Athens Institute for Education and Research
ATINER

ATINER's Conference Paper Series
STA2016-2126

Teaching College Probability for Higher Achievement

Irini Papaieronymou
Lecturer
Michigan State Univeristy/ University of Nicosia
USA / Cuprus

An Introduction to ATINER's Conference Paper Series

ATINER started to publish this conference papers series in 2012. It includes only the papers submitted for publication after they were presented at one of the conferences organized by our Institute every year. This paper has been peer reviewed by at least two academic members of ATINER.

Dr. Gregory T. Papanikos
President
Athens Institute for Education and Research

This paper should be cited as follows:

Papaieronymou, I., (2016). **" Teaching College Probability for Higher Achievement",** Athens: ATINER'S Conference Paper Series, No: **STA2016-2126.**

Athens Institute for Education and Research
8 Valaoritou Street, Kolonaki, 10671 Athens, Greece
Tel: + 30 210 3634210 Fax: + 30 210 3634209 Email: info@atiner.gr URL: www.atiner.gr
URL Conference Papers Series: www.atiner.gr/papers.htm
Printed in Athens, Greece by the Athens Institute for Education and Research.
ISSN: 2241-2891
09/02/2017

Teaching College Probability for Higher Achievement

Irini Papaieronymou
Lecturer
Michigan State Univeristy/ University of Nicosia
USA / Cuprus

Abstract

This paper presents the results of a study which examined the role of particular tasks implemented through two instructional methods on college students' achievement in probability. A mixed methods design that utilized a pre-test (with multiple-choice items) and post-test (with multiple-choice and open-ended items) in treatment and control groups in an introductory statistics course was used. An analysis of pre-test scores indicated that students in the control and treatment groups had comparable initial probability knowledge. The results of the Wilcoxon Signed-Ranks test indicated that the multiple-choice scores of students in the control group were significantly lower on the post-test compared to the pre-test. In the case of the treatment group, student scores on the multiple-choice items did not increase significantly from the pre-test to the post-test. In addition, the results of the Mann-Whitney test showed that the normalized gain scores of the treatment group were significantly different from those of the control group. Relative to the open-ended post-test items, the Mann-Whitney test indicated that the scores of the treatment group on these items were significantly higher than the scores of the control group. Overall, the total post-test achievement of students in the treatment group was significantly higher than that of students in the control group.

Keywords: achievement, college, instructional methods, probability

Acknowledgments: The results presented in this paper form part of the dissertation study of the author. Special thanks go to the author's dissertation committee - Dr Sharon Senk (chair) and Dr Vince Melfi, Dr Jennifer Kaplan, Dr Jack Smith (members) - for their guidance throughout the study.

Introduction

Over the past couple of decades there have been various reform initiatives concerning the content and means of instruction in mathematics classrooms at the college level. In particular, the Mathematical Association of America (MAA, 1998) has supported an increase in the importance placed on probability at the post-secondary level and the use of more active learning methods. To this end, it is recommended that statistics instruction relies less on lecturing and more on active learning that uses group problem-solving, activities and discussions (Cobb, 2000). Given that curriculum reform has brought data handling to the forefront, less emphasis should be placed on "formal" probability and "an empirical frequency-based approach to probability that is also an important foundation for later work in theoretical probability" should be used (Watson, 2006, p.127).

With regards to Cyprus – where this study took place - the results of TIMSS 2007 revealed that at the 8th grade level only 3% of class time was devoted to data and chance (Mullis, Martin and Foy, 2008). Topics relative to this domain are considered to be for the more able students and as such, only 3% of Cypriot 8th graders receive formal instruction in this domain; that is, this 3% of Cypriot students receive the 3% of instructional time devoted to data and chance. Given the importance of probability in everyday life and its recommended increased focus in school curricula, it is encouraging that the Ministry of Education and Culture in Cyprus (MoEC) has recently initiated educational reform efforts which include revisions to the national curriculum in grades K-12 (MoEC, 2008). In 2010, reports published by the MoEC of Cyprus regarding the intended curriculum in mathematics, introduced probability at the elementary school level with increased emphasis as students move through the grades.

