimed at investigating role of gender and grade level on statistical literacy of middle school students, there were several extraneous variables such as students' backgrounds and previous mathematical knowledge which were not controlled for. Furthermore, the study was conducted only in one district of Ankara, Turkey, and therefore, the findings of this study might be limited in its application to a more general population of middle-school students. Yet, the results can be generalized to students whose context and curriculum is similar to this study. Another limitation was that the results of the present study were based on quantitative data collected from participants through the SLT, and so the study might be limited by the representativeness of the items on the Statistical Literacy Test.

REFERENCES

Aoyama, K., & Stephens, M. (2003). Graph interpretation aspect of statistical literacy: a Japanese perspective. *Mathematics Education Research Journal, 15*(3), 207–225.

Bulut, S., Gür, B. S., & Sriraman, B. (2010). Commentary 2 on feminist pedagogy and mathematics. In B. Sriraman & L. English (Eds.), *Theories of mathematics education: Seeking new frontiers* (pp. 455–466). Heidelberg: Springer.

Carmichael, C. S. (2010). *The development of middle school children's interest in statistical literacy.* Unpublished Doctoral Dissertation, University of Tasmania, Tasmania.

Carmichael, C. S. (2013). Gender differences in children's mathematics achievement: Perspectives from the longitudinal study of Australian children. In V. Steinle, L. Ball, & C. Bardini (Eds.), *Proceedings of the 36th annual conference of the Mathematics Education Research Group of Australia* (pp. 130–137). Melbourne, VIC: MERGA.
[Online: http://www.merga.net.au/documents/Carmichael_MERGA36-2013.pdf]

Carmichael, C., Callingham, R., Hay, I., & Watson, J. (2010). Statistical literacy in the middle school: the relationship between interest, self-efficacy and prior mathematics achievement. *Australian Journal of Educational & Developmental Psychology*, *10*, 83–93.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Doyle, P. (2008). *Developing statistical literacy with students and teachers in the secondary mathematics classroom*. Unpublished Masters Thesis, The University of Waikato.

Fennema, E. (2002). Mathematics, gender, and research. In G. Hanna (Ed.), *Towards gender equity in mathematics education* (pp. 9–26). New York: Kluwer Academic.

Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht: Kluwer Academic Publishers.

Leder, G. C. (1992). Mathematics and gender: changing perspectives. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning.* New York: Macmillan.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin*, *136*(6), 1123–1135.

Merriman, L. (2006). Using media reports to develop statistical literacy in year 10 students. *Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute.
[Online: http://iase-web.org/documents/papers/icots7/8A3_MERR.pdf]

Ministry of National Education [MoNE]. (2005). *İlköğretim Okulu Ders Programları: Matematik Programı 6-7-8 (Elementary Curricula Programs: Mathematics Curricula Program for Middle Grades)*. Ankara: MEB.

Ministry of National Education [MoNE]. (2013). *Ortaokul Ders Programları: Matematik Programı 5-6-7-8 (Middle School Curricula Programs: Mathematics Curricula Program for Middle Grades)*. Ankara: MEB.

Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlation: Uses in assessing rater reliability. *Psychological Bulletin*, *86*(2), 420–428.

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association, 88*(421), 1–8.

Watson, J. (1997). Assessing statistical thinking using media. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107–122). Amsterdam: IOS Press.

Watson, J. (2006). *Statistical literacy at school: Growth and goals*. Mahwah, NJ: Lawrence Erlbaum.

Watson, J., & Callingham, R. (2003). Statistical literacy: a complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3–46.
[Online: http://iase-web.org/documents/SERJ/SERJ2(2)_Watson_Callingham.pdf]

Watson, J., & Kelly, B. A. (2008).The vocabulary of statistical literacy. *International Journal of Science and Mathematics Education*, *6*(4), 741–767

Watson, J., & Moritz, J. (2000). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior*, *19*(1), 109–136.

Watson, J., & Moritz, J. (2003). The development of comprehension of chance language: evaluation and interpretation. *School Science and Mathematics*, *103*(2), 65–80.

