

UN CONTRÔLE SURPRISE... POUR L'ENSEIGNANT ! L'ÉVALUATION COMME RÉVÉLATEUR DES PRÉCONCEPTIONS DE LA STATISTIQUE CHEZ LES ÉTUDIANTS

Alain BIHAN-POUDEC¹

TITLE

The assessment as an opportunity to discover statistics students' preconceptions

RÉSUMÉ

Ces propos au sujet d'une expérience d'enseignement partent d'une orientation et d'une conviction. Une orientation : inscrire la statistique dans le quotidien des apprenants ; une conviction : être à l'écoute de ces derniers pour optimiser l'enseignement et *in fine* l'apprentissage de la statistique. L'occasion en est donnée par un contrôle auprès d'étudiants en Licence en Sciences de l'Éducation. Les questions ouvertes laissaient voir un arrière-plan de préjugés et de conceptions erronées. Ainsi, les *verbatim* font-ils apparaître déni de la variabilité dans les échantillons, nécessité de données quantitatives pour évaluer, croyance dans la nécessité d'une précision maximale, etc. Ces observations confortent la nécessité pour l'enseignant d'identifier les conceptions *clandestines* des étudiants.

Mots-clés : enseignement de la statistique, littératie statistique, obstacle épistémologique, pédagogie universitaire.

ABSTRACT

This teaching experience was conducted with two aims: to introduce statistics in students' everyday situations; to be attentive to their talk in order to increase teaching and therefore learning. An example is given by an assessment test carried out with a group of third year students studying Educational Sciences. The opened questions show pre-conceptions and misconceptions: the denial of the variability in samples, the need for quantitative data for evaluating, the belief in the need for accuracy, etc. These *verbatim* reinforce the need for teachers to identify students' prior conceptions.

Keywords: misconceptions; statistical literacy; statistics education research; teaching in university; teaching statistics.

1 Introduction

Les objectifs de l'enseignement de la statistique dépendent assurément du cursus dans lequel il s'inscrit, mais un consensus semble s'établir quant à celui d'acquérir un mode de pensée statistique (Rumsey, 2002 ; Ben-Zvi et Garfield, 2004) : il s'agit notamment de doter les personnes de compétences à décrypter les informations qu'ils reçoivent, et ce afin de favoriser leur esprit critique. En bref, il s'agit de former un *statistical citizen* (Rumsey, *ibid.*, p. 1).

¹ LUNAM Université, Université Catholique de l'Ouest, Département de recherche en Sciences Humaines et Sociales, Angers, France, alain.bihan-poudec@uco.fr

Un contrôle surprise... pour l'enseignant !

La *littératie statistique* (*statistical literacy*) a été l'objet de communications (telle la journée d'études de la SFdS sur Angers, voir par exemple Padieu, 2006), d'articles, de livres. Ce terme peut être défini ainsi :

« The term 'statistical literacy' refers broadly to two interrelated components, primarily (a) people's ability to interpret and critically evaluate statistical information, data-related arguments, or stochastic phenomena, which they may encounter in diverse contexts, and when relevant (b) their ability to discuss or communicate their reactions to such statistical information, such as their understanding of the meaning of the information, or their concerns regarding the acceptability of given conclusions. These capabilities and behaviors do not stand on their own but are founded on several interrelated knowledge bases and dispositions. » (Gal, 2002, p. 2-3)²

Au niveau des modalités de formation à la littératie statistique, Gattuso préconise une proximité des données utilisées lors des cours avec les informations présentes dans le quotidien. En effet, selon elle :

« L'enseignement de cette discipline ne peut se limiter à la transmission de techniques de calcul, mais doit mettre l'accent sur la compréhension de concepts et l'acquisition d'une pensée statistique à travers la pratique des méthodes utilisées dans cette discipline. [...] Trop souvent, les statistiques sont présentées hors de leur contexte "normal", c'est-à-dire les données réelles, et plutôt comme une procédure que comme une analyse globale d'une situation à l'aide d'outils statistiques. [Or] il est indispensable que les élèves soient en mesure d'apprécier le rôle du raisonnement statistique qui permet de lire et interpréter des phénomènes réels sans vouloir apporter des réponses exactes mais en donnant toutefois la possibilité d'avoir une vision de la réalité qui ne soit pas que subjective. » (Gattuso, 2003, p. 2)

Qu'en est-il réellement ? Si telle est l'intention de l'enseignant, quelles garanties a-t-il quant à l'acquisition effective de cette littératie statistique ? Ainsi, Hubbard (1997) montre bien comment il est difficile pour l'enseignant de reconnaître chez l'apprenant la maîtrise ou non de la démarche statistique.

