Improving the Diagnostic and Didactic Meaningfulness of Mathematics Assessment in France

Antoine BODIN

IREM - University of FRANCHE-COMTE - FRANCE

Annual Meeting of the American Educational Research Association AERA - NEW-YORK April 10, 1996

This paper has been given in a symposium titled : Developing Valid Large Scale Mathematics Performance Assessments That Provide Diagnostic Information (AERA Divison D Symposium)

Symposium coorganised and shared with :

Gillian CLOSE, Margaret BROWN, Mike ASKEW - University of London
Marja Van Den HEUVEL-PANHUIZEN - University of UTRECHT

Objectives of the research : better understand what we are assessing and how assessments might be improved (see part 3 below).

Purposes : link the results gathered in large scale assessment to the current developments in research in mathematics education to use them as tool in assessment purposes (see part 3 and 4).

Methods and techniques : those belonging to the *Didactique* of mathematics field as well as measurement and classificatory ones (*implicative* analysis and IRT) (see part 4).

Data sources : national large scale assessments (see part 2).

Educational goals : helping teachers to become more professional and particularly to change from making intuitive judgements about students' knowledge to making informed judgements (see part 1).

Scientific interest : linking two separate research fields : research in evaluation and measurement on the one hand and research in mathematics education on the other one (see part 4).

****

The research presented has been undertaken within a specific educational context and leans largely on specific large scale assessment settings. It also makes large use of ideas, concepts, and results developed and obtained around the so called "French School of *Didactique* of Mathematics"

For a better understanding of the presentation it seemed more appropriate in this paper to give greater emphasis to part 1 and part 2 : the context of the research and the description of the large scale assessments involved. Meanwhile it is part 3 which is aimed to be the core of the presentation.

1 - Context of the research

In the educational field at least, France has constantly resisted using and spreading measurement ideas, theories, and tools. The tradition is that most assessment consists of marking classwork, with the underlying idea that where teaching is good, assessing is not an issue.

Reasons may be found on the teacher side as well as on the measurement one. It is well acknowledged that, in France at least, teachers would not allow being deprived of any part of their right to judge students independently. Much empirical research supports this view, showing holistic teachers' assessments more reliable than usual marking methods.

Another reason i the fact that, for a long time, research focus has been put on reliability (from PIERON (1930) to NOIZET (1980)...). The name coined for that research field (*docimologie*) addresses the only measurement issue teachers have been usually aware of and that contributed to move them from other measurement interests.

It seems useful to recall that, in France, MCQ formats have been steadily discarded for assessment purposes. They have meanwhile made recent but still rare appearances in some mass examinations as well as in some competitions (i.e. : KANGOUROU), but until now the baccalaureate exam has resisted (and it is well acknowledged that the French educational system is Baccalaureate driven !).

2 - New assessment needs identified

In the late 80s, apart from casual classroom assessments and official examinations, needs and room for other types of assessments became clear and more likely to be accepted by teachers. Two goals were identified : the first one addresses needs for evaluating the curriculum, the second one addresses the possibility of providing teachers with assessment tools likely to help them diagnosing students' needs. Independently organized, but linked in several ways, two National assessment settings are providing important information and are acknowledged to change teachers' assessment practices and conceptions. At the same time new research questions have emerged as well as needs for ways to better understand what is ultimately at stake when assessing (what does such or such observed fact really mean ?)

First have a glance at those two settings (see chart) :

1 - The EVAPM observatory

> For about 10 years, the French Association of Mathematics Teachers (APMEP), associated with several research institutes (IREM,...), and with several official bodies, has undertaken a series of large scale assessments at secondary level (grade 6 to 12). The first aim has been the evaluation of a new national mathematics curriculum (implemented, one year at a time, from 1986 to now). The second has been to provide teachers and officials with valid students' achievement indicators as well as with "good" assessment items and designs (the meaning of*good* has evolved and is exemplified below, and in more details in the EVAPM reports).

