

Zeitschrift: Études pédagogiques : annuaire de l'instruction publique en Suisse
Band: 61/1970 (1970)

Artikel: Docimologie et orientation : l'insuffisance de nos moyens d'évaluation
Autor: Bain, D.
DOI: <https://doi.org/10.5169/seals-115816>

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. [Mehr erfahren](#)

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. [En savoir plus](#)

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. [Find out more](#)

Download PDF: 09.02.2026

ETH-Bibliothek Zürich, E-Periodica, <https://www.e-periodica.ch>

Docimologie et orientation

L'insuffisance de nos moyens d'évaluation

Importance et actualité du problème

« La conférence des maîtres déconseille l'inscription de cette élève au 10^e degré en section latine. »

Lors d'une récente enquête sur l'évolution au dixième degré¹ des élèves sortant du Cycle d'orientation genevois, nous avons trouvé la remarque ci-dessus dans onze dossiers. Ces jeunes filles, malgré ce conseil, s'étaient inscrites à l'école secondaire supérieure en section latine. En relevant leurs résultats à la fin du dixième degré, nous avons eu la surprise de constater que seules deux d'entre elles confirmaient le pronostic pessimiste de leurs maîtres en échouant; les neuf autres réussissaient, la majorité (six) obtenant même le certificat.

Il serait abusif et malhonnête de vouloir caractériser l'orientation dans une école à partir d'un exemple particulier. Toutefois, si nous nous sommes permis de le citer, c'est qu'il ne représente pas un accident isolé et qu'il illustre, de façon un peu caricaturale, les difficultés d'appréciation que les maîtres rencontrent tous les jours dans leur activité d'orientateurs.

Les professeurs qui ont examiné le cas de ces onze élèves en conseil d'école disposaient pourtant non seulement des notes annuelles dans toutes les disciplines, mais encore des résultats des deux années précédentes, sans parler des informations fournies par les instituteurs primaires.

D'autre part, ces enseignants n'ont pas donné un pareil conseil à la légère. Nous verrons que ces erreurs de pronostic peuvent difficilement être mises sur le compte de l'instabilité qui caractériserait le comportement de l'adolescent. Il nous faut donc mettre en cause nos moyens d'évaluation dont l'élaboration procède actuellement d'un empirisme dangereux. En employant cet adjectif, nous songeons aux élèves qui se sont conformées à l'avis de leurs maîtres et ont renoncé à s'inscrire dans une section latine, ignorant la fragilité du conseil qui leur était donné².

¹ Non publiée.

² Notre étude qui portait sur le 10^e degré ne nous a pas permis de les recenser.

Définitions

Le terme de *docimologie* a été proposé par Piéron pour désigner l'étude systématique des *examens* (10)¹. Pour notre part, suivant une tendance actuelle, nous étendrons le domaine de la docimologie en appliquant ce terme à l'étude des *moyens de contrôles pédagogiques* en général.

Nous donnons par ailleurs à *orientation scolaire* un sens très large, incluant la sélection scolaire que la pratique actuelle a beaucoup de peine à dépasser (6). Il s'agit de la tâche pédagogique (dévolue actuellement surtout aux écoles moyennes) consistant à guider les écoliers dans le choix de leurs études en fonction de leurs capacités, de leurs aptitudes et de leurs goûts. Nous nous contenterons de cette définition sommaire pour ne pas entrer dans des développements qui nous entraîneraient trop loin. Concrètement, nous pensons surtout à l'activité du maître d'un cycle d'orientation ou d'un établissement similaire, qui est non seulement chargé d'instruire ses élèves mais aussi de les aiguiller dans les différentes sections que comportent nos systèmes scolaires dès le niveau secondaire. Dans les considérations qui suivent, nous nous limiterons aux aspects *pédagogiques* de l'orientation, ce qui ne signifie naturellement pas que nous sous-estimions le rôle, en orientation scolaire, de la psychologie et des psychologues.

La note, principal moyen d'appréciation et d'orientation ; ses fonctions

NOMBREUSES sont les études docimologiques qui ont paru cette dernière décennie. Devant cette abondante bibliographie, nous sommes frappés par leurs médiocres répercussions sur la pratique journalière de l'enseignement. Les maîtres se sentent souvent peu concernés par des travaux qui portent principalement sur des examens, moyens de contrôle qui ont généralement été élaborés par d'autres qu'eux: inspecteurs ou spécialistes de la pédagogie expérimentale. D'ailleurs la tendance actuelle est de supprimer ces examens (pour des raisons souvent plus affectives que scientifiquement fondées), d'où l'importance des *notes* attribuées par les enseignants tout au long de l'année. A Genève par exemple, ce sont elles qui déterminent pratiquement la promotion et l'orientation de l'école enfantine à la maturité. Peut-on affirmer qu'elles échappent aux défauts des examens?

Avant de répondre à cette question, nous devons définir les différentes *fonctions* que remplissent les appréciations du maître en général et les notes en particulier. Nous en distinguerons trois: fonctions diagnostique, de motivation et pronostique. Nous traiterons rapide-

¹ Les chiffres entre parenthèses renvoient à la bibliographie placée à la fin de l'article.

ment et seulement partiellement les deux premières qui concernent tous les types et tous les ordres d'enseignement, pour développer surtout la troisième, caractéristique de l'orientation.

La fonction diagnostique et les problèmes du contrôle des acquisitions

L'enseignant a pour mission de faire acquérir à ses élèves un certain nombre de connaissances, de techniques, de savoir-faire, d'habitudes et d'aptitudes, c'est-à-dire de favoriser chez eux certains apprentissages, au sens psychologique du terme. Quelles que soient les méthodes didactiques qu'il emploie ou la pédagogie à laquelle il se réfère, il doit disposer de moyens pour mettre en évidence l'évolution des conduites que suppose tout apprentissage.

