

# Les qualités psychométriques de l'évaluation de l'enseignement par les étudiantes et les étudiants : impacts des modalités pédagogiques et de passation

Léon Harvey, leon\_harvey@uqar.ca, Université du Québec à Rimouski

Marie-Hélène Hébert, marie-helene\_hebert@uqar.ca, Université du Québec à Rimouski

---

**Résumé :** Les évaluations de l'enseignement par les étudiants se déroulent sous différentes conditions pédagogiques et de passation. Le but de l'étude est de vérifier l'impact de la modalité de passation (*en ligne* et *papier*) sur les qualités psychométriques d'un questionnaire d'évaluation de l'enseignement lorsque celui-ci est adapté pour les *cours réguliers*, les *cours médiatisés*, les *laboratoires*, les *stages* et les *cours en tutorat*. Le corpus de données est constitué des évaluations réalisées durant les années 2007-2008 (format *papier*) et 2010-2011 (format *en ligne*). Quelle que soit la « modalité pédagogique » considérée, les résultats attestent que la modalité de passation a une influence sur les taux de réponses et, dans certains cas, sur les taux de satisfaction de l'enseignement, mais qu'elle n'influence ni la fiabilité ni la dimensionnalité des évaluations. Les implications de ces résultats sont discutées.

---

**Mots-clés :** évaluation de l'enseignement, modalités de passation et pédagogiques, fiabilité, dimensionnalité

## 1. Contexte

L'évaluation de l'enseignement par les étudiants (EEE) désigne « ...les procédures par lesquelles les étudiants prennent part au jugement porté sur des cours ou des enseignements spécifiques, attribués à des enseignants ou à des équipes d'enseignants au sein d'un programme ou d'une formation » (Romainville & Coggi, 2009, p. 13). Actuellement, la grande majorité des universités aux États-Unis (Dommeyer, Baum, Hanna, & Chapman, 2004), au Canada et en Europe évaluent régulièrement la qualité de l'enseignement à partir de l'opinion des étudiants inscrits aux activités d'enseignement. Les résultats de ces évaluations sont utilisés pour des fins académiques, mais sont également évoqués lors de décisions liées à la carrière professorale (Donovan, Mader, & Shinsky, 2006 ; Fontaine, 2009 ; Younès, 2009). Ils sont, par conséquent, une source de controverses dans les milieux universitaires (Dommeyer et al., 2004 ; Younès & Romainville, 2012). Trois aspects parmi les plus controversés concernent le construit qui est effectivement mesuré par ces évaluations, la baisse du taux de réponses lorsque l'évaluation est réalisée *en ligne* plutôt qu'en classe en utilisant un format *papier*, et l'impact des différentes modalités pédagogiques sur ces évaluations.

## 2. Objectif

L'étude consiste à vérifier l'impact de la modalité de passation (*en ligne* et *papier*) sur les qualités psychométriques d'un questionnaire d'évaluation de l'enseignement lorsque celui-ci est adapté pour les *cours réguliers*, les *cours médiatisés*, les *laboratoires*, les *stages* et les *cours en tutorat* (appelés ci-après « modalités pédagogiques »). Le cas des *cours réguliers* a fait l'objet d'une publication antérieure (Harvey & Hébert, 2012). Quatre hypothèses de recherche découlent de cet objectif :

*Hypothèse 1 - Taux de réponses.* La modalité de passation aura un effet sur les taux de réponses pour l'ensemble des modalités pédagogiques.

*Hypothèse 2 - Taux de satisfaction.* La modalité de passation n'aura pas d'effet sur les taux de satisfaction de la qualité de l'enseignement pour l'ensemble des modalités pédagogiques.

*Hypothèse 3 - Fiabilité.* La modalité de passation n'aura pas d'effet sur la fiabilité du questionnaire d'évaluation de l'enseignement pour l'ensemble des modalités pédagogiques.

*Hypothèse 4 - Dimensionnalité.* La modalité de passation n'aura pas d'effet sur la dimensionnalité du questionnaire d'évaluation de l'enseignement pour l'ensemble des modalités pédagogiques.

### 3. Méthodologie

Dans les lignes qui suivent sont explicités le questionnaire et les échantillons considérés, ainsi que le plan d'analyses de l'étude.

