A Teaching Idea:
Learning Statistics Through Real Problems

Wong Ka-Ming
Department of Curriculum Studies,
The University of Hong Kong

Statistics has been a substantial component in the local school mathematics curriculum for decades. But its teaching does not usually arouse much discussion among mathematics teachers and educators, perhaps because it is rather commonly perceived as a relatively "soft" part in the mathematics curriculum. First, the mathematics involved in statistical computations, for the most part, does not present particular difficulties to an average student, especially when the calculator nowadays takes up almost all routine computations. Second, most teachers find it straightforward to teach statistics by introducing formulae, demonstrating examples and working on exercises. In any case, more complicated formulae, like the computation of the standard deviation, need not be explained or proved, as far as the curriculum targets are concerned. Third, public examination questions in this topic at the Certificate of Education level are usually routine tasks in which students can exhibit relatively satisfactory performance.

As a matter of fact, many teachers, when under time constraint, are most likely to choose to omit teaching units in statistics for their mathematics classes, especially those units perceived to be not having much mathematical content (e.g., the abuse of statistics) or not conveniently assessable in usual tests and examinations (e.g., comparing the use of different measures of central tendency).

Insofar as my personal experience in teaching supervision and curriculum discussion is concerned, most local mathematics teachers do not pay particular attention to reflecting upon statistics teaching, as there are seemingly no particular pedagogical difficulties in this area. [Note 1] That most mathematics teachers themselves have not received any formal, substantial training in statistics may be a strong factor contributing to this taken-for-granted attitude towards teaching school statistics. Simply to put, their instructional focus is on the mathematical aspect of statistics only. On the other hand, as a literature search can easily tell, there has been a dearth of local research studies in teaching statistics, showing that statistics, as part of our mathematics curriculum notwithstanding, has not attracted much attention from local mathematics educators either.

Admittedly, when I was a school mathematics teacher more than a decade ago, the best idea in teaching statistics I had (as different from teaching other mathematical topics in the curriculum) was to use, as far as possible, more realistic data from various daily life contexts or fields of application to introduce and explain the statistical notions. And this belief was (and is) undoubtedly shared by many teachers, at least at a theoretical level, if not actually realised in their day to day teaching. When a group of colleagues and I were writing a series of school mathematics textbooks (in the mid-1980's), we were thinking mostly about the mathematics involved in these units on statistics. Thus, we fabricated "mathematical" examples rather than "statistical" examples. For instance, we asked students (at Secondary 5) to compare the dispersion of two distributions by considering the steepness of the cumulative frequency curves. Of course, this was a legitimate mathematical problem in statistics, though some teachers might find it hard for their students. Yet there could be a lot more statistical considerations made on the two distributions posed. The

problem with our approach then was, with hindsight, that we had not placed enough emphasis on the non-mathematical aspects of statistics.

For this remark, I am actually echoing a point made in a survey report on local statistics teaching published in the 1980's, which is surprisingly the only large-scale survey (1321 teachers from 338 secondary schools) we have in the local research literature on statistics education until now. In this report, Cheung, Lam, Siu and Wong (1986) pointed out four major areas of concern:

1. The subject matter does not bear enough relevance to students.
2. The subject matter is not taught in a lively manner.
3. There is a wrong emphasis on techniques rather than concepts.
4. The non-mathematical aspect of the subject is ignored. (pp. 176-178).

With present day colourful textbooks usually updated with more realistic data from daily life situations and with the availability of modern information technology for classroom teaching, I tend to believe that the first two areas may have seen some improvements now, though the extent is uncertain. But I suspect the last two problems still persist, if not more seriously, as informed by my personal experience in teacher education since mid-1990's. In this short paper, I am not in the position to discuss these questions further, not to mention some recent concerns. But interested readers are referred to a recent paper in this regard, namely, Wong (2000).

To remedy some of the said problems, project work, *inter alia*, has been advocated (Cheung et al., 1986), and this is to provide students with more hands-on experience in working with real statistics, starting from the stage of conceptualisation of a statistical study and then collecting relevant data and so on. As a means to promote this teaching approach, the autumn of 1986 saw the inauguration by the Hong Kong Statistical Society of an annual activity of

statistical project competition for secondary school students (Shen, Li, & Lam, 1995).

