WHAT DID THEY LEARN? STATISTICS SKILLS: FROM FRENCH SECONDARY SCHOOL TO UNIVERSITY

Alain Bihan-Poudec[1] & Philippe Dutarte[2]
[1]Université Catholique de l'Ouest, Angers, France
[2]Inspection Académique de Créteil, France
alain.bihan-poudec@uco.fr, philippe.dutarte@ac-creteil.fr

The statistics curriculum in the French Secondary schools is evolving. According to the French Education Ministry's Instructions, the statistics courses are cursory and essentially descriptive in middle school, while the high school curriculum is oriented towards inferential statistics. However, the limited amount of time available for teaching statistics and the teachers' lack of training lead to the following question: what do students learn about statistics? Our own practice teaching statistics at university and our previous research lead us to put forward a few answers. Overall, undergraduate students in Humanities and Social Sciences associate statistics with mathematics, numbers, calculus, percentages, though interesting exceptions do exist. Regarding the basic notions themselves, it seems that: algorithmic conceptions prevail over the meaning of concepts; and difficulties in learning and understanding statistics are linked with the secondary school curriculum.

INTRODUCTION

The French secondary school system includes:

- The "collège" (French middle school) which teaches one common 4-year curriculum; ages 11-14, the pupils attend their sixth to ninth years of schooling;
- The "lycée" (French high school) can be general, technical or professional; the students reach their tenth, eleventh and, in most cases, twelfth years of schooling.

In this framework, how has the teaching of statistics in secondary school evolved for the last 30 years? According to Henri (2011), the evolution of statistics teaching has been considerable and progressive, and differs according to the level of schooling.

A Three Step Evolution

In the period 1983-1993, middle school statistics was reduced to a few methods of calculation (frequency, mean, median, notion of dispersion) and of graphic representation (histogram, box plot). Data (generally in small quantity) was considered exhaustive. Problems related to sampling and randomness were not treated. In high school, a synthesis of these statistical notions was done and completed by the notion of standard deviation and, in some advance courses, the two-variable statistic. The teaching of probabilities, related to combinatorics and counting, has been disconnected from the teaching of statistics. However, the 1991 11th grade programs introduced a link between statistics and probabilities, through the observation of a "relative stability of an even frequency when the experiment is repeated a large amount of time". We should notice, however, that inferential statistics was introduced in some specific programs in the technical high school.

With the application of the new programs in 2000-2002, we can see an important qualitative change in teaching statistics and probabilities in the second half of the high school curriculum. The general and technological 10th grade program thus stated that "the goal is to make the students think about the nature of the explored data and to make their choice of synthesis based on graphic representations." To think and to make choices in statistics was, in fact, something new at that time in French secondary school teaching. That same program required the teacher to treat "data in a sufficiently large amount so that it justifies a statistical study" and to suggest "study and simulation topics according to the students' interest, the news and their tastes." The sampling fluctuation and the simulation were thus introduced in 10th grade; in advanced science programs in high school senior year (2002), the laws of discrete probability (law of Bernoulli, binomial law) and continuous probability (uniform law, exponential law) were introduced. An initiation to the chi-square test was taught, presented as data fitting an equal repartition law (this is taken off the program ten years later). The programs now in application appeared between 2003 and 2013. Table

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

1 summarizes the evolution of the statistics curriculum for both middle and high school. Under the pressure of training citizens (party related to disappointing results at the PISA international evaluation), the teaching of statistics in middle school was no more limited to a simple methodology but added a dimension of reflection. Besides, the notion of probability was introduced in the course of mathematics in 8-9th grade. To this respect, the 8-9th grade program (2008) became particularly clear: "The mathematics education here meets –through the study of statistics series – the education of citizens: to get used to wonder about the meaning of the number used, about the information brought by a statistics abstract. In the same way, it is to allow the citizen for approaching uncertainty and randomness in a rational perspective that the primary notions about probability are introduced." This introduction of probabilities is especially done through the statistic observation of random experiments.

