

Le problème des enseignants: quel niveau d'exigence avoir ?

Gérard Piquerez

Présentation de l'article de feu Gérard Piquerez

Jean Cardinet

«La science est un travail de fourmi» disait modestement Gérard Piquerez, en présentant son travail. En vrai scientifique, il savait qu'une bonne théorie a d'abord besoin de faits bien établis, par des recherches forcément pointues. Personne ne devra donc reprocher à l'auteur du texte qui suit de manquer de largeur de vue parce qu'il se restreint à faire deux constatations, sans développer leurs implications idéologiques: les enseignants notent tantôt trop haut, tantôt trop bas, (1) systématiquement d'année en année, mais (2) différemment d'une branche scolaire à l'autre.

L'ambition de Gérard Piquerez était très claire. Il voulait apporter la preuve objective d'une maladie de nos systèmes scolaires suisses: l'arbitraire du jugement des enseignants. Mais on peut tout de même deviner dans son texte d'autres motivations implicites.

Mieux que quiconque, Gérard Piquerez savait que l'école est le creuset de la structure sociale et que les procédures de sélection scolaire décident de l'avenir des individus. Partiellement en charge de l'application de la loi scolaire du Jura, il sentait tout le poids de cette responsabilité. En son âme et conscience, il essayait de la faire fonctionner au mieux.

C'est pourquoi les problèmes de docimologie le préoccupaient tant. Il ne pouvait se satisfaire de voir des enseignants juger souverainement de la valeur des travaux de leurs élèves, sans qu'ils aient de points de repères autres que leur notion personnelle subjective du niveau d'exigence qu'il convenait de maintenir. Qu'ils soient trop sévères, ou trop peu, et ils prêteraient ou favoriseraient alors indûment leurs élèves par rapport à d'autres, dans la course à la formation.

Mais lutter pour plus de justice obligeait d'intenter un procès, et pour l'instruire, il fallait établir les faits de façon indiscutable. C'est ce que Gérard Piquerez a voulu réaliser, tout au long d'une patiente enquête de près d'une dizaine d'années.

Lorsqu'il a pu rassembler des preuves indubitables de la partialité systématique (bien qu'involontaire, évidemment) des enseignants impliqués dans les procédures de

sélection, il a rédigé cet article, qui vise deux buts:

- D'une part, ce texte justifie a posteriori la procédure de correction partielle que l'auteur avait fait adopter dans son canton (où les moyennes de notes de chaque enseignant étaient ajustées à ce que prédisait la moyenne de leur classe à l'épreuve objective).
- D'autre part, dans une perspective future, cet article donne une justification objective pour demander au moins le maintien à l'avenir, dans le Jura, de cet ajustement socialement indispensable. Il suggère en même temps d'appliquer aussi ailleurs la même procédure.

Le «grain de sable» que Gérard Piquerez a voulu apporter à la construction de l'édu-métrie n'a donc rien d'anodin. Son auteur y a travaillé, on peut dire, jusqu'à son dernier souffle, laissant ainsi une sorte de testament qui met enseignants et autorités scolaires suisses devant leurs responsabilités. Confrontés à l'évidence scientifiquement établie de l'injustice des notes, que vont-ils faire? Maintenir ce système inique, de jugement sélectif sans avocat et sans appel? Ou ouvrir enfin les yeux, en regardant ce qui se fait ailleurs, dans les pays où tous les enfants ont le droit de continuer à apprendre à leur rythme?

Le problème des enseignants: quel niveau d'exigence avoir ?

Gérard Piquerez

Notre étude porte sur les « corrections de notes » d'enseignants jurassiens de 6P, calculées par le truchement de tests standardisés. Malgré le grand nombre de variables contextuelles non maîtrisées, nos analyses montrent que les barèmes de notes de chaque enseignant sont très stables dans le temps et qu'ils sont propres à chaque branche. Un modèle simple de correction de « l'équation personnelle » des enseignants est proposé ici.

Introduction

Bessel (1823), l'astronome qui calcula le premier la distance de la Terre aux étoiles, puis Piéron (1963) l'inventeur de la docimologie, ont affirmé, chacun à leur manière, que toute mesure ou évaluation était entachée d'une erreur, appelée par Bessel¹ « l'équation personnelle » et définie ici comme la façon dont un observateur, en raison de ses caractéristiques personnelles, organise ou déforme des faits qu'il rapporte ou apprécie. Cette équation semble bien dépendre de l'examineur, de l'instrument de mesure, mais aussi de l'objet et/ou du sujet évalué et de son environnement. L'erreur d'un astronome braquant les yeux sur la Lune avec un télescope n'est pas la même lorsqu'il observe un autre corps céleste avec ce même instrument. Des différences apparaissent surtout si l'on compare les observations des mêmes faits, effectuées par les mêmes personnes, ceci sur de longues périodes. Bessel a mis en évidence que les mesures d'astronomes qualifiés et soigneux sont biaisées par des facteurs personnels sans que ces personnes en prennent conscience. Il ne fut pas seul à le constater.