Incidentement, un contrôle en statistique auprès d'étudiants en Licence en Sciences de l'Éducation peut être l'opportunité d'aborder ces questions.

2 Présentation du contexte

C'est dans cet esprit de proximité avec les données de tous les jours que nous avons proposé aux étudiants un contrôle qui partait de la capture du classement sur internet d'un film. Cette centaine d'étudiants, répartis sur deux sites, suivaient la Licence en Sciences de

² « Le terme de "*littératie statistique*" se réfère de façon générale à deux aspects interdépendants, d'abord a) la capacité des gens à interpréter et à évaluer de façon critique l'information statistique, les arguments liés aux données ou les phénomènes aléatoires, qu'ils peuvent rencontrer dans différents contextes, et le cas échéant (b) leur capacité à discuter ou à communiquer leurs réactions au regard de ces informations statistiques, telle leur compréhension de la signification de l'information, ou leur vigilance quant à l'acceptabilité des conclusions formulées. Ces capacités et comportements ne viennent pas de soi mais se fondent sur plusieurs dispositions et connaissances de base interdépendantes. » (trad. personnelle)

A. Bihan-Poudec

l'Éducation, formation qui n'existe en France que sous la forme d'une troisième année, et qui, pour ce qui nous concerne, accueille principalement des étudiants venant d'Information et Communication, de Sociologie et de Psychologie, où ils bénéficient d'un parcours en éducation. S'y adjoignaient quelques BTS et DUT³. Hormis les étudiants de Sociologie et de Psychologie à qui ont été enseignées la statistique descriptive, la statistique inférentielle et la pratique des tests, les étudiants n'ont eu pour enseignement de la statistique que les éléments présents lors de leurs études dans le secondaire. Aussi, le programme de statistique est-il uniquement introductif à cette discipline et se résume à la statistique descriptive.

2.1 Le texte du contrôle

L'idée de ce contrôle vient de la fréquentation sur internet des sites dédiés au cinéma. Outre les horaires des séances, figurent souvent les avis et commentaires des spectateurs, ainsi que le classement de ces films sous forme d'attribution d'étoiles. Ce dernier point nous a interpellé car nous ne savons pas *a priori* si le nombre d'étoiles mis en exergue désigne l'avis de la plupart des spectateurs (et s'apparenterait ainsi au mode), ou celui de la moitié d'entre eux (médiane), ou enfin la moyenne de leurs avis. C'est cette interrogation que nous avons partagée avec nos étudiants. Voici le texte du contrôle :

« La statistique descriptive permet à son utilisateur d'organiser ses données, de les résumer et d'en tirer une vue d'ensemble. Elle devrait permettre aussi d'avoir un regard critique sur les informations que l'on reçoit.

Le sujet ci-dessous se situe dans cette perspective et prend comme point de départ une information banale telle que vous pouvez la trouver dans des quotidiens ou sur des sites Internet dédiés au cinéma. Ainsi avez-vous souvent, à côté des films présentés, une évaluation faite par les spectateurs comme celle-ci :



Madagascar 2

réalisé par Eric Darnell, Tom McGrath

avis des spectateurs ★★★★★

Animation (1h 35min)

Il en ressort que le film a reçu quatre étoiles sur cinq possibles. Toutefois, pour un lecteur averti, cela n'est pas sans poser de problème car il ne sait pas si cet indice correspond à un mode, une médiane ou une moyenne.

Question 1 : quelle information chacune des caractéristiques de position mentionnées ci-dessus apporte-t-elle ? Montrez-en les différences.

³ BTS (Brevet de Technicien Supérieur) et DUT (Diplôme Université de Technologie) sont deux diplômes français qui sanctionnent des formations professionnalisantes préparées en deux années après le baccalauréat.

Un contrôle surprise... pour l'enseignant !