Au premier abord le problème paraît simple. Le plan d'études définit les notions à enseigner et le manuel les précise et les illustre. Mais le professeur débutant se rend vite compte de l'insuffisance de ces précisions; elles ne lui fournissent pas d'indications sur ce que nous appellerons les *normes et niveaux d'apprentissage*. Quelles sont les notions importantes? Quel niveau d'acquisition (5) faut-il exiger: l'élève doit-il par exemple savoir simplement énoncer une loi physique ou la comprendre, l'appliquer, la situer parmi d'autres lois, voire critiquer sa formulation et la remettre en cause? Ce niveau supposé défini, quand peut-on estimer qu'une notion est acquise, quelles fautes doivent être considérées comme rédhibitoires, quelle proportion d'erreurs est tolérable à un moment déterminé, étant donné les révisions que comportent implicitement ou explicitement tous les programmes? Faut-il contrôler strictement ce que les élèves ont appris, sous la forme où ils l'ont appris, ou est-il licite de leur poser des problèmes analogues sous d'autres formes pour voir s'ils sont capables de «transférer» leurs connaissances et les techniques acquises à d'autres matières ou situations? Comment contrôler et apprécier des acquisitions sans avoir reçu préalablement une réponse à ces questions? Nous pensons donc qu'avant d'aborder les aspects proprement docimologiques de l'enseignement, c'est-à-dire ceux relatifs aux modalités des contrôles, il s'agirait d'en définir les objectifs avec précision, en donnant des exemples des comportements visés, ce qui revient à fournir aux maîtres une série d'épreuves traduisant de façon adéquate ces objectifs. Ceux-ci doivent naturellement être fixés expérimentalement en tenant compte notamment des données de la psychologie du développement.

Nécessité et utilité d'épreuves diagnostiques

La plupart des tests de connaissance ou d'acquisition (achievement tests) édités sont en fait des épreuves de *niveau*. Leurs barèmes

situent la performance d'un élève par rapport à celles de camarades du même degré ou/et du même âge, de la même section, etc. Ils permettent tout au plus d'évaluer les chances de réussites ou d'échecs ultérieurs, ce qui nous renvoie à une autre fonction des appréciations et de la note. Nous manquons à l'heure actuelle d'épreuves qui situent la production d'un élève par rapport aux *notions à acquérir et au processus d'apprentissage*. Pour en citer cependant un exemple, le professeur de français qui veut revoir avec sa classe l'accord du participe passé en recourant à un cours programmé comme celui de G. Métraux (7), commencera par appliquer un pré-test de type diagnostique comportant un échantillon des principaux cas d'accord et tenant compte des erreurs les plus fréquentes (confusions participe passé — infinitif ou participe passé — passé simple, par exemple). Le total de points obtenus importera moins que le type de fautes commises. Grâce à des formules de dépouillement *ad hoc*, chaque élève peut composer lui-même son programme de révision en fonction des lacunes mises en évidence. On constate que ce genre de tests n'a sa raison d'être que s'il est complété par un système de cours d'appui ou de rattrapage; il suppose en outre un enseignement individualisé.

Les problèmes et défauts de nos épreuves d'acquisition ; fidélité des notes

Faute d'une analyse claire des objectifs et des points importants du programme, le maître qui compose une épreuve est souvent embarrassé lorsqu'il doit choisir les points à contrôler et la forme de ce contrôle. Pour des raisons pratiques, de temps disponible notamment, il procède par sondage. Le débutant n'est jamais sûr que son échantillonnage est représentatif de l'ensemble des notions enseignées; il ignore d'autre part le degré de difficulté de ces questions, la façon dont elles discriminent les élèves forts et faibles. Dans l'impossibilité d'essayer ses épreuves à blanc, il ne contrôle pas toujours la clarté de la formulation des consignes. Dans ces conditions, l'épreuve ressemble un peu à une loterie, au moins pour les élèves moyens et faibles. La réussite à un « item » ne signifie pas nécessairement l'acquisition de la notion qu'il est censé représenter et surtout l'échec n'est pas toujours significatif: formulé un peu autrement et portant sur d'autres nombres tel problème de mathématiques aurait peut-être été résolu. Etant donné le caractère aléatoire de certaines questions, le professeur qui appliquerait une nouvelle fois la même épreuve ou une autre épreuve sur le même sujet n'est pas certain de trouver les mêmes résultats. Le manque de *fidélité* ou de *constance* des contrôles scolaires et des notes qui leur sont attribuées a été signalé par maints travaux. Ce défaut se révèle surtout gênant lorsqu'on veut apprécier les progrès d'un élève: il est téméraire de prétendre qu'une moyenne mensuelle qui a passé de 3 à 4 atteste une nette amélioration; sans

s'en rendre compte, le maître a pu modifier ses exigences ou son barème.

Appréciations et notes comme moyens de motivation

Les études en psychologie de l'apprentissage ont mis en évidence la nécessité pour tout sujet qui apprend d'être informé régulièrement sur ses résultats.

Par nos appréciations, quelle qu'en soit la forme: remarque, note, attribution d'un rang, nous agissons donc sur la motivation de nos élèves.