#### 3.1 Questionnaire

Le questionnaire qui est d'usage à l'Université du Québec à Rimouski (UQAR) pour évaluer la qualité des enseignements se décline en cinq versions. Il existe des versions pour les *cours réguliers* (26 items), les *cours médiatisés* (14 items), les *laboratoires* (34 items), les *stages* (25 items) et les *cours en tutorat* (19 items). En général, à travers chacune des versions, six dimensions sont évaluées au départ d'un certain nombre d'items. Elles se déclinent comme suit pour le questionnaire consacré aux *cours réguliers* : *contexte du cours* (5 items), *organisation et clarté* (8 items), *dynamisme, intérêt et habiletés d'enseignement* (4 items), *interaction avec les étudiants et étudiantes* (3 items), *évaluation et rétroaction* (4 items) et *appréciation générale* (2 items). Pour les autres versions, les items nichés sous chacune des dimensions sont adaptés aux spécificités de la modalité pédagogique. Certaines versions peuvent contenir des dimensions supplémentaires ou bien en omettre. En effet, pour la version *laboratoire*, une dimension *travaux pratiques* (7 items) est ajoutée, tandis que pour les *cours médiatisés* et *en tutorat*, la dimension *dynamisme, intérêt et habiletés d'enseignement* est omise.

Pour chacune des versions du questionnaire, l'étudiant sondé doit marquer son degré d'accord (ou de désaccord) aux items au moyen d'une échelle de type Likert à cinq modalités, soit 0 (ne s'applique pas), 1 (tout à fait en désaccord), 2 (plutôt en désaccord), 3 (plutôt d'accord) et 4 (tout à fait d'accord). Les réponses ne s'appliquant pas (0) sont considérées dans le cadre de l'étude comme des données manquantes et sont en conséquence exclues des analyses.

Les cinq versions du questionnaire ont été administrées aux étudiants dans un format *papier* jusqu'en 2007-2008. Depuis, la modalité de passation *en ligne* est privilégiée.

#### 3.2 Échantillons

Le corpus de données provient de huit départements et de deux campus de l'UQAR, situés à Rimouski et Lévis. Au total, 42 960 étudiants ont contribué à l'évaluation de l'enseignement pour les années 2007-2008 (format *papier*) et 2010-2011 (format *en ligne*). La répartition du nombre de classes et de répondants pour chacune des modalités pédagogiques et de passation est montrée au tableau 1.

**Tableau 1** : nombre de classes et de répondants pour chacune des modalités pédagogiques et de passation

Modalité pédagogique	Modalité de passation	Nombre de classes	Nombre de répondants	Année
Cours régulier	En ligne	1 094	16 432	2010-2011
	Papier	868	20 245	2007-2008
Cours médiatisé	En ligne	135	1 011	2010-2011
	Papier	31	420	2007-2008
Stage	En ligne	146	814	2010-2011
	Papier	132	1 222	2007-2008
Laboratoire	En ligne	86	781	2010-2011
	Papier	43	1 702	2007-2008
Cours en tutorat	En ligne	105	261	2010-2011
	Papier	30	72	2007-2008
Total		2 670	42 960	

### **3.3 Plan d'analyses**

*Dimensionnalité.* Des analyses factorielles confirmatoires sont réalisées sur chacune des versions du questionnaire (*en ligne* et *papier*) afin d'évaluer sa dimensionnalité. Les items ayant trait à la dimension *contexte du cours* ne sont pas considérés dans les analyses parce qu'elle n'est pas considérée dans le calcul du résultat associé à la ressource professorale. Pour la version *cours régulier* du questionnaire, quatre modèles ( $M_1$ ,  $M_5$ ,  $M_{51s}$ ,  $M_{51h}$ ) sont pris en considération. Les modèles  $M_1$  et  $M_5$  permettent de déterminer si la structure de la version du questionnaire est unidimensionnelle ( $M_1$ ) ou multidimensionnelle ( $M_5$ ). Dans le cas où elle est multidimensionnelle, les modèles ultérieurs précisent s'il existe un facteur de second ordre ( $M_{51s}$ ), ou un facteur hiérarchique ( $M_{51h}$ ). Ces modèles s'apparentent à ceux développés par Brunner, Nagy et Wilhelm (2012) pour modéliser l'intelligence. Pour les autres versions du questionnaire, seul le modèle qui a le mieux décrit la version *cours régulier* est mis à l'épreuve. Pour une présentation détaillée des modèles testés dans le contexte de l'étude, voir Harvey & Hébert (2012).