2 - The DEP assessments

> Each year, since 1989, the Evaluation and Future Planning Department in the Ministry of Education (DEP) has organized diagnostic tests for all the 3th, 6th, and 10th grade students. The tests are accompanied by teacher training in use of the results. No individual marks are given but students' profiles are issued to encourage teachers to individualize their teaching and to trace more closely each personal student's development.

On the whole, these settings have been providing us with lots of data, reports, analyses and findings, as well as directions for further research. An computerized item bank (EVAPMIB) contains all items, results obtained in large scale studies and further analysis. It is now used for developing new tests and for math teachers' and math-inspectors' training (see example figure 1 and 2 : EVAPMIB items out of about 2 000 items available).

This presentation aims to give more information on what has been drawn from those studies concerning the overall quality of assessment items and tests and on the methods used (or presently developed) to assess that quality.

3 - Research questions and findings

The main questions address the meaning of students' observed answers to assessment items along with the possibility to improve the usefulness of the information gathered in assessments by better designing assessment items and tests.

But usefulness for what purpose ? First to better assess the state of students' personal knowledge, second to gain better ideas on how one might help to improve that knowledge. The first issue here is to assure internal validity.

That may sound like an impossible dream : being sure that what you are looking at in an assessment fits the best observation choice you could have planned according to what you have to decide, and at the same time be informed of the profound meaning of what you are observing.

But this general idea quickly splits into more acceptable research questions : what kind of dependencies do exist between observed facts ? Does that mean anything for subsequent knowledge development ? statistically or compellingly ? What is the effect of items context, vocabulary, sentence structure, mode of testing,... ? The final but not easier problem is to relate the results to the foundamental issues.

4 - Methods and theories involved

First initiated in France, but more and more spread and developed in other countries, the so called "Research in Didactique of Mathematics" field provides us with concepts and models that have proved their usefulness in much research on Mathematics Education. In assessment matters, as well as in other educational topics, we surely have to make "a *step aside*" (CHEVALLARD) and cease confusing observations with reality. For that the use of diverse didactic concepts as intermediate tools has proved valuable.

It would be necessary here to make specific references to some "didactic" concepts such as : "didactic variables", "didactic contract" (BROUSSEAU, G., 1986), "conceptual field" (VERGNAUD, G., 1990), "didactic transposition" (CHEVALLARD, Y., 1985),...

Several complementary methods are used to analyze assessment items and tests :

Didactic Item Analysis

The Didactic Item Analysis Method tries to trace the different meanings it is possible to draw out of assessments on epistemological and *didactic* points of view. The systematic tracing of some given items across grades helps in this way, but it is far from being sufficient.

The method proposed is aimed to link the observation made with what has been taught and with the context in which the teaching takes place (curricular aspects). It also takes into account change expected or observed in results when assessment items are modified without altering their formal content.

One of the item qualities is their potential to reveal something about the kind of relation the student has with knowledge. Knowledge type and quality are themselves described using current theories about teaching and particularly about math teaching.

In addition, the Implicative Analysis method (GRAS R, 1979), as well as Item Response Curves and IRT parameters, are currently used to detect items of interest, item bias, information provided by particular items,...(see example figure 3). Many items that may cause problems when aggregated for scaling purposes provide useful information about the student's state of knowledge. Such items can also be useful for exploring the configuration and boundaries of conceptual fields.

On the whole, the general idea of using mathematics items as revealers of the type and the quality of students' acquired knowledge seems to work.

