

Étude de fiabilité d'un instrument d'observation des comportements de l'élève en classe

Mariel Leclerc, Richard Bertrand et Jocelyne Roberge-Brassard

Volume 5, numéro 3, automne 1979

URI : <https://id.erudit.org/iderudit/900116ar>

DOI : <https://doi.org/10.7202/900116ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Revue des sciences de l'éducation

ISSN

0318-479X (imprimé)

1705-0065 (numérique)

[Découvrir la revue](#)

Citer cet article

Leclerc, M., Bertrand, R. & Roberge-Brassard, J. (1979). Étude de fiabilité d'un instrument d'observation des comportements de l'élève en classe. *Revue des sciences de l'éducation*, 5(3), 359–372. <https://doi.org/10.7202/900116ar>

Résumé de l'article

Après avoir décrit l'instrument et le schéma utilisés pour l'observation des comportements de l'élève en classe, on présente les résultats de l'analyse statistique effectuée pour s'assurer de la fiabilité des données. Pour chacun des trois comportements observés, un coefficient d'entente inter-juges a été calculé pour évaluer dans quelle mesure la mésestimation inter-juges pouvait limiter la fiabilité des données et un coefficient de fiabilité a été calculé pour mesurer la fiabilité avec laquelle les comportements des élèves pouvaient être observés. Les valeurs du coefficient d'entente inter-juges se sont montrées très élevées alors que les valeurs du coefficient de fiabilité se sont montrées très faibles. Ce faible coefficient est expliqué en grande partie par l'instabilité des comportements de l'élève d'occasion en occasion : c'est-à-dire par la trop grande variabilité des données intra-élève par rapport à la variabilité inter-élèves. Les auteurs concluent de l'importance des situations homogènes d'apprentissage dans l'observation des comportements des élèves en classe.

Étude de fiabilité¹ d'un instrument d'observation des comportements de l'élève en classe

Mariel Leclerc, Richard Bertrand et Jocelyne Roberge-Brassard*

RÉSUMÉ

Après avoir décrit l'instrument et le schéma utilisés pour l'observation des comportements de l'élève en classe, on présente les résultats de l'analyse statistique effectuée pour s'assurer de la fiabilité des données. Pour chacun des trois comportements observés, un coefficient d'entente inter-juges a été calculé pour évaluer dans quelle mesure la mésestimation inter-juges pouvait limiter la fiabilité des données et un coefficient de fiabilité a été calculé pour mesurer la fiabilité avec laquelle les comportements des élèves pouvaient être observés. Les valeurs du coefficient d'entente inter-juges se sont montrées très élevées alors que les valeurs du coefficient de fiabilité se sont montrées très faibles. Ce faible coefficient est expliqué en grande partie par l'instabilité des comportements de l'élève d'occasion en occasion : c'est-à-dire par la trop grande variabilité des données intra-élève par rapport à la variabilité inter-élèves. Les auteurs concluent de l'importance des situations homogènes d'apprentissage dans l'observation des comportements des élèves en classe.

* Leclerc, Mariel : professeur, INRS-éducation
Bertrand, Richard : professeur, INRS-éducation
Roberge-Brassard, Jocelyne : professeur, INRS-éducation

Peut-on se fier aux données obtenues à l'aide d'un instrument d'observation ? Certains comportements de l'élève en classe varient-ils d'une occasion à l'autre au point qu'on ne puisse pas se fier aux résultats de l'observation, justement à cause de leur instabilité ? Voilà le genre de questions que nous nous posons². Après avoir décrit l'instrument et le schéma d'observation, cet article présente l'analyse statistique des données recueillies au moyen de cet instrument.

L'INSTRUMENT D'OBSERVATION

En 1975-76 et 1976-77, dans le cadre de l'évaluation de SAGE³, nous avons utilisé l'instrument d'observation de la classe développé par Jane Stallings et ses collaborateurs⁴. En plus d'en présenter une traduction, nous avons fait rapport d'une étude de l'applicabilité de cet instrument en milieu francophone⁵.