Dans cette perspective, nous avons étudié le biais personnel des enseignants jurassiens de 6e primaire dans la mise de notes semestrielles. En nous appuyant sur une épreuve commune standardisée, nous avons déterminé le niveau objectif de chaque classe et calculé la note moyenne qu'elle « devrait » recevoir (en faisant abstraction des nombreux facteurs qui modulent légitimement cette valeur attendue).

L'écart entre cette note prédite et la note moyenne effective a été considéré par nous comme la mesure de l'équation personnelle de l'enseignant en question. En

soustrayant cet effet de chaque note, il était possible d'annuler cette source d'erreur. C'est pourquoi nous avons appelé l'écart ci-dessus «la correction de notes».

La question que nous nous sommes posée a été de savoir si cet effet de l'équation personnelle de chaque enseignant était stable et systématique ou variable et aléatoire. Nous avons aussi voulu déterminer s'il était général, ou bien lié à la branche scolaire évaluée.

Contexte, buts de l'étude et méthode d'analyse

La structure de la scolarité obligatoire du canton du Jura présente six années d'école primaire et trois d'école secondaire I. La Loi scolaire de ce canton suisse stipule que le sixième degré a pour fonction particulière d'observer et d'orienter les élèves en vue des enseignements différenciés pratiqués à l'école secondaire. Cette dernière comprend dès la septième des cours à niveaux (A, B, C) en allemand, mathématiques et français. Les autres leçons y sont données dans le cadre d'options ou de cours hétérogènes.

L'ordonnance qui porte application de la loi scolaire précise qu'en 6^e les élèves sont soumis à trois séries d'épreuves communes dans les disciplines de base, français, mathématiques et allemand. Selon cette ordonnance, les résultats de ces épreuves, ceux des bulletins scolaires ainsi que l'avis des parents fondent l'appréciation des élèves pour accéder aux cours à niveaux de l'école secondaire dans les branches citées ci-dessus. Les textes légaux stipulent que les épreuves communes (EC) sont standardisées et corrigées conformément à un barème cantonal. Trois groupes de rédacteurs construisent ces épreuves constituées uniquement de QCM à 5 choix dont l'un est exact. Le programme scolaire sur lequel portent les EC est clairement cerné et délimité pour les mathématiques et l'allemand, mais présente plus d'approximations en français. Lors des épreuves de fin janvier et de juin, les élèves répondent - pour chaque branche testée - à 40 ou 50 questions posées par écrit, mais aussi oralement, ceci au moyen d'un enregistrement sur cédérom. Ils reportent leurs réponses sur des feuilles préétablies lues par la suite à l'aide d'un lecteur optique.

Les résultats obtenus aux EC et les notes de l'année doivent selon l'ordonnance être pris en compte de manière équivalente, ceci pour l'orientation vers les cours à niveaux de l'école secondaire I. Toutefois, avant la mise en place de cette orientation, on a fait l'hypothèse que certains maîtres ont tendance à sous-évaluer globalement leurs élèves, alors que d'autres ont l'inclinaison contraire. Comment corriger cette pension?

Cardinet et Bader (1958) apportent une réponse à cette interrogation après avoir posé la question suivante: «Faut-il estimer les connaissances scolaires des enfants d'après les notes brutes données par leur instituteur ou bien faut-il évaluer leurs résultats d'après la moyenne des notes habituellement données par cet instituteur?». La solution qu'ils proposent vise à combiner les avantages de cha-

cune de ces deux approches. Elle postule l'existence d'une droite de régression commune à toutes les classes entre le prédicteur (résultat à une épreuve commune) et la note donnée par l'enseignant, qui est alors le critère; les auteurs proposent d'utiliser cette fonction linéaire pour modifier les notes données par chaque enseignant au moyen d'une correction appropriée, valable pour tous les élèves de sa classe.

En application de cette proposition, chaque enseignant jurassien reçoit une «correction» de chaque note semestrielle qu'il a donnée à ses élèves: une pour chaque branche testée. Grâce à cette correction, les barèmes appliqués dans chaque classe pour mettre des notes sont en accord avec les résultats obtenus par la classe aux épreuves communes. L'information spécifique que possède l'enseignant du fait qu'il a la possibilité de comparer ses élèves entre eux reste néanmoins conservée et exploitée.

Plus précisément, les résultats bruts de la deuxième EC et les notes à la fin du premier semestre (janvier) sont pris en compte dans le calcul des corrections de notes de fin janvier [= c1-]. On fait de même avec la troisième épreuve (juin) et les notes du 2^e semestre [= c2-]. Ces éléments sont ensuite utilisés pour élaborer une proposition quant à l'accès aux cours à niveaux. Ceci constitue le premier pas d'une orientation qui se veut plus personnalisée dans le cadre du secondaire I. Signalons que le bien-fondé de cette façon d'orienter les élèves ne sera pas discutée ici.