References

BODIN, A. 1993. What does to assess mean : investigations into assessment in mathematics education. In, An ICMI study, ed M. NISS. Dordrecht : Kluwwer Academics Publishers

BROUSSEAU, G. 1986. Fondements et méthodes de la didactique des mathématiques. Recherches en Didactique des Mathematiques, 7/2

DIDAY, E., LECHEVALIER, Y., SCHADER, M., BERTRAND, P., BURTTSCHY, B. Eds.) : (1995) : New Approaches in Classification and Data Analysys - Springer-Verlag

DOUADY, R. - MERCIER, A. (1992) - Research in "didactique" of Mathematics - selected papers - La pensée Sauvage, Publisher

FREDERIKSEN N., MISLEVY R. J., BEJAR I. I. editors (1993) : Test Theory for a new generation of tests - Educational Testing service - Lawrence Erlbaum Associates

GRAS, R. (1992) : L'analyse des données : une méthodologie de traitement de questions de didactique. Recherches en Didactique des Mathematiques, 12/1

ROMBERG, T. A. ; STEWART, D.M. (1988) : The Monitoring of School Mathematics - Wisconsin Center for Education Research.

VERGNAUD, G. (1990). La théorie des champs conceptuels. Recherches en Didactique des Mathematiques, 12/1

VIGOTSKY L. S. (1935) - Interaction between learning and development - Mind in society (1968) - Harvard University Press.

NOLAN K. (1994) : Mathematics in France, 1994 - in international bentchmarking, New Standards Project - University of PITTSBURG.

HERBST, P. (1994) : The current curriculum for secondary chool in France. Research paper - University of ATHENS - GEORGIA

AFT/NCISE (1994) : A profile of the French Education System, What College-bound students abroad are expected to know about... American Association of teachers

DEP (French Ministry of Education), (1989 through 1996) : series of reports published in "Éducations et formations"

APMEP, (1987 through 1996) : series of EVAPM reports

*****Antoine BODIN is one of the designers of the EVAPM observatory and has been leading the EVAPM team from the very beginning. At the same time he is involved in the mathematics DEP team and in the mathematics curriculum design board in the French Ministry of Education. In addition he is a member of the TIMSS Subject Matter Advisory Commitee.***

He also belongs to the IREM National network (Instituts de Recherche sur l'Enseignement des Mathématiques) and to the CNRS (Centre National de la Resherche scientifique) research group : Didactique et acquisition des connaissances scientifiques.

Bodin@math.univ-fcomte.fr

*** Many thanks to Gillian CLOSE and Margaret BROWN (Kings' College London) for their help in reviewing the english used in this paper.*

Observatoire EVAPM

Equipe associée à l'INRP
L'Observatoire EVAPM est Organisé par l'APMEP
Avec le concours des IREM de BESANÇON et de POITIERS
et du Groupement de Recherche Didactique du CNRS

After a 40% price increase, the new price of a given object is 84 F.

How much did this object cost before the increase ?

Show your work and justify your answer

Correct answer

University entrance 1985 (humanities) : 30 %

Grade 8, 1989 : 05 %
Grade 9, 1990 : 22 %
Grade 10 entrance (general high school) : 24%

Grade 11, 1993 :
Scientific track : 80 %
Mathematics option : 93%
Economics track : 59 %
Literature track : 49 %
Technical track : 53 %

Question not attempted

Grade 11, 1993 :
Scientific track : 04 %
Economics track : 10 %
Literature track : 11 %
Technical track : 11 %

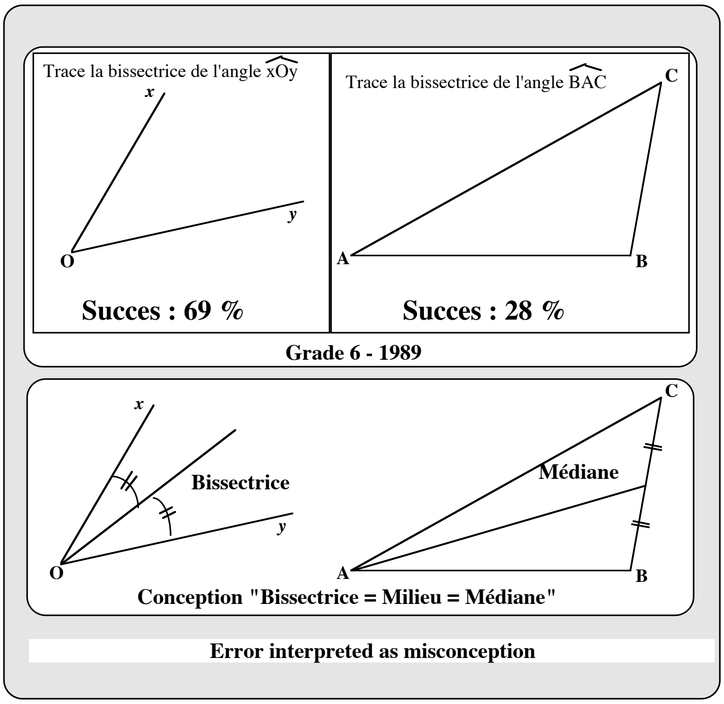
Direct use of the 1.4 multiplicative coefficient