En 1977-78, toujours dans le cadre de l'évaluation de SAGE, nous avons voulu limiter à quelques variables nos observations des comportements de l'élève en classe. Nous voulions de plus simplifier l'utilisation d'un instrument d'observation assez complexe, selon nous, et éviter, par le fait même, les causes d'erreur. Ainsi, en centrant nos observations sur trois variables, nous souhaitons nous rapprocher de la réalité et présenter des résultats fiables.

Définition des variables

— La persistance au travail

Elle est définie comme étant le fait pour l'élève d'être engagé dans un travail d'auto-instruction. Si l'élève converse avec quelqu'un durant son travail, il ne s'agit plus de persistance au travail.

La manière pour l'observateur d'enregistrer la fréquence de ce comportement consiste à donner la valeur de une (1) fréquence à chaque intervalle de cinq secondes où le comportement est présent (l'unité d'analyse étant une unité de temps dont la durée est de cinq secondes). Sur sa feuille, l'observateur écrit le chiffre 4 (chiffre correspondant à la catégorie 4 de l'instrument d'observation de Stallings et signifiant que l'élève est en situation d'auto-instruction — soit qu'il lise, écrive, se serve d'un objet ou d'un appareil —, cette activité étant cependant proprement académique et non verbale).

— La coopération entre pairs

La coopération est définie comme tout travail accompli en commun par deux élèves ou plus. L'observateur, à la manière du paragraphe précédent, inscrit une (1) fréquence pour chaque cinq secondes de présence de ce comportement en notant le chiffre 5. (Dans l'instrument d'observation de Stallings, le travail en commun accompli par deux élèves ou plus est codé selon le cas, 4 — instruction, 5 — remarque générale, 6 — remarque reliée à la tâche et 12 — observation, attention ; nous avons choisi le chiffre 5 pour regrouper les comportements caractérisant cette variable : un élève en instruit un autre (4) ; il fait des commentaires reliés à la tâche à d'autres élèves (6) ; les élèves

sont engagés dans des activités communes (4,5,6), ils écoutent, ils observent (12) en travaillant à deux ou à plusieurs.)

— L'absence de ces deux comportements (persistance au travail et coopération entre pairs)

Donc *l'instrument d'observation* comprend :

- a) la catégorie 4 (persistance au travail) ;
- b) la catégorie 5 (coopération entre pairs) ; et,
- c) la catégorie 0 (absence de ces deux comportements).

L'unité d'analyse

L'unité d'analyse est définie comme suit : les juges indiquent une fréquence pour l'un ou l'autre des comportements à toutes les cinq secondes d'observation. Ainsi, l'addition de ces fréquences au cours de chaque période de cinq minutes d'observation, et cela autant pour les catégories 4, 5 ou 0, constitue les données de l'analyse statistique (voir appendice).

LE SCHÈME D'OBSERVATION

Pour répondre aux exigences de l'analyse statistique, nous avons élaboré le schème d'observation que présente le tableau I. Ce schème d'observation est une adaptation du schème proposé par McGaw, Wardrop et Bunda⁶. Il a été élaboré en vue de l'étude de fiabilité dont nous allons maintenant traiter.

TABLEAU I
Schème d'observation⁷

Elèves	E ₁				E ₂				...	E _n			
	0 ₁₁	0 ₁₂	0 ₁₃	0 ₁₄	0 ₂₁	0 ₂₂	0 ₂₃	0 ₂₄	...	0 _{n1}	0 _{n2}	0 _{n3}	0 _{n4}
J ₁	x ₁₁₁	x ₁₂₁	x ₁₃₁	x ₁₄₁	x ₂₁₁	x ₂₂₁	x ₂₃₁	x ₂₄₁	...	x _{n11}	x _{n21}	x _{n31}	x _{n41}
J ₂	x ₁₁₂	x ₁₂₂	x ₁₃₂	x ₁₄₂	x ₂₁₂	x ₂₂₂	x ₂₃₂	x ₂₄₂	...	x _{n12}	x _{n22}	x _{n32}	x _{n42}