In the last couple of decades a considerable number of studies have been carried out on the teaching and learning of probability (Jones, Langrall & Mooney, 2007). However, most of this research relates to students' thinking of probability with a minimal amount having been carried out relative to instructional methods (Hirsch & O'Donnell, 2001). In addition, Shaughnessy (1992) indicated a lack of research in probability learning and teaching outside of western countries.

Purpose and Research Questions

With the above issues under consideration, this study examined the role of particular tasks on college students' achievement in probability. The research questions addressed in this study involved the use of two instructional methods:

Instructional Method A: Using lectures and small-group cooperative learning sessions during which students solve probability problems from the course textbook and

Instructional Method B: Using lectures and small-group cooperative learning sessions during which students complete activities involving probability experiments that generate real data in order to make connections between experimental and theoretical probability.

The research questions addressed were: What are the effects of using each of these instructional methods on college students' achievement in probability? Does Instructional Method B have a better effect on college students' achievement in probability than Instructional Method A?

Literature Review

Cooperative Learning and Student Achievement

Cooperative learning is defined as a "structured, systematic instructional strategy" (Cooper & Mueck, 1990, p. 68) in which students are assigned specific roles into small groups of two to ten students and work towards a common goal while being responsible for their own learning. Research identified five elements of cooperative learning groups: i) *positive interdependence*; students perceive that they must make a joint effort and that each member has a unique contribution they can make to the group (Johnson & Johnson, 1985; Effandi and Zanaton, 2007); ii) *individual and group accountability*; the group is accountable for achieving the set goal and each member is accountable for making a contribution to the group (Johnson & Johnson, 1985). Individual accountability may be structured by giving individual exams (Effandi and Zanaton, 2007); iii) *face-to-face interactions* during which students help each other complete a task, explain orally to one another how to solve problems and discuss the strategies used (Smith, Douglas and Cox, 2009); iv) *teamwork skills* which may be introduced by assigning students different roles in their groups (Smith, Douglas and Cox, 2009) and v) *group processing*; instructors should provide students with a specific task which is complex enough to warrant a group, and provide sufficient time for them to work in groups.

Research points towards the benefits of instructional methods that use cooperation among multiple students (Franklin & Garfield, 2006; Zieffler et al., 2008). In particular, student achievement may be improved through cognitively demanding tasks set by teachers (Giraud, 1997; Potthast, 1999). Dees (1991) found that students in a college remedial mathematics course performed the same under the cooperative learning and traditional methods of instruction on algebra tasks that did not involve complex thinking. However, students using the cooperative learning approach performed better on measures that tested higher cognitive skills. Meta-analyses indicated that "the average student taught by cooperative learning performs better than the average student taught with competitive and individualistic methods" (Johnson & Johnson, 1989; as cited in Potthast, 1999).

Experiments with Real Data

The statistics education community recommends that students in statistics classrooms have access to real data as well as experiences collecting, analyzing and using such data. The American Statistical Association (ASA) indicates that real data comes in various forms: archival data, data generated in the classroom, and data generated through simulations (Aliaga et al., 2005; Franklin & Garfield, 2006). Among the benefits of using such data sets is that they allow students to appreciate the difference between empirical and theoretical approaches to explaining and predicting phenomena under conditions of uncertainty (Batanero & Sanchez, 2005).

Research by Shaughnessy (1977) with college undergraduate students in an elementary probability and statistics course indicated that "college students can learn to discover some elementary probability models and formulas while working on probability experiments in small groups" (p. 313-314). In his work, Steinbring (1984; 1991; see Jones and Thornton, 2005) examined probability both from its empirical and theoretical forms and emphasized the connection between the two. Due to this connection, he supported that probability instruction should be carried out as follows:

> … learning begins with personal judgments about a random situation; comparisons are made between the empirical situation and conjectured theoretical models, and finally these comparisons lead to generalizations and more precise characterizations of the random situation (Jones and Thornton, 2005, p. 78-79).