Yolcu, A. (2012). *An investigation of eighth grade students' statistical literacy, attitudes towards statistics and their relationship.* Unpublished Masters Thesis, Middle East Technical University, Ankara.

AYŞE YOLCU
Hacettepe University
Faculty of Education, Elementary Education
06800 Beytepe Ankara TURKEY

APPENDIX: STATISTICAL LITERACY TEST
translated from Turkish by the author;
asterisks (*) indicate correct answers in the multiple-choice questions

1. "A study is conducted where the sample is mathematics teachers worked in Ankara". What do you understand by the word "sample" in this sentence?

2a. "Last year, an average of 20 people died each month due to traffic accidents." What do you understand by the word "average" in this sentence?

2b. Which one of the following is *not* a method for finding that an average of 20 people died each month due to traffic accidents last year?
(I) Add the number of people who died in each month and then divide by 12.
(II) Put the number of people who died each month in numerical order and then choose the one in the middle number.
(III) Find the most frequent number of people who died each month of the year in traffic accidents.
(IV) Subtract the smallest number of people who died in traffic accidents each month of the year from the largest number.*

3. The data below represents what a 5 TL (Turkish Lira) lunch includes, and the price of each item: 2 TL main meal, 0.5 TL soup, 1.5 TL desert, 1 TL salad
Which one of the following graph types would best represent this information?
a) Pie Chart*   b) Histogram   c) Line Graph   d) Bar Graph

4. Which of the following is/are random selections?
(I) Selection of red marbles after putting them in a bag and mixed
(II) Selection of any two marbles after putting them in a bag and mixed
(III) Selection of every fifth marble without putting in a bag
a) Only I   b) Only II*   c) I and II   d) I and III

5. Which of the following data sets is most variable? Provide your answer without calculation.
   a)   10, 11, 12, 13, 14, 15 *
   b)   13, 13, 13, 13, 13, 13
   c)   11, 12, 12, 13, 13, 14
   d)   10, 12.5, 12.5, 12.5, 12.5

6. Ali is a member of library club in an elementary school and he wants to investigate the number of books at students' homes. Which one of the following might be a representative sample of the students from his school for this research?
   (a) 30 students chosen randomly from the library club
   (b) 30 students chosen randomly from the whole school *
   (c) 30 students chosen randomly from Ali's class
   (d) 30 female students chosen randomly from the whole school

7. A researcher who lives in a town consisting of 50 families has found the mean of children per family as 2.2. Which one of the following must be true?
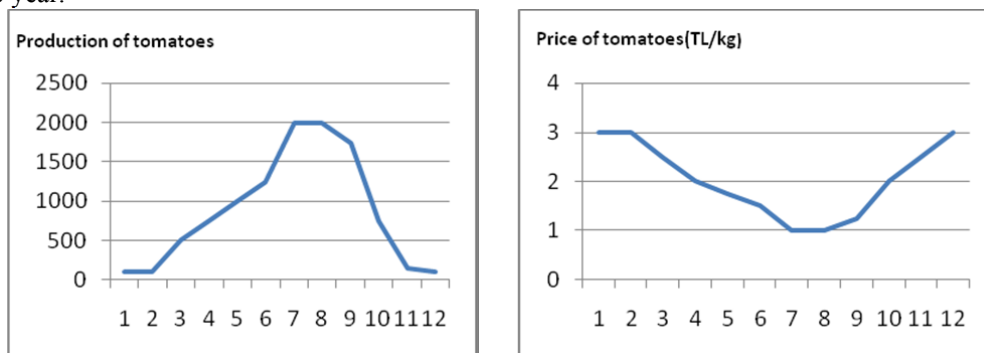   (a) Half of the families in this town have two children.
   (b) There are more families with 3 children than families with 2 children.
   (c) There are 110 children in this town.*
   (d) The mean number of children per adult is 2.2.