Légende : J₁, J₂ : juges ou observateurs
 E_i : i^{ème} élève 1 ≤ i ≤ n , n=35
 0_{ij} : j^{ème} observation du i^{ème} élève, 1 ≤ j ≤ 4
 x_{ijk} : fréquence de l'un des comportements, 4, 5 ou 0, inscrite par le k^{ème} juge à la j^{ème} observation du i^{ème} élève, 1 ≤ k ≤ 2

Qu'entend-on par étude de fiabilité ? Une série de données est dite « fiable » si la différence moyenne entre deux données d'observation du même objet (élève, enseignant, classe...) est plus petite que la différence moyenne entre deux données d'observation de deux objets distincts⁸. Par exemple, la trop grande variation intra-objet des comportements observés par rapport à la variation inter-objets est un élément déterminant de la non-fiabilité des données.

Historiquement les études de fiabilité ont d'abord été appliquées à l'analyse des résultats de tests. Des techniques telles que la *test-retest* ou le *split-half* ont été alors employées à cette fin. Or, comme le soulignent Frick et Semmel⁹, pour être utilisées adéquatement ces méthodes doivent satisfaire certaines hypothèses d'équivalence. Ceci est théoriquement possible lorsqu'on étudie la fiabilité des résultats de tests. Cependant, la situation est fort différente si on étudie la fiabilité des données obtenues à l'aide d'un instrument d'observation en classe. Car, dans ce cas, les hypothèses d'équivalence (des juges ou des observateurs) ne peuvent pas tenir. C'est pourquoi plusieurs chercheurs dont Medley et Mitzel, Cronbach, Gleser, Nanda et Rajaratnam, McGaw, Wardrop et Bunda, et enfin Frick et Semmel ont suggéré d'employer des coefficients de généralisabilité (que nous appelons ici coefficients de fiabilité), calculés à l'aide d'une ANOVA, pour mesurer la fiabilité des données d'observation recueillies en classe. Un coefficient de fiabilité n'est autre qu'un coefficient de corrélation intragroupe (*intra-class correlation coefficient*)¹⁰. Sa forme générale est donc :

$$\rho = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_e^2}$$

où σ_A^2 : variance attribuable au facteur A,
 σ_e^2 : erreur.

Dans le cas qui nous intéresse, nous étudierons plus particulièrement deux types de coefficients de corrélation intragroupe. Le premier, appelé coefficient d'entente inter-juges,

$$\rho_1 = \frac{\sigma_I^2}{\sigma_I^2 + \sigma_J^2}$$

où σ_J^2 : variance inter-juges,
 σ_I^2 : variance totale moins variance inter-juges,

qu'on retrouve entre autres dans Frick et Semmel¹¹, permettra de mesurer jusqu'à quel point la mésentente inter-juges limite la fiabilité des données obtenues à partir des comportements observés. La valeur minimale de ρ_1 est 0 (σ_J^2 beaucoup plus grand que σ_I^2) : cas où la mésentente inter-juges limite complètement la fiabilité des données. La valeur maximale de ρ_1 est 1 (σ_J^2 beaucoup plus petite que σ_I^2) : cas où la mésentente inter-juges ne limite pas du tout la fiabilité.

Le second, appelé coefficient de fiabilité,

$$\rho_2 = \frac{\sigma_E^2}{\sigma_E^2 + \sigma_e^2}$$

où σ_E^2 : variance inter-élèves,

σ_e^2 : erreur,

dû à McGaw, Wardrop et Bunda¹², permettra de mesurer la fiabilité avec laquelle le comportement des élèves peut être observé. La valeur minimale de ρ_2 est 0 (σ_e^2 beaucoup plus grande que σ_E^2) : cas où la fiabilité des données est nulle. La valeur maximale est 1 (σ_e^2 beaucoup plus petit que σ_E^2) : cas où la fiabilité des données est parfaite.

La technique statistique utilisée pour déterminer ces coefficients de fiabilité est l'analyse de la variance à trois dimensions ou facteurs (juges, élèves, occasions) : où l'on a deux interactions, élèves \times juges et occasions \times juges, et un facteur niché (occasions) sous 'élèves'. Dans le cas du coefficient ρ_2 , on définit alors :

$$\sigma_e^2 = \sigma_J^2 + \sigma_{0(E)}^2 + \sigma_{EJ}^2 + \sigma_{rés.}^2$$

où σ_J^2 : variance inter-juges,

$\sigma_{0(E)}^2$: variance due aux occasions (facteur niché sous 'élèves'),

σ_{EJ}^2 : variance due à l'interaction élèves \times juges,

$\sigma_{rés.}^2 = \sigma_{0(E)J}^2$: variance due à l'interaction occasions \times juges.