La note a toutefois un statut particulier parmi les moyens d'appréciations. Elle comporte d'abord un jugement moral: bon, médiocre, nul, qui primitivement concerne la *production* d'un élève, mais qui tend finalement à qualifier sa *personne*, dans l'esprit de l'enfant et parfois dans celui du maître ou des parents. D'autre part, bénéficiant de tout un conditionnement précoce, la note manifeste, à l'emploi, une efficacité certaine sur la plupart des élèves. A tel point que les maîtres dont les disciplines ne bénéficient pas d'une sanction par la note (musique, par exemple) se plaignent parfois de ne pouvoir obtenir un rendement suffisant de leurs élèves et souhaiteraient que l'on introduisît une appréciation chiffrée également dans leur branche. Il est évident que cette efficacité joue à court terme. Illustrons les limites et les dangers d'un « chantage à la note » par cette anecdote que nous rapportait un collègue. Victor avait de très mauvaises notes en physique; son père avait décidé de l'obliger à faire des progrès en l'aidant personnellement, ce qui semble judicieux, et en lui promettant un véloMOTEUR s'il obtenait un 6 à la prochaine épreuve générale de physique. Stimulé par cette promesse, Victor obtint son 6 et son véloMOTEUR. Mais conjointement l'allemand descendait en flèche et tout était à recommencer. Ne pouvant promettre une voiture à son fils, le père dut se rabattre sur un autre type de motivation: menacer de mettre le véloMOTEUR sous clef.

Jouant sur cet aspect motivationnel, certains professeurs infligent de mauvaises notes à leurs élèves, pour les obliger à travailler, quitte à organiser des épreuves de « repêchage »; d'autres accordent des notes dites d'encouragement. Dans ces deux cas, on peut se demander ce que signifie une moyenne mensuelle ou annuelle dans laquelle on a mélangé les notes-constats et les notes-motivation.

Fonction pronostique et normes de promotion

Interrogés sur la signification et la fonction de leurs notes, les enseignants répondent qu'elles représentent un constat des acquisitions ou lacunes de leurs élèves et admettent parfois qu'elles consti-

tuent un moyen de pression pour faire travailler certains d'entre eux. En revanche, ils se défendent souvent de leur donner une valeur pronostique, ne voyant pas la nécessité d'une pareille prévision, l'essentiel étant de contrôler l'évolution actuelle de l'apprentissage scolaire. Nous sommes alors frappés par le fait que, dans la plupart des écoles, la moyenne annuelle qui résume une série de constats, prend tout à coup une signification pronostique puisqu'elle décide de la promotion ou de l'échec. Tout se passe comme si, en dessous de telle norme fixée arbitrairement plutôt qu'expérimentalement par le règlement, on estime que l'élève n'a aucune chance de réussir la suite de ses études dans la branche considérée. Pour justifier cette pratique, on peut alléguer l'opinion de psychologues comme Tilton (15) qui sont d'avis que le meilleur pronostic des acquisitions futures est fait à partir du niveau des acquisitions antérieures, donc du constat actuel. Si ce point de vue est déjà discutable en ce qui concerne les constats obtenus grâce à des tests bien étudiés, il devient contestable quand on s'appuie sur les notes; nous en donnerons une preuve ci-dessous.

Fonction pronostique et orientation

Une école unique et indifférenciée comme l'école primaire a recours aux fonctions diagnostique et motivationnelle des appréciations, mais a rarement l'occasion ou l'obligation de poser un pronostic, exception faite du cas signalé au paragraphe précédent. En revanche, on attend en plus d'un cycle d'orientation qu'il aide ses élèves à découvrir leurs aptitudes. Nous n'épiloguerons pas sur le concept d'aptitude qui a été défini ou critiqué par des auteurs plus autorisés que nous (9). Quelle que soit la définition donnée, cette notion est généralement liée étroitement avec celle de possibilités d'apprentissage. L'orienteur ne peut donc pas se contenter d'un *constat* des acquisitions à un moment donné; il doit étudier ce que André Rey appelait le *processus actuel d'acquisition* (13) pour en tirer des conclusions sur les potentialités de l'individu. Concrètement, en orientation scolaire, les maîtres ont souvent pour mission de déterminer avec les intéressés s'il est utile ou indiqué pour un élève de commencer ou continuer à suivre des cours dans telle branche à tel niveau de difficulté; dans quelle discipline est-il plus « rentable » pour lui d'investir son énergie et son temps, compte tenu naturellement de ses goûts et intérêts? Le choix du latin ou de la physique, du dessin ou de l'allemand, d'une section générale plutôt que scientifique engage l'avenir scolaire et professionnel. Le conseiller, psychologue ou enseignant, doit donc disposer de moyens d'appréciations qui aient une valeur pronostique suffisante, une bonne *validité* comme disent les psychotechniciens. Est-ce le cas de nos résultats et notes scolaires? Nous donnerons ci-dessous des exemples qui justifient le titre du

présent article en montrant les faiblesses et insuffisances de nos moyens d'appréciation. Nous en chercherons ensuite les causes pour finalement esquisser quelques solutions.

*Validité des normes de promotion et d'orientation:
l'admission des élèves de 7^e générale en 8^e moderne.*

Nous résumerons ici une étude dont nous avions été chargés en 1969 par la direction du Cycle d'orientation genevois (1).

En septembre 1968, notre école créait en 8^e une nouvelle section: la moderne. De type prégymnasial, elle mettait l'accent sur les langues modernes (français, allemand et anglais) et visait une culture générale centrée sur le monde d'aujourd'hui. Si l'on avait voulu garantir dès le début un niveau prégymnasial à cette section, il aurait fallu n'admettre que des élèves promus de 7^e latine ou scientifique ou susceptibles d'y être transférés en fin d'année. Cette mesure a paru peu compatible avec la mission d'orientation de notre école; de ce point de vue, elle aurait représenté une régression par rapport à la structure antérieure. Après consultation des organes intéressés, la direction générale décida d'ouvrir largement cette section aux élèves de 7^e générale, en fixant des normes provisoires relativement basses (4,0) ne différant des normes de promotion que de 0,5 et ce dans les deux branches qui paraissaient primordiales: le français et l'allemand; des dérogations à ces normes pouvaient être accordées par le conseil d'école¹.