En fait, les quatre étoiles ont été obtenues à partir de la distribution suivante :

Nombre d'étoiles	Nombre de spectateurs
2	6
3	8
5	20

Tableau 1 : distribution A

Question 2 : déterminez-en les caractéristiques de position (nous supposons qu'il existe une unité de mesure constante).

Un autre site fournit la distribution suivante (B) pour le même film :

Nombres d'étoiles	Nombre de spectateurs
0	6
1	1
5	27

Tableau 2 : distribution B

Question 3 : là aussi, déterminez les caractéristiques de position. En les confrontant avec celles que vous avez obtenues à la question 2, que constatez-vous ? Quelle évaluation (nombre d'étoiles) figurerait sur le site Internet ? Si quelqu'un vous disait que les sites donnent la même information quant à l'évaluation du film, que lui répondriez-vous en prenant en compte les distributions ?

Question 4 : déterminez les caractéristiques de dispersion pour les deux distributions (A et B). Quelles conclusions pouvez-vous en tirer de leur comparaison ?

Question 5 : la plupart des sites calculent un nombre moyen d'étoiles. Ce faisant, ils supposent qu'il existe une unité de mesure constante, l'étoile. D'après vous, est-ce justifié ? À partir des caractéristiques de position et de dispersion que vous avez déterminées pour les deux distributions (A et B), le fait de considérer l'évaluation comme quantitative apporte-t-elle un plus par rapport à une évaluation conçue comme ordinale ?

Un troisième site fournit les éléments suivants. Madagascar : avis des spectateurs



Nombres d'étoiles	Nombre de spectateurs
1	6
2	8
4	20

Tableau 3 : répartition des avis des spectateurs (distribution C)

A. Bihan-Poudec

En sont déduits les éléments suivants :

<i>Caractéristiques de position</i>	mode : 4 étoiles	médiane : 4 étoiles	moyenne : 3 étoiles	
<i>Caractéristiques de dispersion</i>	étendue : entre 1 et 4 étoiles, soit un écart de 3	écart interquartile : entre 2 et 4 étoiles, soit un écart de 2	écart moyen : 1,18 étoile	écart type : 1,24 étoile

Question 6 : en comparant ces indices avec ceux de la distribution A, quelles conclusions pouvez-vous tirer ?

Question 7 : à l'issue de ce contrôle, quelle remarque finale pouvez-vous faire sur la validité des évaluations des films par les spectateurs ? Quelles améliorations pourriez-vous suggérer ? »

2.2 Remarques

Rien ne permettant sur le site de dire si les quatre étoiles obtenues par le film *Madagascar 2* correspondaient au mode, à la médiane ou à la moyenne des avis des spectateurs, il était demandé aux étudiants de différencier ces dernières quant à leurs sens et propriétés. Il s'agissait de voir si, au-delà des calculs pour obtenir ces indices, l'étudiant en comprenait le sens.

Était soumise ensuite aux étudiants une distribution pour laquelle ils devaient déterminer les indices de position ; une distribution différente leur était alors soumise pour laquelle ils devaient opérer les mêmes calculs. Les résultats ne différant pas, il leur était demandé si les sites donnaient toutefois la même information, si les distributions étaient identiques. L'objectif était de valider leurs compétences à calculer les indices de position mais aussi de les amener à s'apercevoir que ces indices ne résumaient que partiellement les distributions, étaient insuffisants à rendre compte des différences observables entre les deux sites.

Les étudiants étaient invités alors à déterminer les indices de dispersion (qui, eux, différaient) et à nouveau à comparer les distributions.

Une cinquième question interrogeait le présupposé fait par les sites de considérer l'étoile comme étant une unité de mesure constante. Les étudiants pouvaient y répondre théoriquement (notion de niveau de mesure) et aussi pragmatiquement, en notant que les conclusions ne différaient pas que le classement du film soit conçu comme relevant d'une échelle ordinale ou envisagé comme étant une échelle de rapport.

Ensuite, une troisième distribution était proposée, déjà caractérisée : elle ne différait de la première que par les indices de position (translation d'une unité). L'idée était au final que l'étudiant se rende compte qu'indices de position et indices de dispersion devaient être indissociables. N'en déterminer qu'une catégorie exposait l'étudiant à des conclusions erronées, car partielles.