In a study carried out by Amit and Jan (2006) with high-achieving students in grades 6-9 in Israel, game tasks were used in the acquisition of probability concepts. Participants had no formal background in probability and no formal teaching intervention was implemented during the study. As a result of actively participating in small groups during games, students gained insight that there is a difference between theoretical and experimental probability and that a link exists between probability and sample size.

Methods

In this study, participants included 44 students in an introductory statistics course taught by the researcher at a college in Cyprus. Twenty of them were male and twenty-four were female. Most of these students (95%) had attended a public secondary school in Cyprus and most of them (91%) were native Cypriots. Thirty six (82%) of the participants indicated that they had received some instruction on probability in high school. All of the participants were able to communicate orally in Modern Greek, the native language in Cyprus, and 93% of them were using Modern Greek as their first language and English as their second language.

A mixed methods design was used in this study. This included a treatment group which was taught using Instructional Method B and a control group which was taught using Instructional Method A. Formal instruction on probability occurred during the second half of the semester and lasted for seven weeks.

The placement of students in course sections was decided upon by the course coordinator in consultation with the academic board at the college based on students' English language proficiency. As a result, the treatment group included 27 students of moderate-to-high English language proficiency whereas the control group included 17 students of low-to-moderate English language proficiency.

The instruments used in this study included a background questionnaire, a pre-test comprising of 14 multiple-choice items and a post-test comprising of 15 multiple-choice items and 2 open-ended items. The pre-test comprised of a set of multiple-choice items from the web ARTIST (Assessment Resource Tools for Improving Statistical Thinking) project and in particular from the Probability Scale created by the ARTIST investigators (delMas et al., 2006; Garfield et al., 2006), as well as multiple-choice items on probability from the TIMSS studies for grade 8 (TIMSS & PIRLS International Study Center, 1995, 2001, 2007, 2009). All items on the pre-test were embedded on the post-test.

During the unit on probability students in the treatment group worked in small groups during class to complete four activities which covered the following topics: i) Law of Large Numbers; ii) Conditional probability and independence; iii) Discrete probability distributions; and iv) the Binomial distribution. The activities aimed to bring about the bidirectional relationship between experimental and theoretical probability and to prepare students for the study of theoretical probability. Each activity required the students to perform a probability experiment, to collect and analyze data and reach some conclusions. An attempt was made to include in each group students of varying mathematical ability in order to create opportunities for scaffolding (Vygotsky, 1978; as cited in Giraud, 1997). During this time period, students in the control group worked in small groups on four sets of probability problems assigned from the course textbook. The problems covered the following topics: i) basic concepts of probability; ii) conditional probability and independence; iii) discrete probability distributions and iv) the Binomial distribution.

The instructional sequence followed in the control group was different to that used in the treatment group. In the control group, the study of a probability topic began by looking at the connection between the probability topics previously learned and the probability topic to be studied that week. This was followed by a study of theoretical probability on the topic using a lecture format which included looking at the bidirectional relationship between experimental and theoretical probability as a whole class using examples from the textbook. Following the above instructional sequence, the instructor solved a few probability problems from the textbook on the particular topic. Then, students were placed in small groups and were asked to solve problems from the textbook. In the treatment group students first worked in small groups on a probability activity prior to formal instruction on a topic. In both the treatment and control groups, students were assigned 2-3 problems from the textbook to work on individually as homework. These were corrected by the instructor and handed back to the students on the next class session.

In this study, cooperative learning was used as part of instruction in both the treatment group and the control group while students worked in small groups. In both the treatment and control groups, students were assigned specific roles and received a shared grade for work completed in groups. These roles were rotated between small group sessions. Groups received feedback on their performance and a shared score for the group work submitted to the instructor but were examined individually on the post-test.