Mais en juin 1969, devant la forte proportion d'élèves de G en échec en 8^e moderne, les maîtres et responsables du CO se demandèrent s'il ne fallait pas modifier les conditions de passage, en exigeant un 4,5 (à la place d'un 4,0) de moyenne annuelle en allemand et en français. Avant de prendre une pareille mesure, il s'agissait de supputer ses conséquences et son efficacité. Nous pouvions les évaluer en appliquant les nouvelles normes à la première volée d'élèves de 7^e G entrée en 8^e moderne en septembre 1968. Certes, nous n'avions pas le droit de comparer sans autre les résultats d'une modification simulée à ceux d'une application réelle de nouveaux critères; dans le second cas, la population est sélectionnée et les conditions d'échec et de réussite en sont modifiées. Toutefois, nos conclusions ne devaient guère en être affectées, si les résultats enregistrés étaient assez nets. On en jugera en consultant le tableau ci-dessous. Il devait nous permettre de répondre aux questions suivantes: les élèves qui, en septembre 1968, satisfaisaient aux nouvelles normes ont-ils tous réussi, les autres ont-ils échoué pour la plupart?

¹ Conférence réunissant tous les maîtres enseignant dans un même degré et un même collège.

TABLEAU I. Réussites ou échecs en fonction des nouvelles normes:
4,5 de moyenne annuelle en français et en allemand
(nombre d'élèves et pourcentage)

7 ^e My annuelle	Réussites	8 ^e moderne		Totaux	
		Echecs			
F. et A. $\geq 4,5$	a) 47	23,0%	b) 20	9,8%	67 32,8%
F. ou* A. $< 4,5$	c) 65	31,9%	d) 72	35,3%	137 67,2%
Totaux		112 54,9%	92 45,1%	204	100,0%

$q = 0,20$
 $X^2 = 8,5$ TS
 * ou disjonctif.

Les lettres *a*) à *d*) désignent les 4 cas possibles:

- a*) les élèves qui satisfaisaient aux nouvelles normes (4,5 au moins en allemand *et* en français) et ont réussi la 8^e M;
- b*) les élèves qui satisfaisaient aux nouvelles normes et ont échoué au cours ou à la fin de la 8^e M;
- c*) les élèves qui ne satisfaisaient pas aux nouvelles normes (français ou/et allemand $< 4,5$) et ont réussi la 8^e M;
- d*) les élèves qui ne satisfaisaient pas aux nouvelles normes et ont échoué au cours ou à la fin de la 8^e M.

Les cas *a*) et *d*) étaient favorables à une augmentation des normes de passages: ces élèves avaient réussi la 8^e M *et* satisfaisaient aux nouveaux critères ou échouaient *et* n'avaient pas obtenu en 7^e G le 4,5 de français et d'allemand; ils ne représentaient cependant que le 60 % de l'ensemble de la population considérée. Les cas *b*) et *c*) et le coefficient d'association ($q = 0,20$) démentaient l'existence d'une liaison causale systématique entre la réussite en 8^e M et l'obtention d'un 4,5 au moins en français et en allemand. Un élève sur dix échouait alors qu'il aurait été admis sous le nouveau régime (*b*). Ce cas était moins gênant que le suivant (*c*): 65 élèves, soit un sur trois, auraient été éliminés, alors qu'en réalité ils ont été promus en 9^e moderne; on les aurait donc empêchés de réussir en leur appliquant la norme de 4,5.

Nous craignons que ces constatations faites à partir d'un cas particulier ne s'appliquent aux critères de promotion en général.

En fixant des normes de passage, le législateur scolaire estime que le succès scolaire futur dépend presque exclusivement du succès passé. Cette affirmation est sujette à caution parce qu'elle suppose:

a) connus et pris en considération tous les facteurs influençant la réussite scolaire;
 b) valable la façon d'apprécier cette réussite (= les notes)
 Or, ces conditions sont rarement satisfaites.

La validité des notes et des épreuves générales

Le tableau I nous révélait déjà la médiocre corrélation existant entre les résultats de 7^e G et 8^e M. Mais quel est le rôle que jouent les différentes branches (principales) dans le pronostic, quelle confiance peut-on accorder aux moyennes de fin d'année et aux épreuves communes, standardisées et théoriquement plus objectives?

TABLEAU II. *Coefficient de corrélation bisérial entre les notes de 7^e G et la réussite ou l'échec en 8^e M*

Branches	r bis	r ²	N
Français	0,31	TS	0,09
Allemand	0,16	TS	0,02
Mathématiques	0,19	TS	0,04
Sciences	0,16	S	0,02
Somme des 4 branches	0,30	TS	0,09

TABLEAU III. *Coefficient de corrélation (r bis) entre les scores (en points) des épreuves communes de 7^e G / II^e ou III^e trimestre et la réussite ou l'échec en 8^e M*

		r	r ²	N
Français	7 ^e EC II / 8 ^e échec ou réussite	0,18	0,03	166
Allemand	7 ^e EC III / 8 ^e échec ou réussite	0,36	0,12	194
Mathématiques	7 ^e EC III / 8 ^e échec ou réussite	0,24	0,05	186

Les chiffres de ces derniers tableaux sont décevants. On peut prendre comme indice de précision du pronostic le r^2 , qui varie de 1 (pronostic sûr) à 0 (aucun pronostic valable possible); on voit qu'il ne dépasse pas 0,12. Il y a certes une relation entre la réussite en 7^e G et le succès en 8^e M, (les corrélations sont significatives), mais cette relation n'est pas très étroite. Une prévision, et par conséquent une décision, qui se fonderaient uniquement sur les notes ou les épreuves communes (sous forme d'examen, par exemple) apparaissent comme très aléatoires.

Il en est de même si l'on cherche à pronostiquer les notes de 8^e M à partir de celles de 7^e dans *la même branche*.

