Using Data Analysis Projects to Promote Statistical Thinking in an Introductory Statistics Course: A Basis for Curriculum Materials Development

ABSTRACT: Statistical thinking has long been a topic of discussion and a generally agreed upon goal for statistics instruction. Statistics involves distinctive and powerful ways of thinking. Statistics is a general intellectual method that applies wherever data, variation, and chance appear. Any introductory course should take as its main goal helping students to learn the basic elements of statistical thinking. Many advanced courses would be improved by a more explicit emphasis on those same basic elements. Those elements were described as: the need for data; the importance of data production; the omnipresence of variability; and the quantification and explanation of variability. The use of data analysis projects provides also students with the opportunity to demonstrate their ability to apply and integrate statistical knowledge and skills in analysing information statistically. This paper will describe the projects and types of statistical analysis that had been selected by 31 counselling students enrolled in an introductory statistics course at the undergraduate level. Content analysis was carried out on their final report of the projects and survey was used to elicit their experiences of working on the projects. These findings will eventually be the basis for the development of curriculum materials to help instructors and their students implement data analysis projects in their respective classrooms.

KEY WORDS: Counselling, projects, survey, statistical thinking, statistics, curriculum materials, and undergraduate students.

INTRODUCTION

Statistical thinking has long been a topic of discussion and a generally agreed upon goal for statistics instruction (Snee, 1999; Chance, 2002; and Rumsey, 2002). In his influential report on the need to reform the teaching of introductory college statistics classes, G. Cobb (1992) wrote that any introductory course should take as its main goal helping students to learn the basic elements of statistical thinking. Many advanced courses would be improved by a more explicit emphasis on those same basic elements. Those elements were described as: the need for data; the importance of data production; the omnipresence of variability; and the quantification and explanation of variability.

Further, D.S. Moore suggested that statistics involves distinctive and powerful ways of thinking. He said that statistics is a general intellectual method that applies wherever data, variation, and chance appear. It is a fundamental method because data, variation, and chance are omnipresent in modern life (Moore, 1998:134). In their landmark paper, C.J. Wild & M. Pfannkuch (1999) provided an empirically-based model of statistical thinking that described the processes involved in the statistical practice of data-based enquiry from problem formulation to conclusions.

For many students, nevertheless, the prospect of taking an introductory statistics class is still daunting. R.E. Kirk (2002) reported that students believe an introductory
statistics course to be demanding, to involve lots of math, and to be irrelevant to their career goals. Thus, R. Snee (1993) advocated changes in the instructional delivery system for statistics education citing people’s lack of understanding of statistical thinking which resulted in a lack of appreciation for statistical thinking. R. Snee further suggested that experiential learning, which includes working with real data and having students work on subject matter in which they take a personal interest, are vital to improving students’ understanding of statistics (Snee, 1993).

Overall, the literature suggests that using projects of some type in an introductory statistics class may positively influence learning of statistics (Garfield, 1995; Bradstreet, 1996; Moore, 1997; and Shaughnessy, 2007). However, often the projects that students in these studies were asked to do were designed by the instructor in some way. Usually, the instructor came up with at least the topic for the project. G. Smith did allow students, with prior instructor approval, to modify or replace a given project. However, he reported that students seldom asked to make these changes (Smith, 1998).

In spite of the increasing support for such projects as sound pedagogical tools for teaching statistics, many instructors still do not incorporate projects into their statistics courses (Landrum & Smith, 2007). This could be an obstacle to developing statistical reasoning and thinking among the students.

CHALLENGES IN DEVELOPING STATISTICAL THINKING

Despite agreement on the need for statistical thinking, there have been no empirically tested instructional materials or methods that can be shown to develop such an important learning outcome. In fact, studies of student outcomes in a first statistics class show an alarming lack of statistical reasoning and thinking (delMas et al., 2007).

The Guidelines for Assessment and Instruction in Statistics Education (GAISE) – endorsed by the American Statistical Association in 2005 – include the goal of developing students’ statistical thinking. The report also offers examples of what this type of thinking may look like and advocates the importance of teaching and modeling statistical thinking in the introductory statistics course (Aliaga et al., 2012).