8. The graphs below represent production of tomatoes and price per kg for each month of the year.

(I) The price of tomatoes is less in summer, more in winter.
(II) The production of tomatoes is more in summer, less in winter.
(III) Since production of tomatoes is less in winter, the price is high.
According to the graphs, which of the statements can be inferred?
a) Only I   b) I and II   c) II and III   d) I, II and III *

9. The following message is printed on a bottle of skin cream: "WARNING: For application to skin areas there is a 15% chance of developing rash. If a rash develops, consult your doctor."Which of the following is the best interpretation of this warning?
(a) About 15 of 100 people who use this medication develop a rash.*
(b) If a rash develops, it involves only 15% of the skin.
(c) There is hardly any chance of getting a rash using this medication.
(d) If you use this cream, apply it to only 15% of your skin.

10. The weight of a baby for each month from birth is shown in the table below.

Age (month)	Weight (kg)
0	3.5
1 month	4.5
2 month	5
3 month	6
4 month	7
5 month	7.5
6 month	8

According to this, predict that how many kg the baby will weigh at the end of the seventh month. Explain how you obtained your answer.

11. Some statistics regarding the mathematics grades for classes8A and 8B in an elementary school are presented in the table below.

	Arithmetic mean	Measure of spread
Class 8A	80	5.2
Class 8B	76	3.5

Which of the following is/are true?
(I) Looking at the arithmetic mean, the grades in class8A are higher than in class 8B.
(II) Looking at the measure of spread, the variation in class 8B is smaller.
(III) Looking at the measure of spread, the variation in class 8A is smaller.
a) Only I   b) Only II   c) I and II *   d) I and III

12. The sample for a study investigating how many hours children watch TV consisted of 5th grade students in School A. The results showed that students who participated watched TV for an average of 3 hours in a day. The results of the study were announced as follows:

"All elementary school students in Turkey watch TV for 3 hours a day."

Do you find this sentence an acceptable statistical claim? Provide a statistical explanation for your answer.

13. The number of problems solved in a mathematics class is counted and represented in the following table.

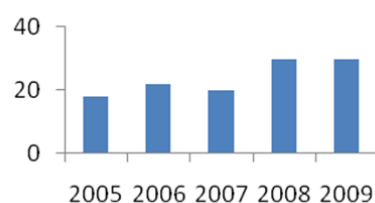| Student	Number of problems solved
A	2
B	6
C	2
D	22
E	3
F	2
G	1
H	2

In order to summarize these data, the mean is calculated and found to be 5.Do you agree with this? Give reasons for your answer.

14. An announcer showed these graphs and stated "Although the numbers of audience members in theatre and cinema differ before 2009, they are almost equal in 2009".
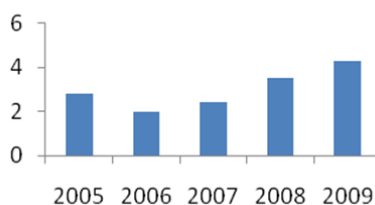
Do you think that the claim announced is acceptable? Provide a statistical explanation for your answer.

15. A study found that those who smoked a pack of cigarettes a day for less than 49 years doubled the risk of premature wrinkling, while for those who smoked for more than 50 years the risk was 4.7 times greater compared to those who do not smoke. The table below summarizes this information:

	Less than 49 years	More than 50 years
Risk of non smokers	A	B
Risk of smokers	2A	4.7B

Is the result of this report acceptable? What kind of questions would you ask to examine the validity of the report?

16. The following information is from a survey about smoking and lung disease among 250 people.

	Lung Disease	No lung disease	Total
Smokers	90	60	150
Nonsmokers	60	40	100
Total	150	100	250

Using this information, a researcher states that "The reason for lung disease is smoking." Do you think that this claim is acceptable? Explain your answer statistically.

17. A group of students noted the highest daily temperature in Ankara during the course of one year. They found that the average highest daily temperature in Ankara was 16°C. Three other students predicted the maximum temperature in Ankara for six different days in a year.

Students	Predicted Temperature
Seda	16, 35, 1, 5, 29, 10
Zeynep	16, 16, 16, 16, 16, 16
Umut	16, 15, 14, 26, 8, 17

Which student gave predicted temperatures with the most realistic variability, Seda, Zeynep or Umut? Explain your answer.