Grade 11, 1993 :
Scientific track : 19 %
Mathematics option : 21%
Economics track : 17 %
Literature track : 21 %
Technical track : 19 %

Question Université Paris 7
EVAPM1/93 Question CC 19-21

Observatoire EVAPM

Equipe associée à l'INRP
L'Observatoire EVAPM est Organisé par l'APMEP
Avec le concours des IREM de BESANÇON et de POITIERS
et du Groupement de Recherche Didactique du CNRS



Il y a quelques années, la population de l'ITALIE était le douzième de celle de l'EUROPE.

La population de l'EUROPE était le sixième de la population MONDIALE.

Quelle fraction de la population mondiale la population de l'ITALIE représentait-t-elle ?

Explications et calculs

7

No answer : 20 %

Success rate : 44 %

La population MONDIALE était alors d'environ 4 **milliards** d'habitants.

Donne, en *millions* d'habitants, une valeur approchée du nombre d'habitants de l'ITALIE à cette époque.

Explications et calculs

8

No answer : 35 %

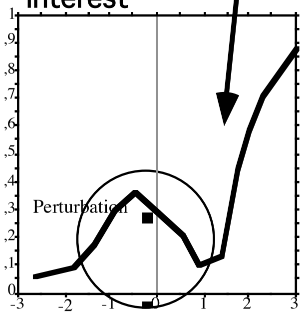
Success rate : 20 %

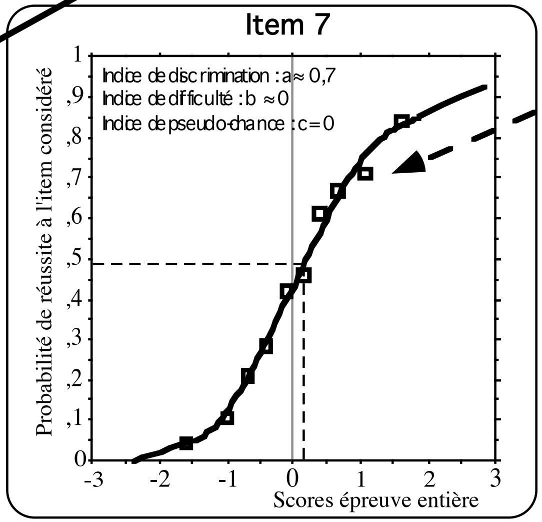
Two items pulled out of a 22 item EVAPM test WA passed in 1992 at the end of grade 9 level : items 7 and 8

The places those item take in the implicative analysis of the entire test.