Furthermore, the report suggests that merely learning statistical content (i.e. terms, formulas, and procedures), even with real data and research studies used as a context, does not appear to lead students to think more like statisticians, along the lines described by C.J. Wild & M. Pfannkuch (1999). What is needed is a radically different approach that is designed to help students to begin to think statistically and to build on this thinking to understand and appreciate the discipline of statistics.

In this paper, I will share my experiences using data analysis projects with my students. The purpose of this study was to explore the extent to which the use of data analysis project in the course where students worked in groups would result in students’ better thinking of statistics. This was done through the listing of the topics that the students themselves had chosen and the types of statistical tests they had selected to answer their research questions. In addition, I also shared what the students had thought of their experiences with the projects and the course itself. Finally, implications of the findings of this study will be discussed as a basis for curriculum materials development for the undergraduate students in the field of social science.

THE SETTING

Beginning the first week of class, students who were enrolled in an introductory statistics course were informed of the data analysis project. For the project, students were asked to choose their own research topics, define their variables, articulate their research questions, devise and carry out a data collection plan, conduct the appropriate analysis on the data, and prepare both a written report and an oral presentation to share the results with the instructor and the rest of the class. Students were given the choice to either work individually or in small teams of 2-4 people. In short, the projects are broadly divided into three phases: the data collection phase, the data analysis phase, and the dissemination phase.
About the Data Collection. During the data collection phase, students gather data from constructing and administering surveys. They may choose to construct their own survey or download and adapt them from the internet. They are also encouraged to use surveys that they had used before in their research methodology class.

About the Data Analysis. The results are produced during the analysis phase of the project, after the data are collected and organized. The requirements for each project depend in part on their research questions for the type of test or analysis to be conducted, which are routinely taught to students during twice weekly class meetings of 2½ hours each for 12-14 weeks.

Regardless of the type of analysis required for a given project, students were strongly urged to produce the appropriate descriptive statistics, including graphical representations (e.g. histograms) of the distribution of their sample data. Students used SPSS (Statistical Package for the Social Sciences) software to accomplish these analyses.

For t-test projects, students must compute and interpret the t statistic and p-value for the test. For linear regression projects, students must create a scatter plot and a graph of the regression line; they must also compute and interpret the value of the correlation coefficient \( r \), coefficient of determination \( R^2 \), and equation of the regression line.

About the Dissemination. When the students have collected and analyzed their data, the last phase of the project was to prepare a written report and in-class presentation. Most frequently, students chose to use Microsoft Word and Power Point for these two tasks.

RESULTS AND DISCUSSIONS
The 31 students in this class have divided themselves into 10 groups of 3 to 4 members each. None chose to work individually.

About the Title of Students’ Research Projects. The topics that they had selected for their group projects are as listed in table 1. Most of the topics chosen are diversified but related to their field of specialization, that is counselling. They can be broadly categorized into four: (1) conflicts and problems; (2) perception studies; (3) effect studies; and (4) teaching and learning.

<table>
<thead>
<tr>
<th>Category</th>
<th>Title of Project</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conflicts and Problems.</td>
<td>College students’ involvement in conflicts with their roommates.</td>
</tr>
<tr>
<td></td>
<td>College students’ financial problems: Causes and strategies to overcome them.</td>
</tr>
<tr>
<td></td>
<td>Social networking addiction among college students.</td>
</tr>
<tr>
<td>Perceptions Studies.</td>
<td>Students’ perception and knowledge of gay, lesbian, bisexual, and transgender</td>
</tr>
<tr>
<td></td>
<td>issues.</td>
</tr>
<tr>
<td></td>
<td>Perception on counselling as career.</td>
</tr>
<tr>
<td>Effect studies (v, vii, viii,</td>
<td>Effects of co-curricular activities on college students’ social development.</td>
</tr>
<tr>
<td>and x).</td>
<td>Effects of cultural diversity on college students’ social interactions.</td>
</tr>
<tr>
<td></td>
<td>Effects of self-esteem on academic achievement.</td>
</tr>
<tr>
<td></td>
<td>Effects of working while studying on academic achievement.</td>
</tr>
<tr>
<td>Teaching and Learning.</td>
<td>Can PBL (Problem Based Learning) be a powerful T&amp;L (Teaching and Learning) tool</td>
</tr>
<tr>
<td></td>
<td>in enhancing humanistic skills among undergraduates?</td>
</tr>
</tbody>
</table>