Results and Discussion

Initial Equivalence of Groups

In order to investigate whether the two instructional methods under consideration had a significant effect on students' achievement in probability, it was necessary to compare the initial probability knowledge of students in the two groups based on pre-test scores. The Kolmogorov-Smirnov and the Shapiro-Wilk tests for normality resulted in p-values lower than 0.05 (0.005 and 0.023 respectively) meaning that the distribution of the overall sample was significantly different from a normal distribution. The Kruskal-Wallis non-parametric test for two samples showed that participants in the control group had comparable initial probability knowledge to participants in the treatment group ($p = 0.37 > 0.05$).

Effects of Instructional Treatment on Students' Achievement

Comparison of Gain Scores on Multiple-choice Items

In the comparison of gain scores, only the 14 multiple-choice (MC) items that were common to the pre-test and post-test were considered. The median pre-test score was 9 (out of 14) in both groups. However, the median post-test score in the control group decreased by 3 points whereas in the treatment group it increased by 1 point.

Table 1. *Changes in Descriptive Statistics from Pre-Test to Post-Test*

Test	Statistic	Control Group	Treatment Group
Pre-test	Mean	8.7	8.96
	Standard Error	0.45	0.53
	Median	9	9
Post-test	Mean	6.47	9.93
	Standard Error	0.55	0.40
	Median	6	10

Moreover, 75% of students in the control group scored below 7 while 75% of students in the treatment group scored above 9. Furthermore, 25% of students in the treatment group scored above 12 on the post-test MC items whereas none of the students in the control group achieved such scores.

Since the data of pre-test MC scores did not follow a normal distribution, the Wilcoxon Signed-Ranks test was used to examine the score gains on the MC items. The control group data included 12 negative ranks meaning that 12 out of the 17 students received a lower score on the post-test MC items compared to the corresponding pre-test items. Only two students (i.e. 12%) in the control group performed better on the post-test MC items. The data from the treatment group included only 4 negative ranks whereas 14 out of the 27 students (i.e. 52%) in this case received higher scores on the post-test MC items than on the pre-test MC items.

Table 2. *Wilcoxon Signed-Ranks Tests for Multiple-Choice Gain Scores*

	Control Group	**Treatment Group**
Negative Ranks	12	4
Positive Ranks	2	14
Ties	3	9
Total (i.e. N)	17	27
z-score	-2.52	-1.71
z-score basis	Positive ranks	Negative ranks
Significance	0.012	0.087

Moreover, the negative z-score (-2.52) for the control group was based on positive ranks, meaning that the MC item scores decreased from pre-test to post-test. Furthermore, the MC scores of students in the control group were significantly lower on the post-test (z-score p-value = 0.012 < 0.05). In the treatment group, the negative z-score (-1.71) was based on the negative ranks, meaning that student scores on the MC items increased from pre-test to post-test. However, the MC item scores of students in the treatment group did not increase significantly (z-score p-value = 0.087 > 0.05).

A final piece of analysis was then carried out in which *normalized* gain scores were computed using the following ratio:

$$g = \frac{posttest \quad score - pretest \quad score}{\max imum \quad score - pretest \quad score}$$

(Bao, 2006)

Particular to this study, 12 students in the control group and 4 students in the treatment group demonstrated negative normalized gain scores. The mean normalized gain in the control group was -0.64 (standard error 0.23) whereas in the case of the treatment group it was 0.066 (standard error 0.11).

Normality tests - Kolmogorov-Smirnov (p = 0.011 and p = 0.000 for the control and treatment groups respectively) and Shapiro-Wilk tests (p = 0.032 and p = 0.000 for the control and treatment groups respectively) - indicated that the data differed significantly from the normal distribution. Subsequently, the Mann-Whitney test was used to compare the normalized gain scores of the control and treatment groups. The test indicated that the normalized gain scores of the treatment group were significantly different from the normalized gain scores of the control group ($p = 0.001 < 0.05$). Note than in the Mann-Whitney test, scores are ranked from lowest to highest. The test generated a mean rank of 14.41 for the control group and a mean rank of 26.96 for the treatment group. This means that the treatment group had a bigger number of higher normalized gain scores in comparison to the control group. In summary, Instructional Method B (treatment) had a significantly better effect on students' probability achievement on the multiple-choice items than Instructional Method A (control).