Bien que le système fonctionne depuis 1994, les données analysées ci-dessous portent sur sept années, de 1998/9 à 2004/5. Une petite trentaine d'enseignants ont été en fonction sans discontinuer pendant les sept ans; ces instituteurs sont en presque totalité des maîtres à la tête de classes à 2 degrés, 5P et 6P. Les constantes de «correction de notes» qui leur avaient été appliquées ont été conservées. Elles ont permis les analyses comparatives présentées maintenant.

Notons que le but premier de cette étude est de suivre la sévérité de chaque enseignant à travers les années et d'apprécier son indépendance par rapport aux volées d'élèves qu'il reçoit. En second lieu, nous cherchons aussi à savoir si ce degré de sévérité est propre à chaque branche ou s'il se retrouve semblable dans toutes les branches.

En ce qui concerne notre première question, l'analyse des données se fera par un calcul des corrélations entre les «corrections de notes» de sept années successives dans les trois disciplines (mathématiques, allemand et français), ceci pour des ensembles de 26 et 27 enseignants. On testera l'hypothèse d'une corrélation nulle par rapport à une corrélation positive. Le seuil choisi pour un test unilatéral est ici fixé à $p < 0.05$.

Afin de répondre à notre seconde question, nous calculerons les corrélations entre les corrections de notes relatives à des disciplines différentes (mathématiques, allemand et français), ceci pour des ensembles de 26 ou 27 enseignants (cf. tableau 2). Nous testerons l'hypothèse d'une corrélation nulle par un test bilatéral fixé à $p < 0.05$.

La stabilité des barèmes des enseignants dans chaque discipline

Le tableau 1 met en évidence les corrélations élevées qui existent entre corrections de notes des enseignants pour la branche des mathématiques.

Tableau 1: *Corrélations entre les corrections de notes en mathématiques*

Corrélations significatives à $p < 0,05$ (test unilatéral) marquées en gras
N=27

	C1-99	C2-99	C1-00	C2-00	C1-01	C2-01	C1-02	C2-02	C1-03	C2-03	C1-04	C2-04	C1-05	C2-05
C1-99	1.00	0.44	0.08	0.34	0.60	0.50	0.58	0.37	0.02	0.25	0.19	0.37	0.41	0.30
C2-99	0.44	1.00	0.20	0.47	0.52	0.63	0.70	0.54	0.29	0.62	0.51	0.41	0.42	0.51
C1-00	0.08	0.20	1.00	0.48	0.27	0.34	0.38	0.05	0.30	0.36	0.39	0.35	0.35	0.49
C2-00	0.34	0.47	0.48	1.00	0.53	0.71	0.45	0.37	0.35	0.60	0.24	0.42	0.25	0.33
C1-01	0.60	0.52	0.27	0.53	1.00	0.64	0.51	0.57	0.21	0.50	0.42	0.47	0.49	0.35
C2-01	0.50	0.63	0.34	0.71	0.64	1.00	0.70	0.41	0.17	0.53	0.46	0.54	0.35	0.51
C1-02	0.58	0.70	0.38	0.45	0.51	0.70	1.00	0.35	0.06	0.46	0.62	0.68	0.60	0.53
C2-02	0.37	0.54	0.05	0.37	0.57	0.41	0.35	1.00	0.38	0.45	0.24	0.41	0.56	0.58
C1-03	0.02	0.29	0.30	0.35	0.21	0.17	0.06	0.38	1.00	0.69	0.28	0.07	0.18	0.39
C2-03	0.25	0.62	0.36	0.60	0.50	0.53	0.46	0.45	0.69	1.00	0.51	0.34	0.27	0.43
C1-04	0.19	0.51	0.39	0.24	0.42	0.46	0.62	0.24	0.28	0.51	1.00	0.57	0.52	0.57
C2-04	0.37	0.41	0.35	0.42	0.47	0.54	0.68	0.41	0.07	0.34	0.57	1.00	0.68	0.48
C1-05	0.41	0.42	0.35	0.25	0.49	0.35	0.60	0.56	0.18	0.27	0.52	0.68	1.00	0.57
C2-05	0.30	0.51	0.49	0.33	0.35	0.51	0.53	0.58	0.39	0.43	0.57	0.48	0.57	1.00

C1-99 = ensemble des corrections correspondant à l'épreuve commune et à la note semestrielle de janvier de l'année scolaire finissant en 1999.

C2-99 = ensemble des corrections correspondant à l'épreuve commune et à la note semestrielle de juin de l'année scolaire finissant en 1999. etc.

On constate dans le tableau 1 que le nombre de corrélations significatives atteint 78%. Ceci prouve que dans l'ensemble des enseignants que nous avons pu suivre pendant 7 années, de 1999 à 2005, ceux qui avaient tendance à noter plus sévèrement que les autres les travaux de mathématiques au début de cette période continuaient à le faire en fin de période. Inversement, les notateurs indulgents tendaient à donner durant les sept années des notes supérieures à ce que le niveau objectif de leur classe justifiait.