Item 7 is the one of the more implicated item a well as more impliquant. Compared with the other items of the test, it also conveys the more amount of information

At the ame time, some items of the test show other type of didactical interest





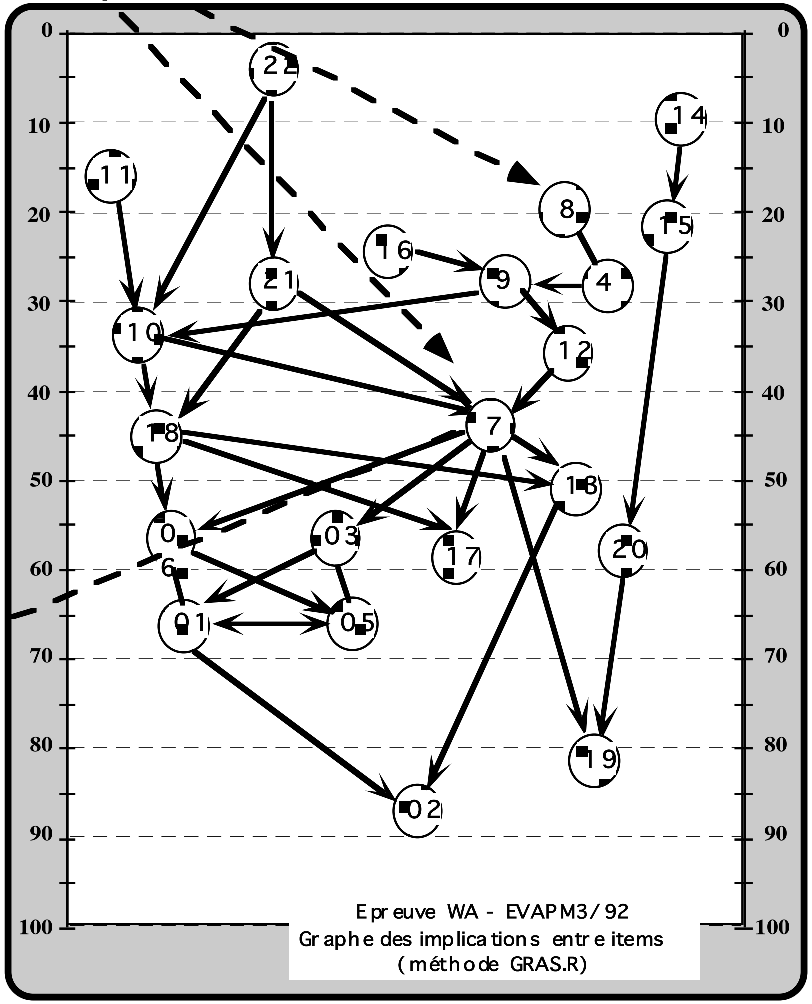
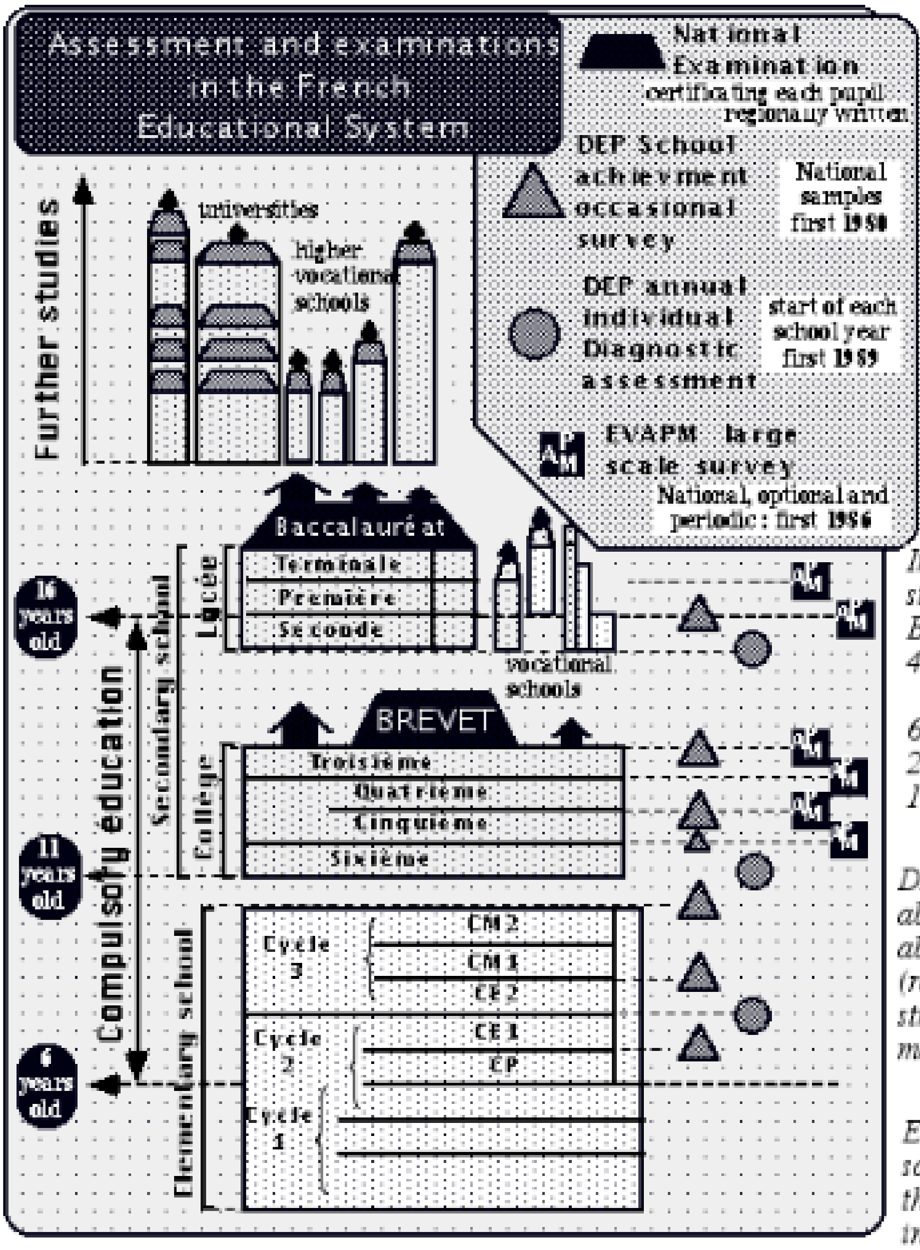