Table 1: Titles of Projects
they members of multiple networking sites?;
(6) What do students perceive as the main
factors contributing to their involvement
in gay, lesbian, bisexual, and transgender
issues?; (7) What are the students’ perceptions
on counselling as career?; (8) Across the
faculties surveyed, students from which faculty
obtained the highest level of self-esteem?;
(9) What are the advantages of extra co-
curricular activities to college students?; (10)
How receptive are local students towards
their international counterparts?; (11) What
are the main reasons for students to work
while studying?; (12) What is the students’
overall perception of PBL, or Problem Based
Learning, as a new approach of learning at the
end of the semester?; and (13) What is the
students’ perception of PBL in developing their
motivation and learning skills?

Second, Research Questions Requiring
Inferential Statistics. For the research
questions requiring inferential statistics,
students’ selection of statistical test is limited
to t-tests, ANOVA, and correlation.

Examples of research questions using t-tests
are: (1) Is there a difference between males
and females with regard to their involvement
in conflicts with their roommates?; (2) Which
gender is more involved in financial problems?;
(3) Which gender is more affected by social
networking addiction?; (4) Which gender is
more involved in gay, lesbian, bisexual, and
transgender activities?; (5) Which gender has a
better perception on counselling as career?; (6)
Which gender has a higher self-esteem: male
or female?; (7) Which gender is more receptive
towards international students?; (8) Which
gender is more involved in working while
studying?; and (9) Is there any difference in the
perception on effectiveness of PBL between
male and female students?

All groups except one had used the
independent t-tests. Further, all the research
questions revolved around gender differences
in the topic they had studied.

Examples of research questions using
ANOVA are: (1) Is there a significant
difference across courses/majors with regard
to conflicts with roommates?; (2) Do students
from different faculties differ in terms of
their financial management?; (3) Is there any
difference between students’ level of study and
their social development?; and (4) Do students
from difficult specializations differ significantly
in terms of their perception on the effectiveness
of PBL or Problem Based Learning?
The independent variables for the research
questions using ANOVA are the different
courses/majors/specializations, faculties, and
level of study.

Examples of research questions using
correlation are: (1) Is there any association
between parents’ income and financial
problems?; (2) Is there any association between
parents’ income and social networking
addiction?; (3) Is there any association
between age and knowledge of gay, lesbian,
bisexual and transgender issues?; (4) Is there
any association between students’ CGPA,
or Cumulative Grade Point Average, and
perception on counselling as career?; (5) Is
there any association between CGPA and self-
estee scores?; (6) Is there any association
between students’ involvement in extra co-
curricular activities and social development?
(7) Is there any association between fluency
in English and social interactions?; and (8) Is
there any association between wage earned
from working and academic achievement?

All the above research questions used
bivariate correlation; and each variable is at
least at the interval level of measurement.

About the Survey Findings and
Interpretations. Upon completion of their
projects, students were asked to complete
questionnaire seeking their satisfaction
working with the data analysis project and
the course itself. Findings from the survey
are divided into two parts: the first being
respondents’ rate of satisfaction with the data
analysis group project; while the second part
elicited how they felt towards the course itself.

First, What do they think of the data
analysis project? Students’ experiences with
the data analysis project were captured through
the questionnaire items. From it, 74.2% of the
respondents indicated their agreement that the
topic of the project their group has chosen is
related to real life situation; nevertheless, the
same percentage of the respondents (74.2%) of
the opinion that the project is challenging.