Tableau 2: Résumé (voir annexes, tableaux 5 et 6)

Signification	Branches	Mathématiques	Allemand	Français
p<0.05 Test unilatéral	7 années suivies (14 corrections de 1998/9 à 2004/05; 2 par année)	27 enseignants	26 enseignants	27 enseignants
		significatif dans 78%	significatif dans 79%	significatif dans 64%

Les corrélations pour chacune des branches sont très élevées. Ce qui démontre clairement notre première hypothèse, à savoir que la sévérité de chaque enseignant à travers les années est stable et indépendante par rapport aux volées d'élèves qu'il reçoit.

On aurait pu penser que ces enseignants, sachant après les premières années que l'on avait corrigé leurs notes dans le sens de plus de sévérité, ou de plus de clémence, auraient rectifié d'eux-mêmes le niveau de leur barème personnel. Leurs «corrections de notes» initiales devraient prédire de moins en moins bien leurs «corrections de notes» en fin de période. Il suffit de regarder l'évolution des corrélations ligne après ligne au-dessus de la diagonale pour vérifier qu'il n'en est rien. Le classement des enseignants du point de vue sévérité reste stable, aussi semblable après six ans qu'après six mois à ce qu'il était lors de leur première notation.

En français, par rapport aux autres branches, on constate un léger affaiblissement dans le pourcentage de corrélations significatives. On pourrait expliquer ceci en notant que les comportements attendus des élèves dans cette branche sont moins bien définis qu'en mathématiques ou en allemand. L'évaluation par notes serait ici plus sujette à certains aléas dans cette discipline. La validité de l'épreuve commune pour prédire la note à l'intérieur de la classe serait ici aussi un peu plus faible.

La stabilité des barèmes des enseignants dans des disciplines différentes

Le tableau 3 ci-dessous met en évidence les corrélations peu élevées qui existent entre corrections de notes des enseignants en allemand et en mathématiques. Le seuil choisi, pour un test bilatéral dans ce cas, est fixé à p<0.05.

Tableau 3: Corrélations entre les corrections de notes des enseignants, en allemand et mathématiques

Corrélations significatives à $p < 0.05$ marquées en gras N=26
 Les colonnes représentent l'allemand et les lignes les mathématiques.

	C1-99	C2-99	C1-00	C2-00	C1-01	C2-01	C1-02	C2-02	C1-03	C2-03	C1-04	C2-04	C1-05	C2-05
C1-99	0.68	0.18	-0.07	0.18	0.50	0.35	0.23	0.06	-0.08	0.10	-0.20	-0.00	0.14	-0.18
C2-99	0.60	0.27	-0.01	0.14	0.52	0.34	0.28	0.29	-0.05	-0.03	-0.06	-0.05	0.29	-0.04
C1-00	0.39	0.00	0.14	0.06	0.30	0.14	0.08	0.07	0.07	0.00	-0.00	-0.11	-0.09	0.01
C2-00	0.48	0.16	0.24	0.27	0.34	0.33	0.21	0.12	0.23	0.15	0.12	-0.07	0.08	0.07
C1-01	0.41	0.07	-0.04	0.03	0.19	0.05	0.31	-0.29	-0.36	-0.05	-0.01	-0.16	-0.03	-0.38
C2-01	0.27	0.13	0.10	0.10	-0.04	0.26	0.17	-0.27	-0.11	0.11	-0.09	-0.22	-0.34	-0.16
C1-02	0.70	0.20	0.09	0.19	0.48	0.19	0.26	0.23	-0.18	-0.02	-0.15	0.04	0.19	-0.02
C2-02	0.52	0.30	-0.09	0.02	0.32	0.09	0.14	0.08	-0.13	-0.01	-0.19	-0.26	0.02	-0.17
C1-03	0.43	0.32	0.29	0.11	0.45	0.28	0.34	0.03	0.17	0.16	0.35	-0.06	0.24	0.26
C2-03	0.52	0.46	-0.00	0.05	0.52	0.26	0.31	0.25	0.07	0.10	0.32	0.03	0.47	0.22
C1-04	0.14	0.05	-0.05	0.07	0.29	0.03	-0.03	-0.03	-0.14	0.02	-0.04	-0.43	-0.35	-0.26
C2-04	0.20	0.11	-0.12	-0.04	0.09	0.10	0.14	-0.19	-0.31	-0.01	-0.21	-0.29	-0.34	-0.31
C1-05	0.43	0.13	0.05	0.22	0.31	0.39	0.12	0.08	0.03	0.22	-0.04	0.08	-0.03	-0.13
C2-05	0.51	-0.01	0.15	0.23	0.41	0.19	0.11	0.15	0.11	0.15	-0.13	-0.07	0.10	-0.10

On constate dans le tableau 3 que le nombre de corrélations significatives ne dépasse pas 11%

Tableau 4: Résumé (cf. annexes, tableaux 7 et 8)

Corrélations significatives à $p < 0.05$ en pour cent, entre les corrections de notes des enseignants pour des branches différentes (test bilatéral)