Table 1. Evolution of statistics teaching knowledge among

	1983-1993	***1993-2003***	***2003-2013***
Middle school ("collège")	1985: Organization and data management. No randomness, no sampling		2008: Statistics reflection. Use of a spreadsheet. Introduction to probabilities in 8-9th grade related to statistical observation
High school ("Lycée ")			
Scientific curriculum	1986: Combinatorics, counting methods to calculate probabilities 1990-1992: Synthesis of the middle school notions. Standard deviation. Probability-statistics link through the observation of frequencies	2000-2002: Sampling, simulation, reflection, statistics, use of real data. Simplified statement of the law of large numbers, modeling. Discrete laws (Bernoulli and binomial) and continuous laws (uniform and exponential). Fitting data to an equal repartition law (particular case of the chi-square test	2009-2012: Exploratory statistics over large data files. Initiation to inferential statistics in the case of a frequency: decision making problem (test) and interval of confidence. Introduction to the binomial law in 11th grade and the normal law in senior year. Withdrawal of the fitting to an equal repartition law
Technical curricula	1989: Teaching of the inferential statistics in superior technical studies		
Professional curricula			2009: Reintroduction of the teaching of probabilities (which had been forgotten) related to statistics. Sampling, simulation. Decision making problems in the case of a frequency (initiation to the hypothesis test)

The high school programs set up between 2009 and 2012, gave a more prominent place to statistics and probabilities than the previous programs (20% to 25% of the allocated time). In descriptive statistics, it is required to use real data under the form of substantial files. Thus the 10th grade program states that "the goal is to make students think about data which is real, rich and diverse (for instance, from a file available at the INSEE), to synthesize the information and to propose relevant representations" (2009). Some notions of inferential statistics are given as soon as

10th grade (interval of fluctuation for decision making, notion of interval of confidence of a proportion) and they are based on the observation of computer simulations. From the 11th grade, the binomial law is used to justify the determination of frequency fluctuation interval and to apply it to the decision making (principle of the statistical test). The introduction of the normal law in senior year allows for a better expression of the fluctuation interval and to think about the question of the interval of confidence

Perspectives

A reform just started (fall 2013) in order to redefine the middle school programs. As this is about obligatory school age, this review is particularly led by the question of the future citizens' statistical literacy. This should lead to a teaching of statistics that is even more data oriented, while following the 4-steps: 1) State a question; 2) Collect data; 3) Analyze, explore data; 4) Interpret the results (*Guideline for Assessment and Instruction in Statistics Education,* GAISE). Moreover, the link between statistics and probability would benefit from being introduced much earlier than it is now (around 15 years old in 9th grade). This evolution would be consistent with what we observed for 30 years in the teaching of statistics in secondary school: the step from teaching calculation techniques to teaching the construction of statistics thought based on data always richer and integrating the randomness aspect on which can be based the data process for decision or estimation (Dutarte, 2011). However, it is far from settled that this content and this way of conceiving statistics are affectively and correctly taught: lack of time, teacher poorly trained to statistics, different approach from classical mathematics (see Raoult, 2013, pp. 59 and 64-65). Teaching is one thing, learning is quite another. As far as we know, there is no systematic survey of statistics skills at the end of the French Secondary School. Failing that, several observations can be made from my experience and research.

UNIVERSITY FRESHMEN

As a teacher in Humanities I work at the "Université Catholique de l'Ouest", in Angers, in Brittany and in French Polynesia. My statistics courses are primarily taken by students majoring in Education. I have nothing particular to highlight with respect to the usual observations made by statistics instructors. Except perhaps that the attitude toward statistics turns out to be boredom and disinterest rather than anxiety or angst. May I tell that teaching statistics tends to become boring, but I have a great interest in the way students learn and understand statistics (I began as a psychologist) Assessments are good, but, following Hubbard (1997), I often ask non-standard questions: then, I observe both conceptual misconceptions and practical mistakes (see Bihan-Poudec, 2010).

Our surveys allow us to deepen these observations. In Angers, we have constituted a research team with academic colleagues, *Statens* (for "Statistiques et Enseignement" / "Statistics & Teaching"). This group contains psychologists, mathematician, biologists and researchers in education. Here some succinct results.

First Survey: Statistics as Calculus

As teachers, we first asked our new students what they know and what they understand about several concepts in statistics: levels of measurement, mode, median, mean, range, standard deviation, etcetera. For each one, the students have to evaluate their degree of ability; furthermore, there were invited to write down how they understood the mean and the standard deviation. This educational concern turned into a research topic because of a) the variability of (mis)conceptions, b) the following paradox: students said they fluently understood the mean but only 22% gave a correct definition (Dubreil-Frémont, Bihan-Poudec & Chevallier-Gaté, 2012). As for the standard deviation, only 4.0% of the answers were correct, 24.4% were mistakes: the rest corresponded to the absence of response… In fact, it seems that statistics is something that a student calculates from a formula.

Second Survey: Statistics Representations

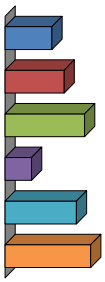
At the very beginning of the freshman year, students were asked to complete a free associations test: "what does the term *statistics* mean to you (use the words which spontaneously

come to your mind)?" They were also asked to specify the program they were following as well as their previous practice of statistics. We emphasize the students have never learned statistics in university before (in Education, first statistics courses begin in the third year).