Enfin, une dernière question les invitait à faire part de leurs remarques quant à la validité des évaluations et à suggérer des améliorations. L'espoir de l'enseignant était que ce détour par la démarche statistique renouvelait le regard de l'étudiant sur des données du quotidien et le convainquait de l'utilité de la discipline.

3 Quand les étudiants écrivent

Globalement, les résultats se sont avérés tout à fait corrects, les calculs rentrant pour 70% dans la notation. Mais des remarques éparses ainsi que les réponses aux questions 6 et 7 nous ont interpellé par leur teneur et assurément aussi parce qu'elles étaient inattendues. Trente-deux copies ont ainsi été sélectionnées ; après lecture et identification des thématiques, les réponses ont été classées en quatre principales catégories (même si parfois des liens apparaissaient entre elles).

TABLEAU 1 – Répartition des fragments de discours⁴ des étudiants selon leur catégorisation

Catégories	La place de la subjectivité	La quantification	La variation comme erreur	L'échantillonnage
Fragments retenus	E1b, E3, E7a, E7b, E9, E11b, E13b, E14b, E16b, E19a, E21, E22B	E1d, E3a, E6a, E7b, E11a, E13b, E16b	E1a, E4a, E6a, E13c, E12a, E15a, E16a, E20a, E22a, E22c, E24a	E8c, E12a, E13a, E20a

3.1 La place de la subjectivité

Une critique récurrente réside dans la dimension *subjective* de l'évaluation du film : « *chaque spectateur a sa subjectivité, son avis qu'il est difficile de quantifier en nombre d'étoiles* » (E1b). Elle se précise : « *on ne peut savoir ce que représente une étoile dans l'esprit des spectateurs* » (E3a), « *pourquoi opter pour une étoile plus que pour 2 étoiles ?* » (E6a), ou encore « *on ne connaît pas la valeur réelle d'une étoile. On ne sait pas à quoi elle correspond vraiment* » (E19a). Un étudiant (E9, fragment b) n'écarte pas le risque que pour le même avis deux spectateurs recourent à deux modalités distinctes.

Ces remarques sont fondées et questionnent même l'idée de mesure (propriétés 1 et 2 du tableau 2). Cependant, les suggestions d'amélioration du mode de recueil des avis montrent que les étudiants ne vont pas jusqu'à y renoncer, mais s'orientent vers deux directions diamétralement opposées comme nous le verrons.

TABLEAU 2 – Propriétés des différents niveaux de mesure (cf. Stevens in Tiberghien, 1984, p. 57-59)

Propriétés de mesure	À chaque individu correspond une modalité et une seule	Les modalités sont distinctes	Les modalités sont ordonnées	L'écart entre modalités a un sens	Le rapport entre modalités a un sens
Niveau de mesure					
Échelle nominale	oui	oui	non	non	non
Échelle ordinale	oui	oui	oui	non	non
Échelle d'intervalles	oui	oui	oui	oui	non
Échelle de rapports	oui	oui	oui	oui	oui

⁴ E suivi d'un nombre correspond à l'identifiant de l'étudiant, la minuscule indiquant le fragment de la copie de cet étudiant. Ainsi E1b indique que le fragment retenu est le deuxième (b) de la copie de l'étudiant E1.

3.2 La quantification

Une des questions du contrôle interrogeait le statut de l'étoile comme mesure : correspond-elle à une unité de mesure constante ? Si oui, nous entrerions dans le domaine du quantitatif ; si non, nous serions en présence d'une échelle ordinale. La plupart des étudiants ont souligné le paradoxe selon lequel ils ne sont pas assurés d'avoir une unité de mesure constante mais que le quantitatif leur offrait la possibilité de calculer des indices de dispersion plus nombreux (écart moyen, écart-type) que s'ils n'avaient affaire qu'à un cas qualitatif. Toutefois, aucun n'est allé jusqu'à reconnaître que la conclusion de différence en hétérogénéité pouvait être obtenue avec l'étendue seulement. Au travers des suggestions d'amélioration estudiantines, plusieurs positions se dégagent :

1. remplacer les étoiles « *par leurs significations* » (E11a), par des mots comme « *j'aime assez bien* » ;
2. remplacer les étoiles par des points (E3a), des notes (E13b, E25a), voire des classes de notes (E21a) ; un argument rencontré est « *afin que tout le monde sache à quoi s'attendre* » (E25a) ;
3. établir une grille de critères (E1d) tels les effets spéciaux, l'intrigue, etc., quitte à « *établir un coefficient d'importance à chaque critère pour enfin faire une moyenne et obtenir un chiffre décimal* » (E16b).