	Significations	Branches	Mathématiques	Allemand	Français
26 enseignants					
7 années suivies (14 corrections de 1998/9 à 2004/05; 2 par année)	$p < 0.05$	Mathématiques	-	sign. dans 11%	sign. dans 18%
	$p < 0.05$	Allemand	sign. dans 11%	-	sign. dans 34%
	$p < 0.05$	Français	sign. dans 18%	sign. dans 34%	-

Les corrélations pour chacune des branches sont basses, sauf entre le français et l'allemand (tableau 7 en annexe), deux disciplines plutôt littéraires. De plus, on voit apparaître des corrélations négatives. Que signifient-elles? Peut-être une sorte d'orientation de l'enseignant qui privilégierait les disciplines littéraires ou scientifiques? On constate aussi dans le tableau 3 que la 1^e colonne (c1-99) et la

5^e (c1-01) ont des corrélations élevées, les autres sont pratiquement nulles. Ceci peut s'expliquer par un changement fondamental dans la méthode d'enseignement de l'allemand. Certaines épreuves communes ont probablement créé des problèmes de prédiction des notes.

En tenant compte de ces remarques, on peut affirmer que ceci démontre formellement notre deuxième hypothèse, à savoir que le degré de sévérité est propre à chaque branche.

Discussion

Confrontons maintenant nos buts aux résultats de l'analyse. La forte proportion de corrélations significatives obtenue dans chacune des branches tend à confirmer notre hypothèse d'un degré de sévérité stable au cours des années. Pour chaque enseignant, lors de chaque phase de mise de notes semestrielles, la «correction de notes» représente une partie relativement stable de la note moyenne de la classe, ceci indépendamment de la «volée» (cohorte) d'élèves.

On peut même étendre ce modèle et l'appliquer à chacune des évaluations durant l'année scolaire. Cette stabilité serait une dimension cachée de la mise de notes, dont l'enseignant ne serait pas conscient. Elle pourrait apparaître comme caractéristique des modes de faire qu'il développe par la pratique. Pourrait-on les interpréter comme des schèmes d'action relatifs à la conduite de sa classe et illustrant des caractéristiques du schème, les invariants opératoires (Vergnaud, 1985)? Les modes de faire seraient alors des savoir-faire.

La stabilité de la sévérité de l'enseignant indépendamment des volées d'élèves pourrait traduire une stratégie didactique. Dans la mesure où ce dernier peut présager du niveau de difficulté des épreuves qu'il applique, il modulerait ses exigences par leur intermédiaire. Il rassurerait ou inciterait à l'effort et maîtriserait ainsi la négociation didactique au sens de Chevallard (1986). Choisir le bon niveau de difficulté d'une tâche est très certainement nécessaire dès les premiers jours de pratique de l'enseignement, un choix inadapté entraînant des perturbations dans le déroulement de la vie de la classe. Des habiletés se développeraient donc en réponse à ce type de contrainte.

On peut, dans ces conditions, imaginer que ce pilotage par le choix du niveau de difficulté des tâches mène à des automatismes progressivement rendus invariables et ritualisés par renforcements successifs internes et externes à la classe et s'actualisant par un degré de sévérité fixé lors des évaluations. Par voie de conséquence, les «corrections de notes» deviendraient ainsi quasi constantes au cours des années.

Le deuxième aspect de notre étude met en évidence que le degré de sévérité est propre à la branche examinée. Selon Felouzis (1997), l'effet de contexte n'explique pas toute la différence entre les classes; cet auteur cite notamment l'importance des caractéristiques propres à l'enseignant (milieu socioculturel, expé-

rience professionnelle, idéologie pédagogique, etc.) Ainsi, l'orientation Mathématiques/Langues qui semble apparaître expliquerait que ses «corrections de notes» soient à la fois stables et différentes d'une branche à une autre. En conséquence, ses comportements d'évaluation et, de fil en aiguille, ses «corrections de notes» seraient un peu différents de celles des autres enseignants.

Conclusion

Bressoux et Pansu (2003) affirment que «pour peu qu'on connaisse le score d'un élève à des épreuves standardisées et le niveau moyen de sa classe, on peut, à partir de ces deux principes, prédire la valeur scolaire que lui attribuera un enseignant». Cardinet et Bader, en 1958 déjà, avaient donné le moyen de relier ces variables. Ce lien donne lieu à ce qu'on appelle ici la «correction de notes». Par ce qui précède, on a étayé l'hypothèse qu'elle exprime la sévérité de chaque enseignant, ceci différemment pour chaque branche examinée: lors de chaque phase de mise de notes semestrielles, la «correction de notes» représente une partie relativement stable de la note moyenne de la classe, ceci indépendamment du groupe d'élèves évalués.

Notes

- 1 Pour Bessel: l'intervalle de temps entre le moment où un phénomène se produit réellement et le moment énoncé par l'observateur (Van Rillaer, 2003).