Table 2. The students and their majors

FORMATIONS

	Nb	%
L1 (first year) Communication	**7**	**12,9**
L1 Education	**9**	**16,1**
L1 Psychology	**13**	**21,3**
L1 Sociology	**4**	**7,5**
L3 (third year) Education	**11**	**19,2**
L3 Education (adult continuing education)	**14**	**23,0**
Tota	**614**	**100,0**



What do we notice? This survey shows that twenty words or so are sufficient to characterize statistics, such as mathematics, percentages, numbers and calculations (Marion & Bihan-Poudec, 2012; Bihan-Poudec, 2013). Better: if we focus on the first two most frequent concepts - mathematics and percentage -, 247 out of 614 students (40%) have mentioned, either one or the other, or both, in their first two associations. Thereby, the association between statistics and mathematics dominates this shared lexical field. The other words mentioned are about statistics and its productions (numbers, calculation, mean, graphics and so on), or about the use of statistics (survey, questionnaire, poll, etc.). This confirms the observation made by Régnier when he writes, "For many of our citizens, statistics is reduced to the statistics and the polls" (2005, p. 9).

Table 3. The first fifteen concepts

Concepts	Frequencies	Percentage of the total number of concepts	Percentage of the total number of students
mathematics	210	8,9	34,2
percentages	204	8,7	33,2
numbers	158	6,7	25,7
calculus	126	5,4	20,5
study	100	4,3	16,3
poll	99	4,2	16,1
average	85	3,6	13,8
data	59	2,5	9,6
graphs	55	2,3	9,0
tables	55	2,3	9,0
survey	46	2,0	7,5
comparison	40	1,7	6,5
analysis	38	1,6	6,2
representation	32	1,4	5,2
diagram	31	1,3	5,0
	1,258	53.6%	
on	2,347	100%	

However, beyond this common speech, differences exist between students in different majors: for some students, the statistics is reducible to statistical indices, the discipline to the

numbers; for others, statistics is conceived as mathematics used in surveys. The statements of other students are distinguished either by the proximity of the statistic to questionnaire methodology, or by their professional use: the reference to mathematics is then significantly reduced.

Moreover, we can add that the affective dimension regarding statistics is barely present in responses given by students, although it is abundantly referred to in the literature: only 22 subjects out of 614 students; 18 responses with negative content, two ambivalent, 2 positive. This set of elements prompted us to move further ahead with a new survey.

Third Survey: A Framework for Statistical Learning

At the beginning of the academic year, 147 others students were asked to respond to a questionnaire. The choice of subjects focused on students in their third year and majoring in Sciences of Education, as they had been taught the same content in the statistics course and had the same teachers: they differed essentially only because of their academic path in pre-service education, for 100 of them, or in adult continuing education (47 of them). The questionnaire, consisted not only of the same association test (a) taken previously, but also questioned the students on:

- Their previous experience with statistics (b),
- Their interest for this discipline (that they had to qualify from passionate to putting off) (c),
- Their participation in the statistics course, whether it was optional and the reasons for their choice (d),
- The conception of statistics, they had to choose from among several definitions (e),
- Their own definition of this discipline (f),
- The interest that students found in it within the framework of their university studies (g), on the one hand and, on the other, for their career (present or future) (h),
- Lastly, some questions of identification: curriculum followed (i), prior experience (j).

In other words, we were trying to find out if possible prior experiences (b), be they academic, be they professional practices, determined the representations of statistics that students have. As a social representation (Moscovici, 1976/1961), statistics, therefore, is studied in its three dimensions: information, the field of representation (items a, e, f) and attitude (c, d, g and h).

At the level of the association test (a), mathematics (numbers, calculations, average, percentages...) appear as vocabulary common to students; however, terms relative to the use of statistics appear more frequently among adult subjects than among young students (this seems corroborated by their use by adults in reports, surveys, etc., in their places of professional exercise).

The analysis of the question of statistics as an option (d) is interesting. If the responses fully justify themselves theoretically as *taking a stand* (Doise, 1985), they inaugurate clear partitions to responses to other questions: a very clear difference between the two populations (i) in favour of a follow-up of courses by students in lifelong education and a refusal by those in pre-service education (chi-square of 32.6, $p < .001$); difference of interest paid to statistics (c) (meaningful with $p < .01$); difference with prior practice (j) (chi-square of 32.6, $p < .05$). This prior practice rather influences the decision not to take this course: a finer analysis would be required to distinguish between school or university experience and professional use, the latter therefore favouring the making of a decision to follow the course.