Plusieurs remarques peuvent être faites. L'orientation vers une mesure ordinale semble reposer sur un glissement sémantique : « *la qualité du film me semble plus important que de quantifier à l'aide d'étoiles* » (E7b). La tendance à aller vers le connu (la notation) est présentée comme si le problème de la mesure se trouvait alors résolu (de manière illusoire, cf. les travaux en docimologie). Quant au recours au quantitatif, il paraît être accompagné par la nécessité d'avoir une pensée analytique et une mesure précise : *quid* alors d'un avis global du spectateur ?

3.3 La variation comme erreur

Plusieurs citations peuvent illustrer cette thématique :

- « *L'écart-type est en fait un indicateur de la "justesse" de ce que reflète en réalité ce nombre d'étoiles. Plus l'écart est grand, plus il y a d'écart entre ce qui est dit et la réalité.* » (E1a)
- « *Ce jugement [troisième distribution] est moins erroné que le premier. Nous avons le même nombre d'étoiles mais les écarts sont beaucoup moins importants (...). Il reflète plus la réalité.* » (E13c)
- « *Il [l'écart-type] ne doit pas être trop élevé pour signifier que la moyenne est correctement établie* » (E16a)
- « *On peut dire que d'une distribution à l'autre l'estimation n'est pas fiable. L'évaluation rend bien compte de l'ensemble de ces estimations mais la seconde distribution comprend des écarts trop importants entre les estimations... Les indices de dispersion éclairent donc les caractéristiques de position et confortent leur mauvaise fiabilité dans le cas présent.* » (E22a et c)
- « *Entre 2 et 4 étoiles existe un écart de 2 étoiles, l'unité de mesure étant une erreur [c'est nous qui soulignons], le rapport entre modalités a un sens.* » (E24a)

Un contrôle surprise... pour l'enseignant !

Tout se passe comme s'il existait une valeur vraie, réelle, telle que les écarts à cette valeur soient des erreurs. Comme s'il ne pouvait pas y avoir des appréciations différentes pour un même film !

3.4 L'échantillonnage

Cette manière de concevoir la variation n'est pas sans conséquences sur la manière de comparer les sites. Ainsi, en observant que deux distributions ont les mêmes caractéristiques de dispersion mais diffèrent quant à leur distribution, est-il conclu que cela « *prouve bien que ces évaluations ne peuvent pas refléter [sic] la qualité des votes* » (E8c). Ou encore que le « *jugement [dans la troisième distribution] est moins erroné que le premier* » (E13c) au motif que les caractéristiques de dispersion y sont moins élevées. Un étudiant suggère – avec pertinence – de regrouper tous les avis des différents sites mais c'est avec l'espoir que « *la note précise pourra être donnée* » (E20a). En d'autres termes, l'« idéal » des étudiants serait l'existence d'une seule valeur, vraie, réelle, absolue. Face à la variabilité des échantillons, soit ils « jettent le bébé avec l'eau du bain » en concluant à l'impossibilité de mesurer, soit en valorisant les échantillons ayant une moindre dispersion, les autres devenant « suspects ».

4 Discussion

Si notre analyse des *verbatim* présents dans des copies d'examen amène à identifier des éléments quant aux préconceptions des étudiants, il ne faut pas écarter le risque que cette analyse amène une cohérence là où elle n'existe peut-être pas. Bref, de construire artificiellement le discours d'un étudiant virtuel, gommant ainsi la possible divergence de conceptions. Mais la cohérence à l'intérieur des copies et entre les copies atténue cette critique.