Figure 3 - AERA 96



A French age classe group is currently about 800 000 ....

In 1994, 777 000 students took the BREVET exam. 553 000 passed...

83% : General Brevet
11 % : Technological...
5 % : Professional...

In 1994, 630 000 students took the BACCALAUREAT exam. 459 000 passed...

60% : General Bacc ...
28% : Technological...
13%, : Professional...

DEP usual samples : about 5 000 students from about 50 schools (representative sample - stratified sampling method)

EVAPM studies : from some thousands to more than 100 000 students involved in every particular study

DEP : Evaluation and future planning department in the Ministry of Education

EVAPM : Mathematics teaching Monitoring system organised by the Mathematics Teachers Association (EVAPM) and sponsored by the Ministry of Education

AERA 96 NEW YORK

French traditional Mathematics exams *Some examples*

BACCALAUREAT Série D 1994 - PROBLEME (sur 12 points)

Ce problème nécessite l'usage de deux feuilles de papier millimétrique

I. Dans cette partie, on étudie principalement les variations d'une fonction f et on trace sa courbe représentative.

Soit f la fonction définie sur R, par $f(x) = 4\frac{e^x}{e^x+1}$

On désigne par $\mathcal{T}$ la courbe représentative de f dans un … orthonormal $(O; \vec{i}, \vec{j})$ (unité graphique : 2 cm)

1) Déterminer les limites de $f(x)$ quand x tend vers … tend vers $+\infty$

En déduire les droites asymptotes à $\mathcal{T}$.

2) Etudier les variations de f et dresser son table…

3) Démontrer que le point d'intersection A de … ordonnées …

4) Donn…

5) Trace… droites a…

II. Dans… courbe r…

1) Soit F… courbe r… graphique 2 cm).

Quel est le sens de variation de F ?