Comparison of Post-test Open-ended Item Scores

Apart from the 14 MC items common to the pre-test and post-test, the post-test included an MC item on discrete probability distributions; an open-ended item on simple and joint probabilities, conditional probability and independence; and an open-ended item on the binomial distribution. With regards to the MC item on discrete probability distributions, 23.5% of students in the control group and 48.1% of students in the treatment group responded correctly.

For the purposes of comparing the open-ended item scores of the control and treatment groups, scoring rubrics were created which allotted numerical values to student responses. Based on these rubrics, the first open-ended item received 18 points and the second open-ended item received 6 points. The minimum possible score on the open-ended items was 0 and the maximum 24.

Descriptive statistics for the data generated by the two open-ended items are presented in Table 3. The mean for the treatment group was 17.48, the standard error of the mean was 0.77, and the standard deviation 3.98. The mean of the treatment group on these items was 5.24 points higher than that of the control group. In addition, the scores of the control group on the open-ended items had a higher variation from the mean in comparison to the scores of the treatment group.

Table 3. *Descriptive Statistics for the Open-Ended Item Scores*

Statistic	**Control Group**	**Treatment Group**
N	17	27
Mean	12.24	17.48
Standard Error	1.34	0.77
Median	14	18
Standard Deviation	5.51	3.98

Based on the results, the control group had a median of 14, meaning that half of the students in the control group received a score lower than 14 (out of 24) on the post-test open-ended items. However, with the exception of four students in the treatment group, the remaining 23 students in this group scored higher than 14. That is, 85% of the students in the treatment group received a score higher than 14 on the post-test open-ended items. Moreover, whereas 75%

of students in the control group scored lower than 16, only 25% of students in the treatment group scored lower than 16 on these items.

The Kolmogorov-Smirnov and the Shapiro-Wilk tests indicated that the distribution of the sample in the case of the control group was not significantly different from a normal distribution but in the case of the treatment group it was significantly different (control group p-values 0.11 and 0.3 respectively on each test; treatment group p-values 0.04 and 0.001 respectively on each test). Subsequently, Mann-Whitney tests were used to compare the post-test results from the open-ended items in the two groups. The results ($p = 0.001 < 0.05$) indicated that the scores of the treatment group on the post-test open-ended items were significantly different from the scores of the control group. The test generated a mean rank of 14.15 for the control group and a mean rank of 27.76 for the treatment group, meaning that the treatment group had a bigger number of higher scores in comparison to the control group.

In summary, students in the treatment group performed better on the additional post-test MC item on discrete probability distributions. Moreover, based on the results of the Mann-Whitney test, the achievement of students in the treatment group on the post-test open-ended items was significantly higher than the achievement of students in the control group on the corresponding items. Therefore, Instructional Method B (treatment) was successful in producing significantly higher achievement scores on the post-test open-ended items compared to Instructional Method A (control).

<u>Comparison of Post-test Total Scores</u>

Overall, the maximum possible post-test score was 39 (15 points from MC items and 24 points from open-ended items). Particular to the treatment group, the mean was 27.89, the standard error of the mean was 1.01, and the standard deviation 5.26. The mean of the treatment group was 8.95 score points higher. In addition, the post-test scores of the control group had a higher variation from the mean in comparison to the post-test scores of the treatment group.

Table 4. *Descriptive Statistics for the Total Post-test Scores*

Statistic	**Control Group**	**Treatment Group**
N	17	27
Mean	18.94	27.89
Standard Error	1.61	1.01
Median	20	29
Standard Deviation	6.64	5.26

Based on the results, the median for the control group was 20 whereas the median for the treatment group was 29. Moreover, in the control group 75% of students achieved a post-test total score lower than 22 whereas in the treatment group 74% of the data was above 27.