This difference between the continuing education student and the pre-service education student is just equally manifest with regard to judgement of the relevance of statistics in the framework of studies (g) and in their professional life (h). While all older students estimate that statistics has its place in the program they are following, this percentage drops to 72.4% in the case of young students (chi-square of 6.03, $p < .001$); the situation is analogous for professional use: almost all adults see the usefulness of statistics, only 52% of the young people who responded to the question consider it (chi-square of 20.5, $p < .001$).The analysis of reasons given will enable us to understand this difference.

CONCLUSION

At university, at least in social sciences and humanities, the students arrive with only a few knowledge, skills and understanding about statistics; we don't claim to be exhaustive but, referring to the typology suggested by Ried and Petocz (2002), the students' conception is at the first level: "Statistics is about individual numerical activity activities"; at its best, "Statistics is about using individual statistical methods." Worse, it seems that the mastery of ready-to-use methods does harm to the comprehension of statistical notions (Bihan-Poudec, 2010, p. 92)…

However, to give final conclusions would be premature right now: the reform is in progress, but it will need a suitable training of mathematics teacher to be added to it, so that the students open their mind to the spirit and the methodology of statistics. Also, Garfield,'s suggestion is still actual: "*Less theory, more data; less lecture, more active learning; less calculation, more technology*" (1997, p. 137).

REFERENCES

Bihan-Poudec, A. (2010). L'enseignement de la statistique : en premier lieu, l'apprenant. *Statistics Education Research Journal, 9*(2), 88-103. http://iase-web.org/documents/SERJ/SERJ 9%282%29_Bihan-Poudec.pdf

Bihan-Poudec, A. (2012). Un contrôle surprise… pour l'enseignant! L'évaluation comme révélateur des préconceptions de la statistique chez les étudiants. *Statistique et Enseignement, 3*(1), 63-72. http://publications-sfds.math.cnrs.fr/ojs/index.php/StatEns/article/view/114/103

Bihan-Poudec, A. (2013). *Des chiffres et des êtres. Étude introductive à l'identification de la représentation sociale de la statistique chez des étudiants de premier cycle en Sciences humaines et sociales en France* (PhD en Éducation). Université de Sherbrooke, Québec.

Bihan-Poudec, A., & Marion, J.-M. (2012). *Regards croisés sur des traitements statistiques de données*. Communication aux 44es Journées de la Société Française de Statistique, Université Libre de Bruxelles, Bruxelles (Belgique)

Doise, W. (1985). Les représentations sociales: définition d'un concept. *Connexions, 45*, 243-253.

Dubreil-Frémont, V., Bihan-Poudec, A., & Chevallier-Gaté, C. (2012). *Quelles conceptions de la moyenne ont les étudiants en Sciences Humaines et Sociales ?* Communication au 3e colloque francophone international sur l'enseignement de la statistique, Angers, 12 septembre.

Dutarte, P. (2011). Évolutions de la pratique statistique dans l'enseignement du second degré en France : du calcul statistique au développement de la pensée statistique. *Statistique et Enseignement, 2*(1), 31-42. http://publications-sfds.math.cnrs.fr/index.php/StatEns/article/view/72/61

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE)report: A preK-12 curriculum framework*. Alexandria, VA: American Statistical Association. http://www.amstat.org/education/gaise/GAISEPreK-12_Full.pdf

Garfield, J. (1997). Discussion about "New Pedagogy and New Content: The Case of Statistics". *International Statistical Review, 65*(2), 137-141.

Henri, M. (2011). Évolutions de l'enseignement du secondaire français en statistique et probabilités. *Statistique et Enseignement, 1*(1), 35-45. http://publications-sfds.math.cnrs.fr/index.php/StatEns/article/view/4

Hubbard, R. (1997). Assessment and Process of Learning Statistics. *Journal of Statistics Education, 5*(1).

Moscovici, S. (2004). *La psychanalyse, son image et son public* (3th ed.). Paris: PUF.

Raoult, J.-P. (2013). La statistique dans l'enseignement secondaire en France. *Statistique et Enseignement, 4(*1), 55-69. http://publications-sfds.math.cnrs.fr/index.php/StatEns/article/view/138

Reid, A., & Petocz, P. (2002). Students' conceptions of statistics: a phenomenographic study. *Journal of Statistics Education, 10*(2). http://www.amstat.org/publications/jse/.

Régnier, J.-Cl. (2005). *Formation de l'esprit statistique et raisonnement statistique. Que peut-on attendre de la didactique de la statistique?* Conférence au séminaire de didactique des mathématiques, janvier 2005, 23p.