2) Expliciter F(x), pour tout x réel.

3) a) Déterminer les limites de F(x) quand x tend vers … tend vers $+\infty$.

b) En déduire l'existence d'une asymptote … Donner son équation.

c) Démontrer que la droite Δ d'équati… asymptote à Γ.

(On pourra remarquer que $e^x + 1 =$ …

4) Résumer les résultats précédents dans u…

5) Tracer G et ses asymptotes sur une deu… millimétrique.

III. L'objet de cette dernière partie est le c… courbe représentative T de f.

1) Soit n un entier naturel non nul.

On définit : $I_n = \int_{n}^{n+1} 4\frac{e^t}{e^t+1}\,dt$

a) A l'aide de la courbe représentative T de f, donn… graphique du nombre I_n

b) Prouver que : $I_n = 4\ln\left(\frac{n+2}{n+1}\right)$

2) On considère S…

a) A l'aide de la co… graphique du nomb…

b) En déduire une…

3) Calculer, en cm… droite d'équation y… Déterminer la limi…

NB : *Il y avait aussi un exercice sur les complexes (4 points) et un exercice sur les probabilités (4 points)*

In any case communication and justification competencies are considered important, as well as procedures used and proofs.

Traditionaly each exam is composed of one "problem" (long question) and two or more exercices (shorter questions). Questions and sub questions are linked together and the problem as well as the exercices must each have an overall meaning

BREVET des COLLEGES - LILLE 1989 – PROBLEME

On n'oubliera pas d'énoncer correctement les théorèmes utilisés. L'unité est le centimètre.

1. Construire un triangle ABC, isocèle de sommet A, de hauteur [AD], tel que BC = AD = 10. On complètera la figure à partir du segment [BC] tracé en premier. Justifier votre construction.

2. Le cercle de diamètre [BC] recoupe [BA] en F et [CA] en E et coupe [AD] en I.
Quelle est la nature des triangles BFC et BCE ?
Démontrer que les droites (BE), (CF) et (AD) sont concourantes en un point H.

3. Justifier que $AC = 5\sqrt{5}$
En exprimant de deux façons différentes l'aire du triangle ABC, démontrer que :
AD x BC = BE x AC.
En déduire la distance BE, puis calculer la distance EC.

4. Exprimer tan $\widehat{B}$ dans les triangles BCE et BDH.
Calculer alors la distance HD et démontrer que H est le milieu du segment [DI].

NB : L'épreuve comportait 5 autres exercices

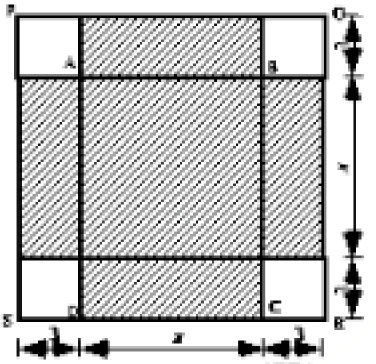
BREVET des COLLEGES - BORDEAUX 1989 – Exercice 2

2. On considère l'application f de IR dans IR définie par:

$$f(x) = (x-3)(2x+5) + x2 - 9.$$

a) Développer, réduire et ordonner $f(x)$.
b) Factoriser $f(x)$.
c) Résoudre dans IR l'équation $f(x) = -24$.

3. Voir schéma ci-contre.
a) Exprimer, en fonction de x, l'aire de la partie hachurée.
b) Si l'aire de la partie hachurée est égale à 133 :
— quelle est l'aire du carré PQRS ?
— quelle est la longueur du côté du carré PQRS ?
— quelle est la longueur x du côté du carré ABCD ?
c) On vérifiera que le nombre x ainsi trouvé est tel que : x2+ 12x = 133.



French teachers as well as public are very attached to this kind of exam. Until now attempts to make them evolve towards more separate questions or an MCQ format have failed. Also moves towards counting assessment of schoolwork have failed.

AERA 96 - NEW YORK