Références bibliographiques

- Bressoux, P. & Pansu, P. (2003). *Quand les enseignants jugent leurs élèves*. Paris: Presses Universitaires de France.
- Cardinet, J. & Bader, R. (1958). Étalonnages différents ou correction du pronostic? *Bulletin du Centre d'Études et Recherches Psychotechniques*, 23 (7), 181-187.
- Chevallard, Y. (1986). Vers une analyse didactique des faits d'évaluation. In J.-M. de Ketele (Dir.), *L'évaluation descriptive ou prescriptive*. Bruxelles: de Boeck.
- Felouzis, G. (1997). *L'efficacité des enseignants*. Paris: Presses Universitaires de France.
- Piéron, H. (1963). *Examens et Docimologie*. Paris: Presses Universitaires de France.
- Van Rillaer, J. (2003). *Psychologie de la vie quotidienne*. Paris: Odile Jacob.
- Vergnaud, G. (1985). Concepts et schèmes dans une théorie opératoire de la représentation. *Psychologie Française*, 30-34, 245-252.

Mots clés: docimologie, tests standardisés, orientation scolaire, sélection, étude longitudinale

Annexes

Tableau 5: Corrélations entre les corrections de notes en allemand

Corrélations significatives à $p < 0.05$ (test unilatéral) marquées en gras
N = 26

	C1-99	C2-99	C1-00	C2-00	C1-01	C2-01	C1-02	C2-02	C1-03	C2-03	C1-04	C2-04	C1-05	C2-05
C1-99	1.00	0.74	0.32	0.39	0.48	0.30	0.66	0.63	0.48	0.54	0.27	0.26	0.45	0.56
C2-99	0.74	1.00	0.59	0.69	0.48	0.30	0.59	0.66	0.58	0.68	0.35	0.31	0.36	0.50
C1-00	0.32	0.59	1.00	0.82	0.34	0.48	0.23	0.51	0.42	0.40	0.46	0.37	0.41	0.56
C2-00	0.39	0.69	0.82	1.00	0.41	0.45	0.38	0.55	0.54	0.47	0.45	0.36	0.59	0.68
C1-01	0.48	0.48	0.34	0.41	1.00	0.52	0.54	0.57	0.45	0.34	0.59	0.75	0.21	0.43
C2-01	0.30	0.30	0.48	0.45	0.52	1.00	0.07	0.38	0.15	0.07	0.37	0.68	0.43	0.22
C1-02	0.66	0.59	0.23	0.38	0.54	0.07	1.00	0.61	0.51	0.35	0.47	0.37	0.35	0.62
C2-02	0.63	0.66	0.51	0.55	0.57	0.38	0.61	1.00	0.52	0.63	0.55	0.52	0.45	0.63
C1-03	0.48	0.58	0.42	0.54	0.45	0.15	0.51	0.52	1.00	0.73	0.49	0.30	0.18	0.35
C2-03	0.54	0.68	0.40	0.47	0.34	0.07	0.35	0.63	0.73	1.00	0.23	0.09	0.14	0.31
C1-04	0.27	0.35	0.46	0.45	0.59	0.37	0.47	0.55	0.49	0.23	1.00	0.65	0.37	0.49
C2-04	0.26	0.31	0.37	0.36	0.75	0.68	0.37	0.52	0.30	0.09	0.65	1.00	0.37	0.30
C1-05	0.45	0.36	0.41	0.59	0.21	0.43	0.35	0.45	0.18	0.14	0.37	0.37	1.00	0.64
C2-05	0.56	0.50	0.56	0.68	0.43	0.22	0.62	0.63	0.35	0.31	0.49	0.30	0.64	1.00

79% de corrélations significatives à $p < 0.05$ ($r \geq 0.33$)

C1-01 = ensemble des corrections correspondant à l'épreuve commune et à la note semestrielle de janvier de l'année scolaire finissant en 2001.

C2-01 = ensemble des corrections correspondant à l'épreuve commune et à la note semestrielle de juin de l'année scolaire finissant en 2001.

On constate dans le tableau 5 que le nombre de corrélations significatives atteint 79% et qu'il est donc aussi élevé que pour les notes de mathématiques. On voit ainsi à nouveau que dans l'ensemble des enseignants que nous avons pu suivre de 1999 à 2005, ceux qui avaient tendance à noter plus sévèrement que les autres les résultats en allemand au début de cette période continuaient à le faire en fin de période. Inversement, les notateurs indulgents donnaient durant les sept années des notes supérieures à celles attribuées par leurs collègues pour des performances comparables.

Ces enseignants n'ont pas rectifié d'eux-mêmes le niveau de leur barème personnel. Si l'on regarde l'évolution des corrélations dans chaque ligne, on voit que le classement des enseignants du point de vue sévérité reste stable à travers les années.

Le tableau 6 ci-dessous présente les corrélations relatives à la troisième discipline évaluée pour l'orientation.