*Fiabilité et autres analyses.* L'indice de fiabilité considéré dans le cadre de l'étude est l'alpha de Cronbach. Il est calculé pour l'ensemble des items et par dimension pour chacune des versions du questionnaire (*en ligne* et *papier*). Les taux de réponses sont calculés, quant à eux, pour chacune des versions du questionnaire. Un test du  $X^2$  est utilisé afin de tester la différence entre les modalités *en ligne* et *papier*. Enfin, pour chacune des versions du questionnaire, les taux de satisfaction de l'enseignement sont calculés par item, par dimension et pour l'ensemble des items. Des tests t (avec correction de Bonferroni) sont calculés pour détecter la présence de différences significatives entre les modalités *en ligne* et *papier*.

## **4. Résultats**

Les principaux résultats de l'étude sont rapportés dans les lignes qui suivent. Ils se présentent en quatre sections distinctes : les taux de réponses, les taux de satisfaction, ainsi que la fiabilité et la dimensionnalité.

### **4.1 Hypothèse 1 - Taux de réponses**

Pour la version *cours régulier* du questionnaire, la modalité *papier* a engendré une plus grande participation chez les étudiants que la modalité *en ligne*. À ce chapitre, Harvey et Hébert (2012, voir tableau 2) rapportent une différence significative de 26,3 % entre les deux modalités à la faveur de la modalité *papier*. Des taux de réponses de 84,9 % et 58,6 % sont respectivement observés pour les modalités *papier* et *en ligne*. La situation est similaire pour les autres versions du questionnaire. Des différences significatives variant de 16,4 % (*cours en tutorat*) à 27,0 % (*stage*) sont enregistrées.

### **4.2 Hypothèse 2 - Taux de satisfaction**

Lorsqu'on considère la version *cours régulier* du questionnaire, les taux de satisfaction de l'enseignement varient très peu. Les taux moyens sont respectivement de 90,3 % et 88,0 % pour les modalités *papier* et *en ligne*. Lorsqu'il s'agit de considérer les différences entre les modalités *en ligne* et *papier* pour les six dimensions de la version et ses 26 items, le portrait demeure le même (voir Harvey & Hébert, 2012, tableaux 3 et 4). Des écarts variant de 2,3 % à 3,0 % pour les dimensions et de 1,5 % à 4,3 % pour les items sont enregistrés à la faveur de la modalité *papier*. Malgré qu'elles soient significatives, ces différences sont considérées comme marginales étant donné la grande taille de l'échantillon d'étudiants considéré pour la version *cours régulier*.

Pour les autres versions du questionnaire dans leur entier, des différences minimales sont également enregistrées en faveur de la modalité *papier*. Elles varient de 1,0 % (*stage*) à 3,8 % (*cours en tutorat*), mais seule la différence associée à la version *cours médiatisé* (3,0 %) est significative. Du côté des dimensions et des items, des écarts significatifs sont aussi à signaler en faveur de la modalité *papier*. Ils concernent les modalités pédagogiques, dimensions et items suivants :

- pour les dimensions :
  - *cours médiatisé* : organisation et clarté (écart de 3,5 %) – interaction avec les étudiants et étudiantes (écart de 3,8 %) – appréciation générale (écart de 3,0 %) ;
  - *laboratoire* : contexte du cours (écart de 3,0 %) – dynamisme, intérêt et habiletés d'enseignement (écart de 3,3 %) – appréciation générale (écart de 3,5 %) ;
  - *stage* : contexte du cours (écart de 2,0 %) – organisation et clarté (écart de 1,8 %) – dynamisme, intérêt et habiletés d'enseignement (écart de 2,0 %) – évaluation et rétroaction (écart de 2,0 %).
- pour les items :
  - *cours médiatisé* : 5 items sur 14 (écart de 3,3 à 4,5 %) ;
  - *laboratoire* : 9 items sur 34 (écart de 2,5 à 7,0 %) ;
  - *stage* : 8 items sur 25 (écart de 1,8 à 4,3 %) ;
  - *cours en tutorat* : 2 items sur 19 (écart de 8,0 et 10,0 %).