L'ANALYSE STATISTIQUE

Les résultats

Considérons les données brutes (voir l'appendice) présentées selon le schéma d'observation (tableau I). On se rend compte que, pour chaque juge (J_1 et J_2), chaque occasion (cinq minutes d'observation) donne lieu à un groupe de trois nombres. Ces trois nombres représentent les fréquences des comportements 4, 5 et 0 telles qu'enregistrées durant cette période d'observation de cinq minutes. Par exemple, pour l'étudiant 1, lors de l'occasion 1, le juge 1 a donné comme fréquences respectivement les nombres 45, 9 et 6 pour les catégories 4, 5 et 0. Comme les juges devaient indiquer une fréquence pour l'un ou l'autre des comportements à chaque cinq secondes d'observation, la fréquence totale pour cinq minutes est toujours 60.

À noter la très grande variabilité des données pour un même comportement, et ce, autant d'occasion en occasion pour un même élève que d'un élève à l'autre. On peut déjà juger de l'instabilité des comportements observés.

Nous reproduisons, dans les tableaux suivants, les résultats des trois analyses de variance (une pour chaque comportement) ainsi que les coefficients d'entente inter-juges et de fiabilité pour chaque analyse.

TABLEAU II

Résultats de l'analyse de variance du comportement 4

	Somme des carrés	Degrés de liberté	Carré moyen (CM)	Variance ¹³ (σ^2)
Juges (J)	6,006	1	6,006	0,0025
Élèves (E)	52379,250	34	1540,566	108,043
Occasions (O(E)) sous Élèves	71421,938	105	680,209	335,284
Élèves × Juges (EJ)	119,117	34	5,651	1,413
Occasions × Juges (O(E)J)	1012,364	105	9,642	9,642

TABLEAU III

Résultats de l'analyse de variance du comportement 5

	Somme des carrés	Degrés de liberté	Carré moyen (CM)	Variance (σ^2)
Juges (J)	20,086	1	20,086	0,087
Élèves (E)	20008,137	34	588,475	19,913
Occasions (O(E)) sous Élèves	44889,668	105	427,521	210,616
Élèves × Juges (EJ)	270,036	34	7,942	1,986
Occasions × Juges (O(E)J)	660,368	105	6,289	6,289

TABLEAU IV

Résultats de l'analyse de variance du comportement 0

	Somme des carrés	Degrés de liberté	Carré moyen (CM)	Variance (σ^2)
Juges (J)	0,356	1	0,356	0
Élèves (E)	78332,938	34	2303,910	216,835
Occasions O(E) sous Élèves	59892,246	105	570,402	280,880
Élèves × Juges (EJ)	254,135	34	7,475	1,869
Occasions × Juges (O(E)J)	907,491	105	8,643	8,643

TABLEAU V
Coefficients d'entente inter-juges et de fiabilité

Coefficients		ρ_1	ρ_2
Comportements			
4		0,999	0,238
5		0,999	0,083
0		0,999	0,427

Discussion

Comme le laissent voir les tableaux II, III et IV, la source de variance de loin la plus importante est due à la variation des occasions (intra-élève). Ceci n'est pas sans rappeler les propos de McGaw¹⁴ et Frick¹⁵ qui affirment que la plus grande source d'erreur de variance provient de l'instabilité de l'objet d'observation d'occasion en occasion.