TABLEAU IV. *Coefficients de corrélation entre les notes de 7^e G et 8^e M : r de Bravais-Pearson*

Branches	7 ^e III ^e trimestre					7 ^e moyenne annuelle			
	8 ^e I	0,42	TS	0,17	207	0,45	TS	0,20	205
Français	8 ^e II	0,33	TS	0,10	153	0,41	TS	0,16	152
	8 ^e My	0,38	TS	0,14	133	0,46	TS	0,21	132
Allemand	8 ^e I	0,56	TS	0,31	196	0,36	TS	0,12	201
	8 ^e II	0,41	TS	0,17	150	0,33	TS	0,10	153
	8 ^e My	0,46	TS	0,21	128	0,36	TS	0,12	131
Mathématiques	8 ^e I	0,30	TS	0,09	175	0,37	TS	0,13	190
	8 ^e II	0,35	TS	0,12	135	0,24	TS	0,05	147
	8 ^e My	0,40	TS	0,16	118	0,30	TS	0,09	130

Le maître le plus confiant en la sûreté de son jugement sur les capacités de ses élèves admettra sans doute volontiers qu'il est difficile de faire des prévisions exactes à longue échéance; trop de facteurs accidentels peuvent influer sur la réussite: circonstances familiales, maladie, accident, etc. En revanche, à court terme, dans les 5 mois qui séparent la moyenne de 7^e III^e trimestre et celle de 8^e I^{er} trimestre, ces facteurs jouant un rôle moins important, le pronostic devrait être relativement sûr. Cette hypothèse est démentie par les chiffres du tableau IV. Le coefficient le plus élevé atteint 0,56 seulement, ce qui correspond à une médiocre précision ($r^2 = 0,31$). Cela signifie pratiquement que des élèves qui ont en 7^e G une moyenne de 4,5 en allemand au troisième trimestre obtiennent à la fin du trimestre suivant (8^e I) des notes qui varient entre 2,5 et 4,5, soit entre une note éliminatoire et celle qui correspond au certificat.

Conséquences pour la sélection et l'orientation

Vu la faiblesse des corrélations, la détermination d'un seuil de sélection est délicate. Dans l'exemple ci-dessus, les notes 4 et 4,5 se situent à la moyenne des notes attribuées par les maîtres aux élèves de 7^e G, ce qui est le cas le plus défavorable. Piéron (10) a « dénoncé maintes fois cette absurdité statistique. Dans une distribution normale, il est impossible de faire une coupure légitime dans

le groupe modal qui a l'effectif maximum; il faut un tirage au sort pour couper ce groupe en deux ». En d'autres termes, la zone du 4 est celle où il est le plus difficile de distinguer des différences de capacités scolaires. C'est dans ce groupe modal que les erreurs de mesure signalées dans les paragraphes précédents peuvent se traduire par des écarts ou différences appréciables quoique non significatifs. Il faudrait donc fixer le point de coupure nettement au-dessous ou au-dessus de la moyenne. Choisissant cette dernière solution, nous avions cherché, dans l'étude décrite ci-dessus, les seuils de sélection qui optimiseraient la proportion de réussite. En exigeant un total de 10 pour les moyennes annuelles d'allemand et de français, nous abaissons le taux d'échec à 19 %, mais il ne nous serait resté que tout juste l'effectif d'une petite classe, soit 16 élèves au lieu de 204.

Pour l'orientation, les conséquences sont tout aussi gênantes. Un élève bon ou faible en 7^e dans une branche ne le sera pas nécessairement également en 8^e; le choix d'une filière scolaire, dans ces conditions, prend un caractère relativement aléatoire.

Les sources d'erreurs

Nous voudrions tout d'abord signaler l'impasse dans laquelle s'engagent toutes nos études de validité. Elles prennent comme critères des données (échec ou réussite, notes) qui sont aussi sujettes à caution que les éléments sur lesquels se fondent les pronostics. Les résultats de 8^e n'ont pas nécessairement plus d'objectivité que celles de 7^e. En dénonçant la médiocre validité des notes primaires, les écoles secondaires postulent que leurs appréciations ont plus de valeur; elles n'ont en fait pour seul mérite que de venir après et de déterminer effectivement l'échec ou la réussite qu'il faut prédire à un certain moment.

Nous avons déjà analysé le *manque d'objectivité et de fidélité* des instruments d'appréciations utilisés par les maîtres. Piéron (10), entre autres, a montré que si l'on demande à un professeur de corriger deux fois, à un certain intervalle, les mêmes travaux, celui-ci pourra leur attribuer des notes très différentes, même lorsqu'il s'agit de branches où l'objectivité de la correction semble acquise (mathématiques ou sciences). Cette variabilité de l'appréciation est d'autant plus grave qu'elle touche le plus souvent des élèves situés de part et d'autre de la limite du suffisant.

L'expérience professionnelle des maîtres enseignant aux différents degrés peut varier. Elle retentit naturellement sur la qualité de l'enseignement reçu par les élèves. Le professeur expérimenté saura mieux distinguer dans le programme l'essentiel de l'accessoire; il en tiendra compte dans ses travaux de contrôle. Ses élèves seront à la fois mieux préparés et mieux jugés.

Les professeurs d'une même discipline n'ont pas tous la même *conception de la branche* qu'ils enseignent. En français par exemple, certains attribueront beaucoup d'importance à la rédaction et à la lecture expliquée, d'autres drilleront l'orthographe et la grammaire. En allemand et en anglais, chacun mettra l'accent de façon différente sur la langue écrite ou parlée, sur l'analyse grammaticale ou l'acquisition intuitive des structures, sur la traduction ou la compréhension globale. D'une année à l'autre, l'élève peut se trouver ainsi devant des exigences bien différentes, que traduiront notamment les épreuves. Les divergences peuvent porter aussi sur l'importance de certaines notions. Nous rappellerons à ce propos l'enquête de M. Reuchlin et F. Bacher dans les classes de 3^e françaises (11). Partant d'épreuves objectives de connaissances relatives au programme de français et de mathématiques de troisième, on a demandé aux professeurs de ces deux disciplines d'indiquer pour chacune des questions posées quel devrait être, d'après eux, le pourcentage d'élèves capables, au moment d'entrer en 2^e, de donner la réponse correcte. Pour un problème comme celui-ci: Math. 15. Un cercle a 7 cm de rayon. Par un point A, situé à 9 cm. du centre, on mène une sécante dont la partie extérieure AB est égale à la partie intérieure BC. Quelle est la longueur AB? (La figure était donnée).