### 4.3 Hypothèse 3 - Fiabilité

La version *cours régulier* du questionnaire dans son entier montre des mesures de fiabilité plus qu'acceptables lorsque celle-ci est estimée par l'alpha de Cronbach (*papier* : 0,95 ; *en ligne* : 0,96). À l'exception de la dimension *contexte du cours*, la valeur calculée pour chacune des dimensions est supérieure au seuil de 0,80 habituellement souhaité (voir Harvey & Hébert, 2012, tableau 5). L'écart observé entre les fiabilités des modalités *papier* et *en ligne* pour l'ensemble des dimensions est très faible, variant de 0,01 à 0,05 à la faveur de la modalité *en ligne*. Le portrait demeure le même lorsqu'il s'agit de considérer les autres versions du questionnaire. Des coefficients voisins ou supérieurs à 0,80 sont enregistrés (sauf pour la dimension *contexte du cours*), et de très faibles écarts sont notés entre les modalités *papier* et *en ligne*.

### 4.4 Hypothèse 4 - Dimensionnalité

Pour la version *cours régulier* du questionnaire, les analyses confirmatoires suggèrent que le modèle hiérarchique est celui qui décrit le mieux l'ensemble des données, qu'elle soit administrée en format *papier* ou *en ligne* (voir Harvey & Hébert, 2012, tableaux 6 à 9). Les indices d'adéquation pour ce modèle sont de 0,94 (*papier*) et 0,93 (*en ligne*) pour l'indice général d'adéquation (GFI) et de 0,059 (*papier*) et 0,062 (*en ligne*) pour l'erreur quadratique moyenne (RMSEA). Les saturations associées au facteur hiérarchique sont élevées et varient entre 0,53 et 0,85 pour la modalité *papier* et entre 0,51 et 0,88 pour celle *en ligne*. En contrepartie, celles associées aux autres dimensions sont faibles. Elles varient entre 0,00 et 0,53 pour la modalité *papier* et entre 0,00 et 0,56 pour la modalité *en ligne*.

Le modèle hiérarchique décrit également bien la plupart des autres versions du questionnaire, et ce, que l'administration soit faite *en ligne* ou *papier*. Les seules exceptions concernent les versions *stage* et *laboratoire* (*en ligne*) et *cours en tutorat* (*en ligne* et *papier*), où les indices GFI sont inférieurs à 0,90. Les tableaux 2 et 3 rapportent les indices d'adéquation de ce modèle appliqué aux autres modalités pédagogiques sous les formats *en ligne* et *papier*.

**Tableau 2** : indices d'adéquation pour les cinq modalités pédagogiques (*en ligne*)

Modalité	Paramètres	GFI	RMSEA	X <sup>2</sup>	dl	p
Cours régulier	63	0,93	0,062	9576	170	0,0001
Cours médiatisé	36	0,97	0,058	171	44	0,0001
Laboratoire	57	0,89	0,078	642	134	0,0001
Stage	69	0,89	0,073	873	209	0,0001
Cours en tutorat	54	0,84	0,088	262	122	0,0001

**Tableau 3** : indices d'adéquation pour les cinq modalités pédagogiques (*papier*)

Modalité	Paramètres	GFI	RMSEA	X <sup>2</sup>	dl	p
Cours régulier	63	0,94	0,059	10652	170	0,0001
Cours médiatisé	36	0,95	0,063	99	43	0,0001
Laboratoire	57	0,94	0,061	844	135	0,0001
Stage	69	0,90	0,068	1144	208	0,0001
Cours en tutorat	54	0,66	0,170	323	120	0,0001

## 5. Discussion et conclusion

Dans le cadre de l'étude, quatre hypothèses sont soulevées. La première qui suggère que les taux de réponses sont affectés par la modalité de passation est confirmée. La différence à la faveur de la modalité *papier* se situe entre 16,4 % et 27,0 % en fonction de la modalité pédagogique considérée. Ce résultat est conforme à ce qui est observé dans la littérature. La seconde hypothèse a trait aux taux de satisfaction de l'enseignement. Des effets qui varient entre 1,0 % et 3,8 % sont observés en faveur de la modalité *papier* lorsqu'on considère les versions du questionnaire dans leur entier. Si certaines de ces différences sont significatives, elles doivent être considérées comme marginales considérant la grande taille de l'échantillon de répondants (effets de grandeur négligeables).