While this activity approach has its advantages and effectiveness in improving the quality of statistics education, there are situational considerations and constraints in the school context which may prevent its wide adoption by mathematics teachers. First and foremost, successful and meaningful project work cannot be conducted in the class without due support and guidance from the teacher from time to time. Request for teachers' advice and suggestions can be quite demanding. Project work is, understandably, likely to take up much class time for initial discussion, progress reports, comments from the teacher, and final presentation of results. In view of the tight teaching schedule and teachers' already heavy workload, this is, therefore, not a realistic option for learning the subject in many schools. As a matter of fact, many practising teachers themselves did not have this project work experience when they were students at school. For this lack of first-hand experience in learning, teachers may feel reluctant to try this mode of learning in their own classroom.

When project work is conducted by student groups, there may be another pertinent issue which concerns the prevalent learning culture. In many local classrooms, students tend to compete (for higher academic achievement) than to collaborate in learning activities. This can be an obstacle to promote students' participation and cooperation in project work. Even if granted a better collaborative classroom atmosphere, students often have grievances and make complaints about very unequal contributions to the group work, especially when it carries a substantial weight in assessment. Furthermore, the open-ended nature of project studies in general also induces more anxiety among students as compared with conventional exercises and assignments. I guess many teachers can readily extend this list of constraints further. Notwithstanding all these

negative factors, I do not mean to disapprove the promotion of statistical project work in school. Rather, teachers' contribution and effort required in this matter should never be underestimated, and pointing out these issues in advance better prepares the teachers to implement it in their teaching in some ways feasible in their own classrooms.

As an alternative to project studies, what I want to propose as a teaching idea in this paper is something less demanding in terms of resources, but can possibly bring about improved learning in statistics education, though the associated learning experience may still be less comprehensive as compared with project work. This idea came about two years ago when I had to teach a course on the teaching of Advanced Supplementary Level Statistics, an elective module in our Postgraduate Certificate in Education (in-service) programme for teachers. Course participants in the past two years were mostly mathematics teachers, with only a few having major or minor in statistics in their undergraduate years. As part of the course assessment, I gave them this exercise to work out, and their performance was discussed afterwards. For lack of a better name, I called it "*Problem Study*." This exercise can be used for individual or group work.

This so-called Problem Study assignment, on its surface, looks like conventional types of statistical problems, but actually consists of several real problems which are adapted in some simplified manner for statistical exploration, analysis and deliberation in the classroom. The problems I used in my course are in fact selected from the book *"Statistics in the real world: A book of examples"* by Larsen and Stroup (1976). They are different from the usual types of problems found in local textbooks on statistics in that other than computational tasks, there are questions which provoke an analysis into the real situations for an appropriate application of the statistical methods and

sometimes also an evaluation of the validity of the conceptualisation of the situation in question. As a matter of fact, each problem begins with a description of the situation and relevant details, which is much more wordy than conventional computational problems. Some questions are open-ended indeed, and the quality of their solutions depends very much on the students' orientation and understanding with respect to the particular settings. Needless to say, apart from mathematical methods, much common sense has to be invoked for the solution. That is to say, students have to make sense of the situation before they can look for some viable solutions. As an assignment for my course, these questions have served the purpose of demonstrating to mathematics teachers a possible type of student work that can cultivate a better understanding of statistical concepts and provide learning opportunities in the non-mathematical aspect of the subject.

In order to convey a more concrete sense of the nature of this exercise and a flavour of the questions involved, let me present herewith an example (pp. 4-10 in Larsen and Stroup, 1976). The problem starts with a passage that describes a puzzle surrounding the authorship of a set of ten letters which appeared during 1861 in an American newspaper, the *New Orleans Daily Crescent*. These letters chronicled the author's adventures as a member of the Louisiana militia in the Civil War, and were signed "Quintus Curtius Snodgrass," a name which seemed to have no real record despite the unquestionable authenticity of the events described. Interestingly, the style of these letters allegedly bears traces of the humour and irony of the famous American writer Mark Twain, who had, during his lifetime, been refusing to reveal his own role in the Civil War. The objective of the exercise is to determine whether Mark Twain could have been the real author of the Quintus Curtius Snodgrass letters, a subject of historical debate in actual fact. As Larsen and Stroup (1976) said, "[t]he analysis in this example shows how descriptive

statistics can be used to solve what appears to be a totally nonstatistical problem – in a totally nonstatistical subject" (p. 5).