The results of both the Kolmogorov-Smirnov and Shapiro-Wilk tests (control group p-values 0.2 and 0.9 respectively for each test; treatment group p-values 0.00 for both tests) indicated that the data for the control group did not differ significantly from the normal distribution whereas the data for the treatment group deviated significantly from the normal distribution. The Mann-Whitney non-parametric test resulted in a p-value of 0.00 (< 0.05) meaning that the post-test scores of the treatment group were significantly different from the post-test scores of the control group. The test generated a mean rank of 12.53 for the control group and 28.78 for the treatment group, meaning that the treatment group had a bigger number of higher scores compared to the control group.

In summary, the overall post-test achievement of students in the treatment group was significantly higher than the overall post-test achievement of students in the control group. Therefore, Instructional Method B was successful in producing significantly higher post-test scores compared to Instructional Method A.

Recommendations

The findings of this study may help mathematics instructors and curriculum developers by providing valuable information on the effect of particular tasks on students' achievement in probability. During the process of selecting the instructional materials to be used in this study, the researcher realized that there is a lack of available activities that could be used during instruction on probability in introductory statistics courses at the tertiary level. Given this, it is important that such activities be developed.

In Cyprus, there is a lack of research on students' achievement and understanding in the area of probability, especially at the secondary and post-secondary levels. Reports on the reformed mathematics curriculum place increased emphasis on new topics that have rarely been taught at the school level, including probability, and in the way these topics are taught i.e. using increased student involvement in the learning process and cooperative learning (Papanastasiou, 2002). Research needs to be carried out relative to the effects of these reformed materials, currently being used in K-12 classrooms in Cyprus, on students' achievement and understanding of probability concepts.

Moreover, relative to the significant differences in achievement between the treatment and control groups in this study, research needs to be carried out with the aim of revealing the factors that help explain these differences. Factors such as gender, the role of language and students' feelings regarding the use of group work should be examined. In addition, findings from a larger group of students at various tertiary education settings could produce trends that were not evident in this study.

References

Aliaga, M., Cobb, G., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, B., Utts, J., Velleman, R., and Witmer, J. 2005. *Guidelines for assessment and instruction in statistics education (GAISE) college report*. DOI= http://www.amstat.org/education/gaise/

Amit, M. and Jan, I. 2006. Autodidactic learning of probability concepts through games. In Novotná, J., Moraová, H., Krátká, M. and Stehlíková, N. (Eds.). *Proceedings of the 30th Conference of the International Group for the Psychology of Mathematics Education*, 2, 49-56. Prague: PME.

Bao, L. 2006. Theoretical comparisons of average normalized gain calculations. *American Journal of Physics*, 74(10), 917-922.

Batanero, C. and Sanchez, E. 2005. What is the nature of high school students' conceptions and misconceptions about probability? In Jones, G. A. (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 241-266). New York, NY: Springer Science and Business Media, Inc.

Cobb, G. 2000. Teaching statistics: More data, less lecturing. In T. L. Moore (Ed.), Teaching statistics: Resources for undergraduate instructors. *MAA Notes*, 52, 3-5. Washington, DC: Mathematical Association of America Inc.

Cooper, J., and Mueck, R. 1990. Student involvement in learning: Cooperative learning and college instruction. *Journal of Excellence in College Teaching*, 1(1), 68-76.

Dees, R. L. 1991. The role of cooperative learning in increasing problem-solving ability in a college remedial course. *Journal for Research in Mathematics Education*, 22(5), 409-421.

delMas, R. C., Ooms, A., Garfield, J. B., & Chance, B. 2006. Assessing students' statistical reasoning. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved July 2, 2009, from http://www.stat.auckland.ac.nz/~iase/publications/17/6D3_DELM.pdf

Effandi, Z. and Zanaton, I. 2007. Promoting cooperative learning in science and mathematics education: A Malaysian perspective. *Eurasia Journal of Mathematics, Science and Technology Education*, 3(1), 35-39.