One of the benefits of the project that
they had undertaken was that the project gave them more confidence to interpret statistical data in other courses (64.5%); and helped them understand how data is processed using the statistical tests (67.7%). About 64.5% of the respondents also felt that through doing the project their learning of statistics was enhanced. Although almost half of the respondents (48.4%) felt that they had difficulty in understanding the statistics involved in the project, they admitted that the projects did capture their interest in learning statistics (42%); and the project had also prepared them to deal with statistics outside the classroom (45.1%).

In addition, all responses to the items are negatively skewed. The mean, standard deviation, number of respondents, and percentage as well as skewness for the items on respondents’ satisfaction towards the data analysis project is provided in table 2.

**Second, What do they think of the course?**

The respondents’ perceptions of the course were found to be negatively skewed, except in two items: “Statistics is an easy subject” and “learning statistics is fun”. Majority of the respondents did not agree with the first statement; that is, they were of the opinion that statistics is not as an easy subject. Hence, learning it becomes no fun.

In addition, it was rather unfortunate that before taking the course, a majority of respondents (77.4%) had heard that statistics is a difficult course; and only 35.5% of the respondents’ initial attitude (that statistics is difficult) changed after attending the course. The good news, however, is that 61.3% of the respondents were of the opinion that they would be using the statistical tests to conduct research in their profession as counsellors; and 58.1% agreed with the statement that learning statistics did prepare them to be smart consumers in today’s society. Finally, almost half of the respondents (48.4%) were confident that they will get very good grades in the statistics project; and 42% of them are now becoming more confident to apply statistics in future.

---

**Table 2:**

Respondents’ Perceptions of Group Project, No (%)

<table>
<thead>
<tr>
<th></th>
<th>Mean (SD)</th>
<th>Strongly Disagree</th>
<th>Neutral</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>The group statistics project captures my interest in learning statistics.</td>
<td>3.06 (1.09)</td>
<td>(6.5)</td>
<td>(29.0)</td>
<td>(22.6)</td>
</tr>
<tr>
<td>The group statistics project is difficult to understand.</td>
<td>3.35 (1.05)</td>
<td>(3.2)</td>
<td>(19.4)</td>
<td>(29.0)</td>
</tr>
<tr>
<td>The group statistics project is challenging.</td>
<td>3.87 (0.92)</td>
<td>(3.2)</td>
<td>(3.2)</td>
<td>(19.4)</td>
</tr>
<tr>
<td>The topic of the project our group has chosen is related to real life situation.</td>
<td>4.10 (0.87)</td>
<td>-</td>
<td>(3.2)</td>
<td>(22.6)</td>
</tr>
<tr>
<td>The project gives me more confident to interpret statistical data in other courses.</td>
<td>3.61 (0.84)</td>
<td>-</td>
<td>(12.9)</td>
<td>(22.6)</td>
</tr>
<tr>
<td>The project helps me understand how data is processed using the statistical tests.</td>
<td>3.55 (1.06)</td>
<td>(3.2)</td>
<td>(19.4)</td>
<td>(9.7)</td>
</tr>
<tr>
<td>Doing the project enhanced my learning of statistics</td>
<td>3.52 (1.06)</td>
<td>(3.2)</td>
<td>(19.4)</td>
<td>(12.9)</td>
</tr>
<tr>
<td>The project has prepared me to deal with statistical situations outside the classroom.</td>
<td>3.06 (1.12)</td>
<td>(12.9)</td>
<td>(16.1)</td>
<td>(25.5)</td>
</tr>
</tbody>
</table>
One of the groups summarized their experience in the course as follows:

Learning statistics is not something very difficult to learn, but it requires a very deep interest and concentration if students want to understand it well. And we really hope that we will continuously get the important information about statistics from time to time until we enter the world of work in the future (Interview with Students Group A, 15/10/2013).

Table 3 provides the mean, standard deviation, number and percentage as well as skewness for the items discussed above.