Il faut noter, en outre, que la variance due aux juges est particulièrement petite par rapport aux autres sources de variance, ce qui est reflété, au tableau V, par un coefficient d'entente inter-juges (ρ^1) très élevé dans chaque cas. Ainsi on peut conclure que le faible degré de mésentente inter-juges ne limite en rien la fiabilité des données. Pourtant, les coefficients de fiabilité couplés à ces coefficients d'entente inter-juges sont très bas, démontrant, par le fait même, le faible degré de fiabilité des données. Cette apparente contradiction, entre les coefficients d'entente inter-juges très élevés et les coefficients de fiabilité très bas, est éclaircie par McGaw¹⁴ et Frick¹⁶ lorsqu'ils affirment que le coefficient d'entente inter-juges est le premier mais non le plus important coefficient à calculer pour mesurer la fiabilité des données. Cronbach¹⁷ mentionne même qu'il n'est pas impossible d'obtenir des données très peu fiables alors que les juges s'entendent à plus de 99%.

Qu'est-ce qui cause le si faible degré de fiabilité des données ? Pour répondre à cette question, nous devons expliciter la formulation du coefficient de fiabilité.

$$\text{Soit } \rho_2 = \frac{\sigma_E^2}{\sigma_E^2 + \underbrace{\sigma_J^2 + \sigma_{0(E)}^2 + \sigma_{EJ}^2 + \sigma_{0(E)J}^2}_{\sigma_e^2}}$$

Ce coefficient sera très petit si σ_e^2 est très grand par rapport à σ_E^2 , c'est-à-dire si l'une ou l'autre des quatre composantes de σ_e^2 est très grande par rapport à σ_E^2 . Or, un regard au tableaux II, III et IV nous montre que dans chacune des trois analyses, c'est presque exclusivement la composante $\sigma_{0(E)}^2$ qui rend σ_e^2 plus grand que σ_E^2 . Ainsi, il s'agit bien de l'instabilité des comportements (4, 5 ou 0) de l'élève d'occasion en occasion qui est

à la source de la non-fiabilité des données : plus précisément il s'agit de la trop grande variabilité des données intra-élève (d'occasion en occasion), *par rapport* à la variabilité des données inter-élèves. Pour illustrer l'importance de cette relation (intra vs inter) dans la détermination du coefficient ρ_2 , nous donnons au tableau VI un exemple de données fictives où la variabilité intra-élève est beaucoup moins sensible que la variabilité inter-élèves : ce qui donne un coefficient de fiabilité très élevé ($\rho_2 = 0,97$).

TABLEAU VI
Exemple des données fiables

Juges	Élèves	E ₁				E ₂				E ₃			
		0 ₁₁	0 ₁₂	0 ₁₃	0 ₁₄	0 ₂₁	0 ₂₂	0 ₂₃	0 ₂₄	0 ₃₁	0 ₃₂	0 ₃₃	0 ₃₄
J ₁		10	11	13	18	25	23	22	30	51	56	58	57
J ₂		12	18	16	15	28	30	25	26	53	60	52	51

CONCLUSION

Après avoir décrit l'instrument d'observation et le schéma d'observation utilisés pour recueillir les données, nous avons présenté les résultats de l'analyse statistique. Un coefficient d'entente inter-juges et un coefficient de fiabilité ont été calculés pour chacun des trois comportements observés. Le coefficient d'entente inter-juges (ρ_1) nous a indiqué dans quelle mesure la mésentente inter-juges a pu limiter la fiabilité des données. Le coefficient de fiabilité (ρ_2) a été calculé pour mesurer la fiabilité avec laquelle le comportement des élèves pouvait être observé.

Nous avons trouvé que les valeurs du coefficient d'entente inter-juges étaient très élevées alors que les valeurs du coefficient de fiabilité se sont montrées très faibles. Ce faible coefficient a été expliqué en grande partie par l'instabilité des comportements de l'élève d'occasion en occasion : c'est-à-dire par la trop grande variabilité des données intra-élève par rapport à la variabilité inter-élèves¹⁸.

Ne pouvant nous fier aux données, même si l'entente inter-juges est très élevée, nous devons conclure que les résultats de nos observations ne nous permettent pas d'évaluer les différences qu'il pourrait y avoir entre les élèves dans leurs comportements (persistance au travail, coopération entre pairs).