Les hypothèses 3 et 4 considèrent respectivement la fiabilité et la dimensionnalité des versions du questionnaire. Dans la mesure où un modèle hiérarchique décrit bien la plupart des versions (*en ligne* et *papier*), il est possible de conclure que la modalité de passation n'a pas d'effet sur la dimensionnalité, ce qui confirme l'hypothèse 4. Quant à l'hypothèse 3, elle est aussi confirmée puisque les indicateurs de fiabilité (alpha de Cronbach) des modalités *en ligne* et *papier* sont quasiment identiques, et ce, pour l'ensemble des modalités pédagogiques. Par contre, Harvey et Hébert (2012) soulignent que la présence d'une dimension hiérarchique a un impact considérable sur la fiabilité des dimensions primaires. En effet, pour la version *cours régulier* (format *en ligne*), les coefficients *oméga* pour chacune des dimensions primaires varient entre 0,04 et 0,29. La faible valeur des coefficients suggère que ces dimensions ne sont pas fiables lorsque l'apport de la composante hiérarchique est pris en compte. Cette dernière est en revanche très fiable avec un coefficient *oméga* de 0,95.

À l'instar de Harvey et Hébert (2012), l'étude suggère que les décisions prises à partir des questionnaires d'évaluation de l'enseignement par les étudiants sont valides et fiables lorsqu'elles sont basées sur le résultat global au questionnaire. En contrepartie, considérant qu'un modèle hiérarchique est ressorti des analyses, les décisions basées sur les résultats aux dimensions spécifiques (de premier ordre) devraient être évitées. La dimension hiérarchique est généralement interprétée comme étant une compétence générale en enseignement et est parfois associée au professionnalisme de la ressource professorale (Spooren & Mortelmans, 2006).

## 6. Références

- Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality, 4*, 796-846.
- Dommeyer, C. J., Baum, P., Hanna, R. W., & Chapman, K. S. (2004). Gathering faculty teaching evaluations by in-class and on-line surveys: Their effects on response rates and evaluations. *Assessment & Evaluation in Higher Education, 29*(5), 611-623.
- Donovan, J., Mader, C. E., & Shinsky, J. (2006). Constructive student feedback: On-line vs. traditional course evaluations. *Journal of Interactive Online Learning, 5*(3), 283-296.
- Fontaine, S. (2009). Des expériences actuelles d'évaluation des enseignements vers des démarches adaptées aux 2<sup>e</sup> et 3<sup>e</sup> cycles. Dans Marc Romainville & Cristina Coggi (Eds.), *L'évaluation de l'enseignement par les étudiants : approches critiques et pratiques innovantes* (pp. 123-144). Bruxelles : De Boeck.
- Harvey, L., & Hébert, M.-H. (2012). Évaluation de la qualité de l'enseignement par les étudiantes et étudiants : qualités psychométriques et comparaison des conditions de passation. *Mesure et évaluation en éducation, 35*(3), 31-60.
- Romainville, M., & Coggi, C. (Eds.) (2009). *L'évaluation de l'enseignement par les étudiants : approches critiques et pratiques innovantes*. Bruxelles : De Boeck.
- Spooren, P., & Mortelmans, D. (2006). Teacher professionalism and student evaluation of teaching: Will better teachers receive higher ratings and will better students give higher ratings? *Educational Studies, 32*(2), 201-214.
- Younès, N. (2009). L'évaluation de l'enseignement par les étudiants comme seuil de changement. Dans Marc Romainville & Cristina Coggi (Eds.), *L'évaluation de l'enseignement par les étudiants : approches critiques et pratiques innovantes* (pp. 191-210). Bruxelles : De Boeck.
- Younès, N., & Romainville, M. (2012). Les transformations actuelles de l'EEE. *Mesure et évaluation en éducation, 35*(3), 175-199.