My present experience suggests that students are more engaged in the course material and learn the material better when involved in the group project than when presented with traditional lectures. However, it should not be viewed as the sole method for improving instruction in a heavily concept-oriented course such as statistics. Course material must be organized to meet clearly defined course objectives, and class activities and instructors must be oriented toward giving students practice in applying difficult concepts.

In incorporating the project into the course, the instructor must also consider the impact on the instructor's in-class and out-of-class time. In courses with large enrolments, the initial establishment of group processes may take some commitment of course instructional time (Maziha Mustapha & Nik Abd Rahman, 2011).

I have found that this time is more than made up by the increased efficiency of assisting students in groups rather than as individuals. Questions may also be answered by group members, freeing the instructor to assist students in developing an understanding of the more abstract and difficult concepts.

An unexpected benefit, I have experienced from this group project, is the day-by-day feedback the instructor gets from reading and grading the material turned in by the groups. Since these papers are relatively few in number, it is not a burden for the instructor to grade them. The instructor finds out what the students understand and what must be

<table>
<thead>
<tr>
<th>Mean (SD)</th>
<th>Strongly Disagree</th>
<th>Neutral</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning statistics is fun.</td>
<td>2.61 (1.20)</td>
<td>7 (22.6)</td>
<td>8 (25.8)</td>
</tr>
<tr>
<td>Before taking this course, I heard that statistics is difficult.</td>
<td>4.03 (1.43)</td>
<td>3 (9.7)</td>
<td>2 (6.5)</td>
</tr>
<tr>
<td>After attending this course, my initial attitude towards statistics has changed.</td>
<td>3.06 (0.89)</td>
<td>2 (6.5)</td>
<td>5 (16.1)</td>
</tr>
<tr>
<td>Many real world examples involving statistics have been included in this course.</td>
<td>4.0 (1.0)</td>
<td>- (9.7)</td>
<td>3 (19.4)</td>
</tr>
<tr>
<td>I will use statistical tests to conduct research in my profession as a counsellor.</td>
<td>3.68 (0.98)</td>
<td>1 (3.2)</td>
<td>2 (6.5)</td>
</tr>
<tr>
<td>Learning statistics prepare me to be a smart consumer in today’s society.</td>
<td>3.58 (0.92)</td>
<td>1 (3.2)</td>
<td>2 (6.5)</td>
</tr>
<tr>
<td>I am more confident to apply statistics in future.</td>
<td>3.06 (1.09)</td>
<td>2 (6.5)</td>
<td>9 (29.0)</td>
</tr>
<tr>
<td>I will get a very good grade in the statistics project.</td>
<td>3.45 (1.34)</td>
<td>4 (12.9)</td>
<td>2 (6.5)</td>
</tr>
</tbody>
</table>

Skewness: 0.885
Skewness: -1.536
Skewness: -0.734
Skewness: -0.185
Skewness: -0.649
Skewness: -0.661
Skewness: -0.049
Skewness: -0.470
Skewness: -0.470
Skewness: -0.470
reviewed or re-taught. This has resulted in an experiential learning process that enables the students to integrate critical thinking, technical writing, and presentation skills into an analytical course thus promoting statistical thinking.

CONCLUSION

In conclusion, three main practical implications can be drawn from this study. First, the teachers and the researcher need to come to a common consensus of what they mean by the term statistical thinking. Second, the teachers need to reflect critically on their current teaching and identify areas which are acting as barriers to the development of their students’ statistical thinking. Third, the identification of the current main barriers to statistical thinking led the researcher to designing a new statistics teaching unit. New activities were chosen on the basis that they would be interesting to the students and would allow sufficient exploration of the data within an empirical enquiry process.

References


Interview with Students Group A, who were enrolled in an Introductory Statistics Course, in Gombak, Malaysia, on 15 October 2013.


Students should be Promoted toward the Statistical Thinking
(Source: www.google.com, 12/5/2014)

Regardless of the type of analysis required for a given project, students were strongly urged to produce the appropriate descriptive statistics, including graphical representations (e.g. histograms) of the distribution of their sample data. Students used SPSS (Statistical Package for the Social Sciences) software to accomplish these analyses.