Comment interpréter de tels résultats ? Le problème étant l'instabilité des comportements d'une occasion à l'autre, il vient spontanément à l'idée qu'il faudrait définir dans le schème d'observation des situations d'apprentissage homogènes à observer à des occasions différentes. Par exemple, les situations suivantes : quand les élèves passent un examen, quand ils discutent en classe, quand ils travaillent en groupe, quand ils corrigent une dictée, quand ils écoutent les explications. Ainsi, la situation pourrait amener une stabilité dans les comportements observés. D'ailleurs certains auteurs, comme McGaw *et alii*, proposent de prévoir dans le schème d'observation des situations semblables ou homogènes qu'on pourrait observer à plusieurs occasions différentes (au cours d'un semestre ou d'une année scolaire).

Dans le cas de la persistance au travail, il faudrait définir cette variable de telle sorte qu'on puisse l'observer dans des situations homogènes. Par exemple, seraient persistants au travail non seulement l'élève qui travaille seul à un contenu académique, mais celui qui discute de ce contenu avec ses pairs ou avec l'enseignant et celui qui s'organise pour mettre en place tous les moyens dont il a besoin pour ce travail. La persistance couvrirait ainsi quatre activités distinctes reliées entre elles par un même focus. Cette nouvelle définition de la persistance au travail permettrait d'identifier plusieurs situations homogènes ou contextes d'apprentissage donnés regroupant différentes activités d'enseignement.

Si, pour la persistance au travail, il est facile de concevoir des situations d'apprentissage en classe regroupées autour d'activités que l'on peut décrire, il n'en est pas de même pour la coopération entre pairs. Nous avons, en effet, affaire à deux variables fort différentes. La persistance au travail a beaucoup plus d'importance en classe, semble-t-il, que la coopération. Elle est reliée au succès académique : la recherche en éducation répète que le temps consacré à l'étude d'un contenu académique est relié significativement au succès scolaire. La coopération entre pairs est une variable dont il est beaucoup plus difficile de justifier l'importance en classe. En soi, les activités de la classe sont centrées autour des matières académiques. La coopération entre pairs est sans doute une attitude sociale très importante à acquérir, mais elle ne se prête pas à une attention continue et presque unique comme toute matière académique : l'enseignant de français enseigne le français et, à l'occasion, il pourra peut-être insister sur la coopération. Celle-ci, qu'on pourrait appeler secondaire dans le milieu 'classe', est difficile à définir en fonction de situations d'apprentissage, au même titre que la persistance au travail. Les situations de coopération en classe ne sont pas nécessaires comme les situations de persistance au travail et, par conséquent, ne sont pas aussi nombreuses et probablement très peu fréquentes comme situations. C'est pourquoi il est sans doute illusoire de les observer dans le but de trouver des différences significatives entre certains groupes ou certaines méthodes d'enseignement.

Ainsi, des deux variables retenues, il n'y a, semble-t-il que la persistance au travail qui puisse offrir la possibilité de résultats fiables. Pour ce faire, la définition de la persistance au travail devra être faite dans un contexte de situations d'apprentissage.

Donc, peut-on se fier aux données obtenues à l'aide d'un instrument d'observation ? Oui, dans la mesure où l'on tient compte de l'importance de situations homogènes d'apprentissage dans l'observation des comportements des élèves en classe, sans oublier les caractères fondamentaux de stabilité et de pertinence des variables à observer.