The statistical procedure has been based on an observation (given in the problem description) that a given author will use roughly the same proportion of, say, three-letter words in his/her writing quite consistently, much like kind of "fingerprints." The same holds true for words of any length. Furthermore, the proportion of *three-letter* words used by an author is very likely to differ from the proportion used by another. Then data on word-length distributions are presented for three sets of letters known to be written by Mark Twain and also for the ten Quintus Curtius Snodgrass letters (combined counts).

So there are the usual type of computational problems, like computation of relative frequency, drawing of relative frequency polygons, calculation of measures of locations or central tendency, etc. Statistical inferences regarding the puzzle on authorship are to be drawn from these computations. Apart from these, there are further thought-provoking questions. For example, students are asked what kinds of words, if any, have to be excluded in making the word counts (i.e., the frequency data to start with) in order that this method of comparison is sound (with good justification, of course). [Note 2] Another question which prompts students to think more deeply into the rationale underlying the method itself is as follows:

> In a paternity suit, certain medical information (blood type, length of pregnancy, and so on) can sometimes be introduced that will prove the defendant is *not* the child's father. But there is no way that such information can prove that the defendant *is* the child's father. Is it likely that statistical tests of authorship

> are similarly "one sided"? Explain. (Larsen & Stroup, 1976, p. 10, italics in the original) [Note 3]

This problem as a whole, as you can agree, is a *real* problem. Students can understand that after this exercise no one, not to mention the class teacher, can know everything for sure, and the mystery (or the unknown as in other cases) may still be taken to persist unless there are new evidences uncovered in future, although the statistical method *does* point to some direction as more likely to be the truth. So in this sense, this problem differs very much from the usual types of school problems. The students work it out, like a statistician, a historian, a scientist, a researcher, or a detective, that is, in some fashionable jargons, performing a form of *authentic activity* (as opposed to mere school learning activity). They are not guessing or searching for "the" answers on the teacher's mind. And it is likely that some students may even embark on sound ideas and possibilities that the teacher may have missed to consider. This is a very good example of the sort of so-called *authentic* tasks.

As a matter of fact, this problem has been adapted from a research paper in statistics, namely, C.S. Brinegar's "*Mark Twain and the Quintus Curtius Snodgrass letters: A statistical test of authorship*" (1963), from which the data on the frequency counts are quoted. This also demonstrates to teachers how this type of real problems can be formulated. Viable sources for re-formulation into school work can be research papers in journals, magazines of various disciplines, and reports and publications on a diverse range of issues published by various research institutions, government departments or public organisations. Of course, as a starting point, Larsen and Stroup's (1976) book is quite a good reference albeit slightly outdated. Writing such real problems for students surely will cost much time and effort on the part of the teachers, but it will pay, I believe.

It must be emphasised that my proposal here is nothing new and I myself have not actually done much further on this. First, I have only used problems selected from the above-mentioned book for my class of in-service teachers. [Note 4] Second, I can also see a close connection of this teaching idea to what Cheung and others (1986) have suggested some 15 years ago, namely, the production of teaching materials in the form of *"capsules."* Given sufficient resources, the real problems proposed here can be worked out in a much more elaborated manner to constitute components for teaching capsules focussing on different statistical topics in the school mathematics curriculum.