Un point intéressant est la conception de la variation comme erreur. Vincke et Depiereux (2010) y voient un obstacle épistémologique et notent que la première perception statistique est la moyenne ; que la variabilité est d'abord conçue « comme une erreur, et non comme une manifestation de la diversité naturelle »⁵. Toutefois, si cette conception est ici inopportune (les spectateurs de *Madagascar 2* ayant bien le droit d'avoir des avis différents), il n'en est pas moins vrai qu'historiquement elle est sans doute première : Boyé et Comairas (2002) nous rappellent que les développements de la statistique doivent beaucoup à l'astronomie et aux problèmes de mesure que celle-ci rencontrait au XVIII^e siècle : les observations d'un même phénomène ne donnaient pas les mêmes résultats. Il convenait alors de chercher un indice, un modèle qui rendait compte au mieux de ces irréductibles erreurs. Ceci aboutit notamment à la distinction que fit Quételet entre moyenne « objective » des mesures d'un individu unique d'une part, et, d'autre part, moyenne « subjective » concernant, elle, les mesures de plusieurs individus. Il n'y a donc pas en soi obstacle épistémologique mais utilisation inappropriée d'une conception de la variation au vu du contexte. Dans le cas du contrôle, la variation est (parfois) vue comme une erreur, comme du bruit parasite, alors qu'elle est le signe d'une variabilité des avis des spectateurs au regard d'un film. Il serait intéressant de voir si cette conception historiquement première l'est aussi au niveau individuel, suivant ainsi la

⁵ Ce qu'un étudiant formule clairement dans un autre contrôle où il s'agissait de calculer moyenne et écart-type à partir de la distribution du nombre d'enfants chez une population singulière, les personnes n'ayant pas de télévision (Bergier, 2010) : la moyenne étant de 1,93 enfant, « *l'écart-type est de 1,65 enfant. Les questionnaires de ce genre ne sont pas totalement fiables parce qu'il y a une erreur de 1,65.* »

A. Bihan-Poudec

suggestion de Piaget et de Garcia (1983), ou si cela est le fruit des enseignements de statistique antérieurs : il serait paradoxal – et redoutable – de s’apercevoir que des étudiants ayant suivi des cours sur la statistique inférentielle ne considèrent la dispersion que comme erreur d’échantillonnage. Avec ce recours systématique, nous aurions là une conception partielle de la variation, soit persistante, soit acquise rétroactivement !

Partir de données du quotidien apparaît donc utile pour faire émerger des conceptions latentes ; cela l’est sans doute pour que les étudiants s’approprient les notions de statistique mais il faut reconnaître que cette confrontation ne va pas de soi, comme si cette discipline, pour eux, se devait d’être reléguée comme un savoir strictement scolaire, sans rapport avec le quotidien : les réactions des étudiants à propos de ce contrôle – tout au moins celles dont il nous a été fait état – sont passées au mieux du sentiment d’incongruité à celui de surprise, voire d’interpellation⁶. Il est du chemin à faire entre les situations-problème et la problématisation !

Nous voulions surtout montrer qu’une simple évaluation pouvait être riche d’enseignements pour le professeur. Sans qu’il ne s’agisse ni d’une expérimentation, ni d’un travail sur les conceptions de la variation chez les étudiants⁷. À la limite, il s’agirait là de l’inverse. Une expérimentation implique une hypothèse ; une recherche, bien souvent, part d’un corpus déjà constitué qu’il s’agit d’enrichir ou d’affiner. Une évaluation peut, à certaines conditions, ouvrir à l’inattendu, à l’in-ouï. Sans doute, classiquement, une épreuve de validation est-elle autoréférentielle : nous cherchons à voir si les compétences attendues des étudiants et pour lesquelles nous avons œuvré, sont bien présentes. Comme le note Hubbard : « *Any responsible instructor should assess what they have been asking their students to learn and in the way they envisage the students have learned* » (*op. cit.*, p. 1). Pour dépasser cela, sans le dénier toutefois, il s’agit d’aménager dans les contrôles de la place à de l’inattendu pour avoir de l’inédit : simplement en formulant des questions naïves⁸, en faisant appel au jugement de l’étudiant, en y aménageant des questions ouvertes.

Quoi qu’il en soit, les observations précédentes montrent bien que le challenge de l’enseignement de la statistique ne se limite pas à pourvoir les apprenants de la maîtrise de procédures mais passe par l’élucidation de conceptions, fussent-elles clandestines (cf. Quilici et Mayer, 1996, p. 144) : comprendre ne se résume pas à réussir (Piaget, 1974). Ces observations montrent aussi, à l’instar de Shaughnessy (2006), que les étudiants eux-mêmes sont une source inestimable pour enseigner.