NOTES ET RÉFÉRENCES

1. Nous traduisons *reliability* par fiabilité.
2. Comme l'indiquent nos références (Medley et Mitzel, 1963 ; Cronbach, 1972 ; McGaw *et alii*, 1972), nous appliquons la théorie telle que proposée dans le domaine de l'utilisation des instruments d'observation en classe.
3. Système d'apprentissage géré par l'étudiant.
4. *Classroom Observation Instrument et Training Manual for Classroom Observation, Follow — Through Evaluation*, Stanford Research Institute, Menlo Park, California 94025, février 1973.
5. *Un instrument d'observation de la classe pour l'élémentaire*, Mariel Leclerc, Jocelyne Roberge-Brassard et Claire Turcotte, Collection Devenir, no 2, 1977, INRS-Éducation. Et *Revue des sciences de l'éducation*, « La mesure des comportements dans un système d'enseignement individualisé », Mariel Leclerc *et alii* pp. 191-205, vol. III, no 2, printemps 1977.
6. McGaw, B., Wardrop, J.L., Bunda, M.A., « Classroom observation schemes : Where are the errors ? » *American educational research journal*, vol. 9, no 1, hiver 1972, p. 23.
7. Ainsi deux juges ont observé, conjointement, trente-cinq élèves à quatre occasions différentes (étant donné les contraintes matérielles, nous n'avons pu observer ces trente-cinq élèves qu'à quatre occasions différentes). Ces élèves ont été choisis comme suit. Il y avait en tout cinq classes expérimentales (classes appliquant SAGE) et deux classes témoins. C'est parmi ces classes qu'on a choisi au hasard cinq élèves par classe. Au total on a effectué 700 minutes (11 heures et 40 minutes) d'observation.
8. Medley, D.M., Mitzel, H.E., « Measuring classroom behavior by systematic observation » in *Handbook of research on teaching*, Chicago : Rand McNally, 1963, p. 250.
9. Frick, T., Semmel, M.I., « Observer agreement and reliabilities of classroom observational measures », *Review of educational research*, hiver 1978, vol. 48, no 1, p. 158.
10. Hays, W.L., *Statistics for the social sciences*, second edition, Holt Rinehart and Winston Inc., 1972, p. 535.
11. Frick et Semmel, *idem*, p. 164.
12. McGaw *et alii*, *idem*, p. 24.

$$\begin{aligned}
 13. \quad \sigma_j^2 &= \frac{1}{140} (CM_J - CM_{EJ}) \\
 \sigma_E^2 &= \frac{1}{8} (CM_E - CM_{0(E)} - CM_{EJ} + CM_{0(E)J}) \\
 \sigma_{0(E)}^2 &= \frac{1}{2} (CM_{0(E)} - CM_{0(E)J}) \\
 \sigma_{EJ}^2 &= \frac{1}{4} (CM_{EJ}) \\
 \sigma_{0(E)J}^2 &= CM_{0(E)J}
 \end{aligned}$$

14. McGaw *et alii*, *idem*, p. 16.
15. Frick et Semmel, *idem*, p. 159.
16. Frick et Semmel, *idem*, p. 158.
17. Cronbach, L.J., Gleser, G.C., Nanda, H., Rajaratnam, N. *The dependability of behavioral measurements : Theory of generalizability for scores and profiles*. New York : Wiley, 1972, p. 190.
18. Voir Calkins, D., Borich, G.D., Pascone, M., Kugle, C.L., Marston, D.T., « Generalizability of teacher behaviors across classroom observation systems », *The Journal of classroom interaction*, vol. 13, no 1, 1977, p. 9.

De même, pour ses auteurs, les faibles coefficients de fiabilité trouvés ne sont pas dus au manque d'entente inter-juges mais bien à la trop grande variation des comportements des objets d'observation.

APPENDICE

Données brutes recueillies selon le schème d'observation

Juges	E ₁				E ₂			
	O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁	45 9 6	21 9 30	0 12 48	29 6 25	31 16 13	33 4 23	1 11 48	16 18 26
J ₂	45 11 4	21 8 31	0 15 45	27 7 26	42 18 0	32 6 22	0 19 41	5 23 32
	E ₃				E ₄			
	O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁	32 1 27	30 3 27	36 4 20	53 0 7	55 0 5	47 3 10	55 5 0	35 4 21
J ₂	33 3 24	39 1 20	36 5 19	59 0 1	56 0 4	54 3 3	54 0 6	31 7 22
	E ₅				E ₆			
	O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁	51 0 9	3 9 48	50 8 2	23 24 13	36 0 24	24 35 1	31 14 15	20 20 20
J ₂	50 0 10	14 9 37	35 23 2	13 42 5	33 0 27	20 36 4	27 19 14	25 18 17
	E ₇				E ₈			
	O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁	24 6 30	0 60 0	28 0 32	0 21 39	36 1 23	0 15 45	28 32 0	1 14 45
J ₂	24 7 29	0 60 0	29 0 31	0 21 39	39 0 21	0 13 47	30 30 0	2 15 43
	E ₉				E ₁₀			
	O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁	31 26 3	11 14 35	6 12 42	35 18 7	3 14 43	22 37 1	60 0 0	28 3 29
J ₂	28 29 3	11 12 37	5 11 44	33 22 5	3 7 50	25 34 1	60 0 0	28 2 30
	E ₁₁				E ₁₂			
	O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁	43 10 7	43 12 5	9 15 36	36 18 6	31 9 20	57 0 3	15 42 3	55 0 5
J ₂	41 19 0	42 4 14	16 18 26	34 20 6	36 2 22	55 0 5	6 53 1	58 1 1