Les proportions variaient, selon les professeurs (qui enseignaient tous en 3^e et 2^e), de 100 % à 0 %. Donc, pour les uns ce problème correspondait à des notions essentielles, nécessaires à tous les élèves, alors que pour d'autres enseignants, ce chapitre de géométrie présentait un intérêt très accessoire pour la suite des études. Pareille enquête menée dans nos écoles ne donnerait probablement pas des résultats fondamentalement différents.

Une même branche ou matière enseignée à des niveaux différents fait appel parfois à des *aptitudes différentes*. C'est probablement le cas pour l'enseignement d'une langue qui passe d'une phase orale à un stade écrit et grammatical. En outre, l'auteur de pittoresques narrations ou descriptions (7^e et 8^e) n'a pas nécessairement les qualités de sophiste indispensables pour réussir une bonne dissertation.

Les études docimologiques montrent que chaque maître a tendance à apprécier les travaux de ses élèves en fonction de son « *équation personnelle* ». On le qualifiera généralement de plus ou moins sévère ou exigeant par rapport à ses collègues. Mais à cela s'ajoute aussi une propension à utiliser une proportion plus ou moins grande de l'échelle des notes. A ce point de vue, on pourrait établir une véritable typologie des enseignants, qui varierait d'ailleurs selon la branche enseignée.

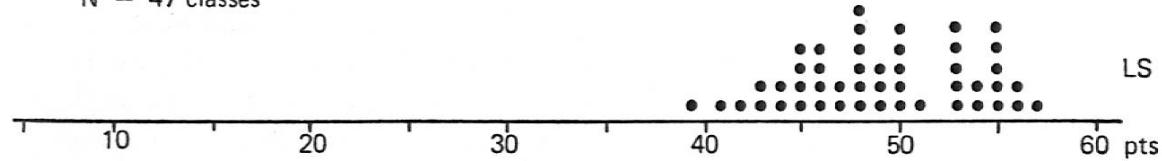
Il est bien connu que l'appréciation du maître est très influencée par *le niveau de sa classe*. Les notes provenant de classes différentes sont-elles donc comparables? A cette question, Karlheinz Ingenkamp

(7), sur la base de ses recherches, répond: « Pour attribuer leurs notes, les maîtres se réfèrent au niveau de performances de leur classe actuelle. La définition des notes (6 = très bien, 5 = bien, etc.) et le souvenir de classes antérieures ne suffisent pas pour assurer une mesure indépendante du niveau de la classe. A l'intérieur de chaque classe, les élèves sont ordonnés de façon relativement concordante par les maîtres et par les tests. Mais dans un ensemble de classes du même degré des notes très différentes correspondent à des performances objectivement égales. Qu'un élève d'un niveau de performance déterminé obtienne un 2 ou un 4 dépend du hasard de son appartenance à telle classe. Ces résultats signifient ni plus ni moins que tout notre système de normes ne repose sur aucun critère objectif. »

Le type d'erreur signalé par Ingenkamp sera d'autant plus important que les niveaux des classes différeront. Une étude récente non publiée, dont nous extrayons le tableau suivant, met en évidence de grandes différences entre les classes, au moins au niveau de la 7^e du CO.

TABLEAU V. *Niveaux moyens des classes de 7^e aux épreuves communes d'allemand du premier semestre par section (1968-1969)*
Histogramme des moyennes de classes (en points)

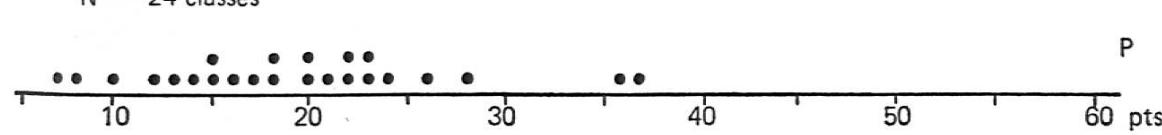
N = 47 classes



N = 43 classes



N = 24 classes



nombre de points maximum : 66 pts

Si nous considérons la section G, nous constatons que la classe la plus faible obtient à cette épreuve commune une moyenne 2,5 fois inférieure à celle de la classe la plus forte (resp. 20 et 49 points). Avant cette épreuve, les deux professeurs d'allemand (enseignant dans deux collèges différents) n'ont probablement pas pu estimer avec précision le niveau de leurs élèves par rapport à l'ensemble de la

section ; ils se seront certainement rendu compte qu'ils avaient beaucoup de résultats respectivement médiocres et excellents, mais nous doutons que leurs barèmes, dans d'autres travaux de contrôle, n'aient pas été influencés par le niveau général de leur classe.