On the other hand, insofar as my personal experience in using this Problem Study is concerned, the result, if not worrying, is not at all optimistic however. The performance of the course participants in the past two years has revealed a very clear trend. Most (mathematics) teachers did perfectly well in explicit computational tasks, but showed incomplete or problematic understanding of the statistical concepts involved. Furthermore, their performance in questions on the non-mathematical aspect of the situations, like the two questions cited above, has been rather poor. This is possibly attributable to the stale or biased statistics education they had received at school (as most teachers had not taken any further courses on statistics at universities), so that they are in fact holding a very narrow or even primitive conception of statistics, and therefore take statistics to mean merely a collection of computational techniques applicable to various situations to obtain statistics and to draw inferences of varied degree of likelihood. The intricacies in understanding the contextual meaning of its various applications are entirely overlooked, not to mention how to cultivate sense-making in their own teaching of statistics. In view of this biased knowledge background in statistics among many mathematics teachers, an initial effort to formulate more real problems useful in

the classroom may have to depend on teachers who have a more thorough training in statistics, or even better, on well-trained statisticians, if we mean to bring about substantial changes to the kind of learning experiences in school statistics.

As a last remark, some careful readers may have noticed that in this paper I have been advocating something *apparently* opposite to what I have been saying so far on other occasions regarding the relevance of authentic tasks for mathematics learning, for example, in Wong (1997, 1998). This is a deep question which I cannot deal with here lightly, as it necessarily pertains to the nature of the two disciplines, mathematics and statistics. While statistics is not currently a separate school subject and has to reside under the roof of mathematics, it may not mean necessarily that approaches to teaching the statistical topics are the same as those to teaching other mathematical topics. And in my humble opinion, the two disciplines are quite different in terms of its nature and subject matter, and so the instructional approaches should be different accordingly too, though observation of classroom teaching usually tells us the opposite story as mentioned in the very beginning. To justify my proposal in a concise, if not simplistic, manner, the real-problem-approach discussed here does not mean to focus on the teaching of statistical computations, but rather on the cultivation of a better sense and understanding in the *non*-mathematical aspect of statistics, and thereby a more comprehensive conceptual understanding of the subject.

Notes

[1] In this regard, readers are referred to some relevant data in a recent large-scale survey on teachers' views on various topics in the mathematics

curriculum and their perception of students' learning difficulties reported in Wong, Lam, Leung, Mok and Wong (1999).

[2] I have rephrased this question. The readers are encouraged to contemplate on this question. (All necessary details have been mentioned.)

[3] Of course, the statements made here (in 1976) may become a bit outdated in view of the rapid advances in DNA studies. But the cited question is still susceptible to some minor revision to the same effect.

[4] Larsen and Stroup's (1976) book is available at the Main Library of The University of Hong Kong. For copyright reasons, I cannot freely reproduce further examples from the book here. But interested readers may contact me via e-mail <k.m.wong@graduate.hku.hk> for further information.

References

Brinegar, C.S. (1963). Mark Twain and the Quintus Curtius Snodgrass letters: A statistical test of authorship. *Journal of the American Statistical Association, 58*, 85-96.

Cheung, P.H., Lam, K., Siu, M.K., & Wong, N.Y. (1986). An appraisal of the teaching of statistics in secondary schools of Hong Kong. *Hong Kong Science Teachers Journal, 14*, 171-186.

Larsen, R.J., & Stroup, D.F. (1976). *Statistics in the real world: A book of examples.* New York: Macmillan.

Shen, S.M., Li, K.Y, & Lam, K. (1995). Activity teaching approach in statistics: Statistical project competition for secondary students. In M.K. Siu (Ed.), *Retrospect and outlook on mathematics education in Hong Kong: On the occasion of the retirement of Dr. Leung Kam Tim* (pp. 157-165) (in Chinese). Hong Kong: Hong Kong University Press.

Wong, K.L. (2000). Statistics in the 2001 mathematics syllabus: Statistical investigations in a technological environment. *Proceedings of the Hong Kong Mathematics Education Conference 2000*, pp. 58-71.

Wong, K.M. (1997). Do real-world situations necessarily constitute "authentic" mathematical tasks in the mathematics classroom? *Curriculum Forum, 6*(2), 1-15.

Wong, K.M. (1998). How authentic should the situations in mathematical word problems and tasks be? *EduMath, 7*, 44-54 (in Chinese).

Wong, N.Y., Lam, C.C., Leung, F.K.S., Mok, I.A.C., & Wong, K.M. (1999). *An analysis of the views of various sectors on the mathematics curriculum: Final report. A research study commissioned by the Education Department.* Hong Kong: Education Department.