		E ₁₃				E ₁₄			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		3 0 57	46 3 11	0 0 60	0 0 60	33 15 12	22 4 34	56 3 1	0 10 50
J ₂		7 0 59	48 0 12	0 0 60	0 0 60	32 17 11	27 3 30	53 5 2	0 8 52
		E ₁₅				E ₁₆			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		45 8 7	49 3 8	36 17 7	24 28 8	43 3 14	7 43 10	0 40 20	13 47 0
J ₂		51 5 4	57 2 1	30 20 10	12 42 6	45 6 9	5 45 10	0 35 25	0 60 0
		E ₁₇				E ₁₈			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		41 9 10	1 6 53	43 6 11	5 50 5	60 0 0	51 9 0	4 0 56	28 5 27
J ₂		43 9 8	4 4 52	42 6 12	4 51 5	60 0 0	42 18 0	6 0 54	27 4 29
		E ₁₉				E ₂₀			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		0 0 60	59 0 1	0 52 8	41 8 11	49 0 11	60 0 0	48 0 12	49 1 10
J ₂		0 0 60	60 0 0	0 53 7	39 9 12	55 0 5	55 0 5	54 0 6	51 1 8
		E ₂₁				E ₂₂			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		21 22 17	59 1 0	26 13 21	49 0 11	0 60 0	52 1 7	30 14 16	51 8 1
J ₂		24 18 18	58 1 1	24 8 28	52 0 8	0 60 0	52 1 7	21 17 22	46 14 0
		E ₂₃				E ₂₄			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		51 9 0	39 14 7	2 47 11	56 4 0	22 10 28	57 0 3	49 0 11	57 0 3
J ₂		50 4 6	40 16 4	0 46 14	52 5 3	13 6 41	58 0 2	44 2 14	60 0 0

		E ₂₅				E ₂₆			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		57 0 3	1 59 0	52 3 5	36 14 10	0 0 60	12 0 48	9 0 51	0 0 60
J ₂		60 0 0	0 58 2	52 3 5	35 7 18	0 0 60	7 0 53	12 0 48	3 0 57
		E ₂₇				E ₂₈			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		0 0 60	0 0 60	5 0 55	0 0 60	0 0 60	16 0 44	0 0 60	0 0 60
J ₂		0 0 60	0 0 60	10 0 50	0 0 60	0 0 60	14 0 46	0 0 60	0 0 60
		E ₂₉				E ₃₀			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		0 0 60	48 0 12	0 0 60	0 0 60	0 0 60	54 0 6	0 0 60	0 0 60
J ₂		0 0 60	47 0 13	0 0 60	0 0 60	0 0 60	55 0 5	4 0 56	0 0 60
		E ₃₁				E ₃₂			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		15 1 44	0 0 60	0 0 60	3 39 18	33 0 27	0 0 60	0 0 60	0 60 0
J ₂		9 0 51	0 0 60	0 0 60	0 35 25	25 0 35	0 0 60	0 0 60	0 60 0
		E ₃₃				E ₃₄			
		O ₁	O ₂	O ₃	O ₄	O ₁	O ₂	O ₃	O ₄
J ₁		21 0 39	0 0 60	23 0 37	0 32 28	8 0 52	0 0 60	31 0 29	0 0 60
J ₂		19 0 41	7 0 53	26 0 34	0 33 27	7 0 53	0 0 60	32 0 28	0 0 60
		E ₃₅							
		O ₁	O ₂	O ₃	O ₄				
J ₁		0 0 60	0 0 60	39 0 21	3 0 57				
J ₂		0 0 60	0 0 60	39 0 21	7 0 53				