Le tableau montre en outre de très larges recouvrements entre sections¹. Comme M. Reuchlin et F. Bacher (11 p. 36) nous soulignons « les conséquences très importantes de ces constatations sur la pratique de l'orientation au cours et à la fin du cycle : un changement de type de section peut fort bien ne pas présenter pour un élève de difficultés d'adaptation au niveau général de la classe qui soient supérieures à celles qu'il rencontrera si, sans changer de type de section (...), il devait seulement passer d'un établissement à un autre ». C'est ce qui se produirait si, par suite d'un déménagement, un enfant de la première classe citée dans notre exemple était inscrit dans la seconde ; en allemand il passerait d'un niveau moyen P à un niveau moyen LS. Un phénomène analogue peut jouer un rôle important lors du passage d'un degré à l'autre ou surtout d'un type d'enseignement à l'autre. Il explique peut-être le cas des élèves cité en guise d'introduction. Ces onze jeunes filles issues de deux collèges exclusivement féminins à l'époque, se sont retrouvées sur les mêmes bancs que des camarades sortant d'établissements mixtes et qui avaient été moins sélectionnées, au moins sur le plan des langues. Nos épreuves communes révèlent en effet que fréquemment les filles l'emportent sur les garçons en français, allemand et latin ; dans les langues, le niveau général d'une classe mixte est souvent inférieur à celui d'une classe ne réunissant que des filles. En outre, dans l'un des deux collèges féminins, les élèves provenaient presque exclusivement d'un quartier résidentiel. De par leur implantation, certaines écoles recrutent la majorité de leurs élèves dans des couches de la population relativement homogènes sur le plan socio-professionnel (2) ; dès les premiers contrôles on constate des différences assez nettes entre les collèges (3) ; pour l'épreuve d'allemand mentionnée ci-dessus, 10 points séparent la moyenne de collège la plus élevée de celle obtenue par l'école la moins bien lotie parce que recevant des élèves de zones populaires et rurales (resp. 52 pts et 42 pts pour l'ensemble des sections).

Dans un cycle d'orientation comportant plusieurs sections, le facteur niveau de classe étudié par Ingenkamp est modifié ou modulé par d'autres variables. Suivant leur poste (répartition des heures d'enseignement dans différentes classes et sections), les maîtres ont ou n'ont pas la possibilité de comparer les performances d'élèves appartenant à des sections différentes, les conséquences de ces comparaisons pouvant d'ailleurs varier d'un maître à l'autre. Le

¹ Il faut cependant tenir compte du fait que l'horaire hebdomadaire de G et P compte 5 heures d'allemand contre 4 pour la LS.

règlement du CO prévoit que chaque section utilise sa propre échelle de notes allant de 0 à 6. Théoriquement donc les notes devraient se distribuer de la même façon en LS, G et P. Un examen des palmarès de fin d'année nous a prouvé que ce n'était pas le cas: en 7^e latino-scientifique (1969-1970), la moitié des élèves (49,7 %) obtenaient un certificat¹, un quart (24,8) seulement des générales recevant la même distinction. A « l'équation personnelle » s'ajoutera donc une « *équation de section* ».

Lors de notre enquête sur les normes de passage de 7^e G en 8^e M, nous avons calculé la proportion d'échecs des élèves provenant de générale à la fin de la 8^e moderne; en moyenne de 45 %, le pourcentage variait de 8 % à 88 % selon les collèges. On peut difficilement attribuer de telles variations seulement à une différence de capacités scolaires liée au recrutement socio-professionnel. Des écoles situées dans des quartiers de même stratification sociale enregistraient des taux d'échecs très différents. Cette constatation et notre expérience nous incitent à émettre l'hypothèse qu'en plus du facteur recrutement, nous avons aussi affaire à *des politiques d'orientation et de sélection* plus ou moins divergentes selon les établissements. Sur le plan très général de l'appréciation des résultats, nous aurions ainsi une « *équation de collège* ».

Les facteurs relevés ci-dessus essaient d'expliquer le manque de *validité et fidélité de la mesure*. Il faut naturellement tenir compte également du manque de *constance des sujets*, en l'occurrence des élèves. Particulièrement dans cette phase de la préadolescence, ceux-ci évoluent, pas nécessairement de la même façon. Les changements qui en résultent influent sur nos corrélations. Le facteur général d'efficience mentale, plus ou moins stable à cet âge, est modulé positivement ou négativement par des facteurs contingents tels que l'effort, l'intérêt, la motivation (14). Cette dernière peut être très fortement influencée chez certains élèves par la personnalité du maître; un changement de professeur entraînera chez eux une modification d'attitude à l'égard de telle branche. Sans parler d'événements accidentels comme la maladie, la séparation des parents, etc.

Enfin, la *faible dispersion* des notes dans certaines branches ne permet pas de bien discriminer les résultats, c'est-à-dire de distinguer l'élève faible de l'insuffisant, ou l'élève brillant de l'enfant moyennement doué. Sur le plan statistique, cette faible dispersion a pour effet de diminuer la corrélation.

Les conditions d'une amélioration; esquisse de solutions

Dans le cadre de cet article nous ne pouvons qu'esquisser certaines solutions qui découlent de l'analyse des paragraphes précédents.

¹ Ce qui correspond à une moyenne générale d'au moins 4,5.

a) Une réforme de nos modes d'évaluation scolaire doit nécessairement toucher l'ensemble de la scolarité (y compris l'université) pour être efficace, les notes, ou toute autre appréciation, données à une certaine période de la scolarité jouant à la fois le rôle d'éléments prédicteurs pour l'étape suivante et de critères pour le stade antérieur.

b) La première démarche consiste en une analyse précise des objectifs, à tous les degrés, précisant sous forme d'exemples détaillés les performances attendues et le niveau visé. Des manuels bien faits ou une méthodologie clairement présentée peuvent contribuer grandement à unifier les conceptions didactiques et docimologiques des maîtres. Lors de notre enquête sur le passage des élèves du CO dans les écoles secondaires supérieures nous avons trouvé des corrélations particulièrement élevées en mathématiques et en allemand en ce qui concerne les sections gymnasiales; or, ces deux branches bénéficient de manuels et d'une méthodologie mis au point conjointement par des professeurs des deux ordres d'enseignement (CO et Collège de Genève).