Références

- [1] Barr, G. V. (1980), Some student Ideas on the Median and the Mode, *Teaching Statistics*, 2, 38-41.
- [2] Ben-Zvi, D. et J. Garfield (dir.) (2004), *The challenge of developing statistical literacy, reasoning and thinking*, Kluwer Academic Publishers, Dordrecht, Boston, London.

⁶ Nous retrouvons le phénomène d’*isolation cognitive* identifié par ailleurs (Bihan-Poudec, 2010, p. 93-94).

⁷ Le lecteur pourra à profit se reporter aux numéros des années 2004 et 2005 du *Statistics Education Research Journal*, <http://www.stat.auckland.ac.nz/~iase/publications.php?show=serjarchive>.

⁸ Ainsi Barr (1980), dans un bref questionnaire remis à ses étudiants, a-t-il eu la surprise de voir que le mode n’était selon eux ni un indice de position, ni un indice de dispersion !

Un contrôle surprise... pour l'enseignant !

- [3] Bergier, B. (2010), *Pas très cathodique. Enquête au pays des « sans télé »*, Érès, Toulouse.
- [4] Bihan-Poudec, A. (2010), L'enseignement de la statistique : en premier lieu, l'apprenant, *Statistics Education Research Journal*, **9**(2), 88-103, http://www.stat.auckland.ac.nz/~iase/serj/SERJ9%282%29_Bihan-Poudec.pdf
- [5] Boyé, A. et M.-C. Comairas (2002), Moyenne, médiane et écart-type. Quelques regards sur l'histoire pour éclairer l'enseignement des statistiques, *Repères*, 48, 27-39, <http://www.univ-irem.fr/commissions/reperes/consulter/48boye.pdf>
- [6] Gal, I. (2002), Adults' statistical literacy: Meanings, components, and responsibilities, *International Statistical Review*, **70**(1), 1-25.
- [7] Gattuso, L. (2003), *Les statistiques, un élément essentiel de la littéracie. Une expérimentation d'enseignement des statistiques dans les écoles italiennes*, XXXV^{èmes} Journées de Statistique, Lyon, 2-6 juin, http://www.sfds.asso.fr/groupe/jslyon03/35JS_Lyon_Linda_Gattuso.pdf
- [8] Hubbard, R. (1997), Assessment and Process of Learning Statistics, *Journal of Statistics Education*, **5**(1), <http://www.amstat.org/publications/jse/v5n1/hubbard.html>
- [9] Padieu, R. (2006), *Pour une évaluation de la culture statistique*, Communication à la Journée du Groupe "Enseignement de la statistique" de la Société Française de Statistique, Université Catholique de l'Ouest, Angers, 22 mars.
- [10] Piaget, J. et alii (1974), *Réussir et comprendre*, PUF (coll. Psychologie d'aujourd'hui), Paris.
- [11] Piaget, J. et R. Garcia (1983), *Psychogenèse et histoire des sciences*, Flammarion (coll. Nouvelle bibliothèque scientifique), Paris.
- [12] Quilici, J. L. et R. R. Mayer (1996), Role of examples in how students learn to categorize statistics word problems, *Journal of Educational Psychology*, **88**(1), 144-161.
- [13] Rumsey, D. J. (2002), Statistical literacy as a goal for introductory statistics courses, *Journal of Statistics Education*, **10**(3), <http://www.amstat.org/publications/jse/v10n3/rumsey2.html> database
- [14] Shaughnessy, J. M. (2006), *Student work and student thinking: An invaluable source for teaching and research*, 3rd International Conference on Teaching Statistics.
- [15] Tiberghien, G. (1984), *Initiation à la psychophysique*, PUF, Paris.
- [16] Vincke, G. et E. Depiereux (2010), Mutation d'un cours de biostatistiques : évaluation d'un dispositif d'autoformation, *Revue internationale des technologies en pédagogie universitaire/International Journal of Technologies in Higher Education*, **7**(3), 6-18, <http://www.ritpu.org/spip.php?rubrique61&lang=fr>