Tableau 6: *Corrélations entre les corrections de notes en français*

Corrélations significatives à $p < 0.05$ (test unilatéral) marquées en gras

	C1-99	C2-99	C1-00	C2-00	C1-01	C2-01	C1-02	C2-02	C1-03	C2-03	C1-04	C2-04	C1-05	C2-05
C1-99	1.00	0.51	0.41	0.29	0.26	0.15	0.18	0.26	0.16	0.05	0.15	-0.00	0.36	0.33
C2-99	0.51	1.00	0.43	0.50	0.56	0.65	0.20	0.36	0.35	0.43	0.43	0.57	0.59	0.45
C1-00	0.41	0.43	1.00	0.65	0.56	0.25	0.01	0.00	0.33	0.11	0.24	0.36	0.47	0.32
C2-00	0.29	0.50	0.65	1.00	0.65	0.46	0.36	0.18	0.36	0.29	0.24	0.61	0.31	0.33
C1-01	0.26	0.56	0.56	0.65	1.00	0.73	0.48	0.17	0.52	0.36	0.44	0.71	0.56	0.43
C2-01	0.15	0.65	0.25	0.46	0.73	1.00	0.37	0.29	0.39	0.62	0.37	0.74	0.52	0.64
C1-02	0.18	0.20	0.01	0.36	0.48	0.37	1.00	0.50	0.60	0.21	0.30	0.34	0.21	0.22
C2-02	0.26	0.36	0.00	0.18	0.17	0.29	0.50	1.00	0.42	0.52	0.29	0.10	0.21	0.13
C1-03	0.16	0.35	0.33	0.36	0.52	0.39	0.60	0.42	1.00	0.42	0.44	0.25	0.39	0.15
C2-03	0.05	0.43	0.11	0.29	0.36	0.62	0.21	0.52	0.42	1.00	0.24	0.49	0.47	0.59
C1-04	0.15	0.43	0.24	0.24	0.44	0.37	0.30	0.29	0.44	0.24	1.00	0.44	0.56	0.30
C2-04	-0.00	0.57	0.36	0.61	0.71	0.74	0.34	0.10	0.25	0.49	0.44	1.00	0.53	0.63
C1-05	0.36	0.59	0.47	0.31	0.56	0.52	0.21	0.21	0.39	0.47	0.56	0.53	1.00	0.64
C2-05	0.33	0.45	0.32	0.33	0.43	0.64	0.22	0.13	0.15	0.59	0.30	0.63	0.64	1.00

On constate dans le tableau 6 que le nombre de corrélations significatives atteint 64%, mais qu'il est un peu moins élevé que dans les deux autres branches, sans pour autant les contredire. On voit donc à nouveau que dans l'ensemble des enseignants que nous avons pu suivre de 1999 à 2005, ceux qui avaient tendance à noter plus sévèrement que les autres les résultats en français au début de cette période continuaient à le faire en fin de période.

Tableau 7: Corrélations entre les corrections de notes en français et en allemand

Corrélations significatives à $p < 0.05$ (test bilatéral) marquées en gras N=26
Les colonnes représentent l'allemand et les lignes les français

	C1-99	C2-99	C1-00	C2-00	C1-01	C2-01	C1-02	C2-02	C1-03	C2-03	C1-04	C2-04	C1-05	C2-05
C1-99	0.45	0.57	0.43	0.46	0.51	0.41	0.36	0.21	0.46	0.48	0.29	0.36	0.49	0.43
C2-99	0.18	0.42	0.41	0.52	0.55	0.48	0.31	0.11	0.52	0.47	0.05	0.36	0.26	0.26
C1-00	0.09	0.19	0.23	0.34	0.16	0.31	-0.04	-0.09	0.25	0.29	-0.16	0.15	0.03	0.31
C2-00	-0.08	0.17	0.36	0.43	0.36	0.43	0.08	-0.07	0.33	0.36	-0.07	0.27	0.07	0.27
C1-01	0.31	0.40	0.36	0.62	0.57	0.50	0.27	-0.03	0.24	0.38	0.42	0.51	0.39	0.53
C2-01	0.05	0.55	0.19	0.32	0.38	0.46	0.03	-0.15	0.05	0.13	0.20	0.49	0.29	0.40
C1-02	0.47	0.22	0.55	0.53	0.48	0.34	0.35	-0.02	0.30	0.24	0.09	0.24	0.19	0.40
C2-02	0.53	0.54	0.57	0.68	0.36	0.32	0.35	0.27	0.31	0.35	0.08	0.28	0.29	0.41
C1-03	0.23	0.41	0.35	0.51	0.49	0.48	0.44	0.34	0.69	0.56	0.22	0.26	0.24	0.33
C2-03	0.33	0.52	0.34	0.44	0.40	0.37	0.49	0.45	0.65	0.49	0.23	0.25	0.22	0.26
C1-04	0.25	0.19	0.50	0.50	0.31	0.30	0.17	-0.01	0.30	0.17	0.22	0.29	0.17	0.36
C2-04	0.26	0.39	0.25	0.39	0.37	0.41	0.09	-0.05	-0.00	0.24	0.27	0.41	0.33	0.44
C1-05	0.12	0.23	0.29	0.07	0.12	0.27	0.08	-0.06	0.23	0.18	0.20	0.18	0.36	0.38
C2-05	0.34	0.14	0.63	0.47	0.40	0.23	0.19	-0.15	0.27	0.18	0.04	0.21	0.33	0.44