c) Une appréciation du rendement scolaire et un pronostic d'adaptation ultérieure supposent une référence à un *modèle psychopédagogique* de l'apprentissage scolaire dans la branche considérée. Le maître doit pouvoir situer son action pédagogique et les comportements de ses élèves par rapport à une épistémologie ainsi qu'en fonction d'une théorie de l'apprentissage et du développement. Nous définissons là une condition de progrès pour la pédagogie en général, dont la docimologie n'est qu'un domaine.

d) Tous les moyens d'appréciations devraient être élaborés et mis au point selon les techniques psychométriques appliquées à la préparation des tests. Il s'agit surtout de « standardiser les conditions d'examen, éliminer les questions ambiguës ou non informatives, simplifier les modes de réponse du sujet, échantillonner méthodiquement les domaines de comportement visés, systématiser la correction pour éviter les divergences d'appréciation, structurer l'interprétation en recourant à des étalonnages et à des équations de régression » (6). En attendant le moment où les maîtres disposeraient de plusieurs batteries de tests pédagogiques, on peut appliquer périodiquement des épreuves générales standardisées fournissant aux professeurs des indications précieuses sur le niveau de leur classe et les aidant à choisir leurs barèmes pour leurs propres travaux de contrôle. A Genève, les épreuves communes ont permis dans certains cas à des enseignants de prendre conscience des différences entre leurs exigences et celles de leurs collègues (4).

e) Pratiquement la préparation de ces tests pédagogiques nécessite l'intervention de spécialistes et d'institutions spécialisées (centres

psychopédagogiques), mais les maîtres devraient être largement associés à ce travail. Nous y voyons la meilleure manière de les familiariser avec ces modes d'évaluation et de les former à les utiliser et interpréter correctement.

f) Dans cette perspective, nous suggérons de supprimer le calcul de moyennes annuelles ou semestrielles comme normes de promotion ; elles ont toujours constitué à nos yeux une hérésie pédagogique : deux moyennes équivalentes peuvent masquer des évolutions très différentes ($2 + 4 + 6 = 6 + 4 + 2$). On craint parfois pareille mesure parce qu'elle favoriserait un certain arbitraire. On oublie que les barèmes dont dépendent les notes procèdent d'un arbitraire tout aussi grand, même s'il est masqué par tout un appareil pseudo-mathématique.

Docimologie et pédagogie ; en guise de conclusion

En discutant avec des collègues des problèmes techniques posés par nos moyens d'évaluation : correction, cotation, barèmes, nous avons été frappés par le fait que presque chaque fois, nous étions obligés de dépasser l'aspect purement docimologique de la question pour envisager — et souvent remettre en cause — la méthodologie de la branche ou même l'ensemble de notre action pédagogique et notre politique d'orientation. Face à l'élaboration d'une épreuve ou à ses résultats, nous sommes contraints d'expliciter nos principes pédagogiques, de les traduire en actes, qui auront des conséquences précises et prévisibles pour nos élèves. Ces discussions nous semblent extrêmement fructueuses, beaucoup plus que des débats très académiques où l'on reste dans le flou et l'indéterminé des idées générales chères à certains pédagogues de l'ancienne école. Ces échanges mettent en évidence non seulement nos divergences d'opinion, mais surtout l'absence de conception claire de notre action pédagogique et de ses objectifs, la pauvreté de notre information dans le domaine psychopédagogique, le besoin de contacts avec les collègues qui recevront nos élèves, etc. La docimologie, au sens où nous l'avons défini, nous apparaît donc le lieu de rencontre par excellence de la pratique et de la théorie, où ces deux aspects se valorisent mutuellement ; la docimologie est un jalon important sur la route vers une pédagogie enfin scientifique.

D. BAIN

Bibliographie sommaire

1. Bain, D. et Chédanne R. G.: *L'orientation des élèves de 7^e générale en 8^e moderne. De la validité des notes scolaires comme critères de sélection.* Centre de recherches psycho-pédagogiques du Cycle d'orientation, Genève, 1969.
2. Bain, D. et Perrenoud, Ph.: *Sélection scolaire et sélection sociale.* CRPP, Cycle d'orientation, Genève, 1968.
3. Bain, D.: *Comparaison entre les collèges du CO à propos d'une épreuve commune de français 7^e/1^{er} trim.* CRPP, Cycle d'orientation, Genève, 1966.
4. Bain, D.: *Les épreuves communes au Cycle d'orientation de l'enseignement secondaire genevois,* CRPP, Cycle d'orientation, Genève, 2^e éd., 1970.
5. Bloom, B. S.: *Taxonomie des objectifs pédagogiques.* Education nouvelle, Montréal, 1969.
6. Cardinet, J.: *Dépasser la sélection scolaire.* GRETI, collection Information, 1969.
7. Ingenkamp, K.: in *Betrifft Erziehung 1969/3*, cité par le *Gymnasium helveticum* 1968-1969, N^o 6.
8. Métraux, G.: Manuel programmé: *Le participe passé.* Service de la recherche section pédagogie, Genève, 1964.
9. Pasquasy, R.: *Les aptitudes et leur mesure.* Editest, 1967.
10. Piéron, H.: *Examens et docimologie.* P.U.F., 1963.
11. Reuchlin, M. et Bacher, F.: *L'orientation à la fin du premier cycle secondaire.* Bibliothèque scientifique, P.U.F., 1969.
12. *Revue internationale: Les sciences de l'éducation*, N^{os} 2-3, 1969: *Docimologie et éducation.* Didier.
13. Rey, A.: *L'examen clinique en psychologie.* P.U.F., 1958.
14. Rey, A.: *De la portée des observations et mesures*, in *Monographies de psychologie clinique.* Delachaux & Niestlé, 1952.
15. Tilton, S.W.: *An educational psychology of learning*, New York, Mac Millan, 1951.