Franklin, C. A. and Garfield, J. B. 2006. The GAISE project: Developing statistics education guidelines for grades pre-k-12 and college courses. In Gail F. Burrill and Portia C. Elliott (Eds.), *Thinking and Reasoning with Data and Chance: Sixty-eighth Yearbook* (pp. 345-375). Reston, VA: The National Council of Teachers of Mathematics, Inc.

Garfield, J., del Mas, R., and Chance, B. 2006. *Assessment resource tools for improving statistical thinking – ARTIST online tests*. DOI= https://app.gen.umn.edu/artist/tests/index.html

Giraud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistics Education*, 5(3).

Hirsch, L. S. and O'Donnell, A. M. 2001. Representativeness in statistical reasoning: identifying and assessing misconceptions. *Journal of Statistics Education*, 9(2).

Johnson, R. T., and Johnson, D. W. 1985. Student-student interaction: Ignored but powerful. *Journal of Teacher Education*, 34(36), 22-26.

Jones, G. A., Langrall, C. W., and Mooney, E. S. 2007. Research in probability: Responding to classroom realities. In Lester, F. K. Jr. (Ed.), *Second Handbook of Research on Mathematics Teaching and Learning: A Project of the National Council of Teachers of Mathematics*. Charlotte, NC: Information Age Publishing.

Jones, G. A., and Thornton, C. 2005. An overview of research into the teaching and learning of probability. In Jones, G. A. (Ed.), *Exploring Probability in School: Challenges for Teaching and Learning* (pp. 65-94). New York, NY: Springer Science and Business Media, Inc.

Mathematical Association of America. 1998. *Quantitative reasoning for college graduates: A complement to the standards*. DOI= http://www.maa.org/past/ql/ql_toc.html

Ministry of Education and Culture, Republic of Cyprus. 2008. *Inclusion in the Cyprus educational system at the beginning of the twenty first century: An overview - national report of Cyprus*. Nicosia, Cyprus: Author.

Mullis, I. V. S., Martin, M. O., and Foy, P. 2008. *International mathematics report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.

Papanastasiou, C. 2002. *TIMSS study in Cyprus: Patterns of achievements in mathematics and science. Studies in Educational Evaluation*, 28(2002), 223-233.

Potthast, M. J. 1999. Outcomes of using small-group cooperative learning experiences in introductory statistics courses. *College Student Journal*, 33(1).

Shaughnessy, J. M. 1977. Misconceptions of probability: An experiment with a small-group, activity-based model building approach to introductory probability at the college level. *Educational Studies in Mathematics*, 8(3), 295-316.

Shaughnessy, M. 1992. Research in probability and statistics: Reflections and directions. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 465-494). Reston, VA: National Council of Teachers of Mathematics.

Smith, K. A., Douglas, T. C. and Cox, M. F. 2009. Supportive teaching and learning strategies in STEM education. *New Directions for Teaching and Learning*, 117, 19-32.

TIMSS and PIRLS International Study Center. 1995. *TIMSS: IEA's third international mathematics and science study: Released item set for the final year of secondary school: Mathematics and science literacy, advanced mathematics, and physics*. DOI= http://timss.bc.edu/timss1995i/Items.html

TIMSS and PIRLS International Study Center. 2001. *TIMSS 1999 mathematics items: Released set for eighth grade*. DOI= http://timss.bc.edu/timss 1999i/study.html

TIMSS and PIRLS International Study Center 2007. *TIMSS 2003 mathematics items: Released set eighth grade*. DOI= http://timss.bc.edu/timss2003i/released.html

TIMSS and PIRLS International Study Center 2009. *TIMSS 2007 user guide for the international database*. DOI= http://timss.bc.edu/TIMSS2007/items.html

Watson, J. M. 2006. *Statistical literacy at school: Growth and goals*. NJ: Lawrence Erlbaum Associates Inc.

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., and Chang, B. 2008. What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education*, 16(2).