34% de corrélations significatives à $p < 0.05$: ($r \geq 0.39$)

Tableau 8: Corrélations entre les corrections de notes en mathématiques et en français

Corrélations significatives à $p < 0.05$ (test bilatéral) marquées en gras N=27
Les colonnes représentent les mathématiques et les lignes le français

	C1-99	C2-99	C1-00	C2-00	C1-01	C2-01	C1-02	C2-02	C1-03	C2-03	C1-04	C2-04	C1-05	C2-05
C1-99	0.33	0.42	0.36	0.31	0.45	0.52	0.30	0.14	0.35	0.49	0.25	0.34	0.27	0.57
C2-99	-0.05	0.45	0.24	0.23	0.25	0.41	0.38	0.21	0.24	0.25	0.31	0.41	0.15	0.30
C1-00	-0.02	-0.04	0.21	0.10	0.14	0.07	0.02	-0.27	0.10	-0.03	-0.22	0.08	0.06	0.23
C2-00	-0.14	0.20	0.43	0.19	0.24	0.31	-0.05	-0.07	0.04	0.36	0.03	0.43	0.23	0.41
C1-01	0.23	0.22	0.47	0.25	0.37	0.26	0.35	0.20	0.51	0.45	0.34	0.34	0.32	0.46
C2-01	-0.17	0.39	0.46	0.08	0.38	0.41	0.02	0.03	0.32	0.41	0.36	0.43	0.51	0.41
C1-02	0.01	0.37	0.24	0.21	0.45	0.58	0.29	0.13	0.27	0.40	0.49	0.43	0.48	0.57
C2-02	0.02	-0.02	0.14	-0.02	0.12	0.09	0.23	0.00	0.06	0.02	-0.07	0.11	-0.05	0.06
C1-03	-0.34	-0.09	-0.16	-0.01	0.04	0.02	0.27	0.07	0.17	0.12	-0.10	0.06	-0.24	-0.14
C2-03	-0.25	0.18	0.04	0.06	0.12	0.21	0.18	-0.05	0.07	0.28	0.31	0.44	0.13	0.34
C1-04	-0.29	-0.01	-0.00	-0.08	0.09	0.15	0.20	0.19	0.39	0.15	0.50	0.08	0.05	0.12
C2-04	-0.04	-0.01	-0.03	-0.18	0.02	0.13	0.17	0.11	0.14	0.13	0.31	0.06	0.26	0.22
C1-05	0.19	0.13	0.09	0.02	0.33	0.25	0.44	0.20	0.37	0.22	0.17	0.13	0.26	0.19
C2-05	-0.04	0.04	0.08	-0.19	0.12	0.16	0.03	0.07	0.11	0.08	-0.09	-0.01	-0.00	0.11

18% de corrélations significatives à $p < 0.05$ ($r \geq 0.38$)

Stabile Einstufungsstandards für eine objektivierte Schüler/innenbeurteilung

Zusammenfassung

Die vorliegende Studie diskutiert die Notengebung und deren „Korrektur“ durch den Bezug standardisierter Tests. Durchgeführt wurde die Studie bei jurassischen Lehrkräften des 6. Schulstufe. Trotz einiger nicht kontrollierter Kontextvariablen, kann die Studie zeigen, dass sich die Notenskalen der Lehrkräfte über die Zeit und bezogen auf unterschiedliche Bereiche als sehr stabil erweisen. Vorgeschlagen wird ein einfaches Modell, um den „persönlichen Vergleich“ bei Lehrkräften als Fehlerquelle zu korrigieren und eine objektivierte Schüler/innenbeurteilung zu ermöglichen.

Schlagnworte: objektivierte Schüler/innenbeurteilung, standardisierte Test, Schüler/innenorientierung, Längsschnittstudie

Una sévèrità stabile nella valutazione o «l'equazione personale» degli insegnanti.

Riassunto

Il nostro studio verte sulle «correzioni di note» degli insegnanti jurassiani del sesto anno di scuola elementare, calcolate attraverso test standardizzati. Nonostante il numero di variabili contestuali non controllate, le nostre analisi mostrano che la scale delle note degli insegnanti sono molto stabili nel tempo e che queste sono specifiche ad ogni singola materia. In questo contributo è proposto un modello semplice di correzione dell' «equazione personale» degli insegnanti.

Parole chiave: docimologia, test standardizzati, orientazione scolastica, selezione, studio longitudinale

Stable grading standards for school marks

Abstract

Grades given by teachers in the course of the 6th year of Swiss region Jura's primary schools are routinely «adjusted» on the basis of standardized tests. These corrections were the object of the presented study. In spite of the many variables in the context that remained uncontrolled, the grading standards of each teacher could be shown to remain stable over time, while differing from one subject matter to the other. A simple correction of each teacher's «personal equation» is proposed, as a means to keep these sources of error in check.

Key words: Evaluation, standardized tests, selection, long-term study