

(Essai) L'évaluation des groupes sous contraintes

Table des matières

Essai	3
1. Introduction.....	3
2. Évaluer pour appréhender le Réel ?	3
3. Une enquête qui a tout changé :	5
4. Généraliser les conditions d'expérience de la première enquête	8
5. Le cadre des évaluations : dimension objective et dimension relationnelle.....	8
6. L'analyse qui précède la construction des questionnaires	9
7. La forme des questions utiles	9
8. La nature des indicateurs générés par des questions	10
9. Panel de données : force de l'évaluation externe	10
10. Panel de données : les qualités particulières des indicateurs relationnels	11
11. Panel de données : créer des plateformes d'évaluations fiables	14
12. Absence de panel de données : l'exploitation des évaluations internes.....	15
13. Précision sur le rôle des normes dans les évaluations	16
14. Les évaluations des individus basées sur des notations	16
15. Les différents types de groupes concernés par l'évaluation du relationnel	16
16. Les différents champs de compétences nécessaire aux évaluations	17
17. Réussir la mise en place pratique des évaluations de terrain	18
18. L'analyse des données recueillies à l'aide des questionnaires	18
19. La diffusion des résultats et l'après évaluation.....	19
Annexes.....	20
1. Annexe 1 : Les différentes formes d'évaluation du relationnel	20
2. Annexe 2 : Évaluation interne et évaluation externe : un peu d'histoire	20
3. Annexe 3 : Un cas « d'école » : la notation une auto-évaluation particulière	21
4. Annexe 4 : L'évaluation de l'individu et le principe de Peter	21
5. Annexe 5 : L'évaluation à la rescousse de la pédagogie des transmissions des savoirs	23
6. Annexe 6 : Quelques mathématiques qui justifient le traitement des données	24

Essai

1. Introduction

Dans le domaine du management des institutions, l'évaluation a longtemps été limitée à la notation des individus (élèves, fonctionnaires, militaires, ouvriers, etc.).

L'école de la République a poussé cette utilisation à l'extrême en utilisant ce type d'évaluation (notation + moyenne + classement) pour mettre en place une méthode d'autoévaluation de l'ensemble du groupe élèves.

Ces pratiques scolaires ont tellement imprégné la psychologie des élites de 19^{ème} et du 20^{ème} siècle qu'elles ont laissé croire, à la grande majorité des acteurs institutionnels, que l'évaluation ne pouvait concerner que celle des individus.

Pourtant une autre évaluation a été mise en place depuis longtemps, celle du bon fonctionnement des institutions à travers le regard des inspecteurs d'État.

Depuis l'apparition de l'informatique et plus récemment de « l'intelligence artificielle » de nouvelles pratiques sont possibles.

Mais ces nouvelles pratiques sont-elles légitimes et les résultats sont-ils fiables ?

C'est à cette question que nous allons répondre dans cet « Essai » en nous basant sur notre expérience (plus de 40 évaluations d'institutions) et notre travail de scientifique que nous avons poussé jusqu'à la mathématisation des données récoltées.

2. [Évaluer pour appréhender le Réel ?](#)

Pour appréhender le Réel nous disposons de l'information apportée par nos sens et du traitement de cette information par notre cerveau. Sans ces mécanismes nous sommes incapables de vivre. Le handicap d'un seul de nos sens (vue, audition, toucher, etc.) nous perturbe gravement. Les pathologies comme la maladie d'Alzheimer sont une perte de la relation que nous entretenons avec le Réel. En ce sens nous n'avons pas accès au Réel mais en avons simplement l'image que nous nous en forgeons. Les opinions font partie de ces images.

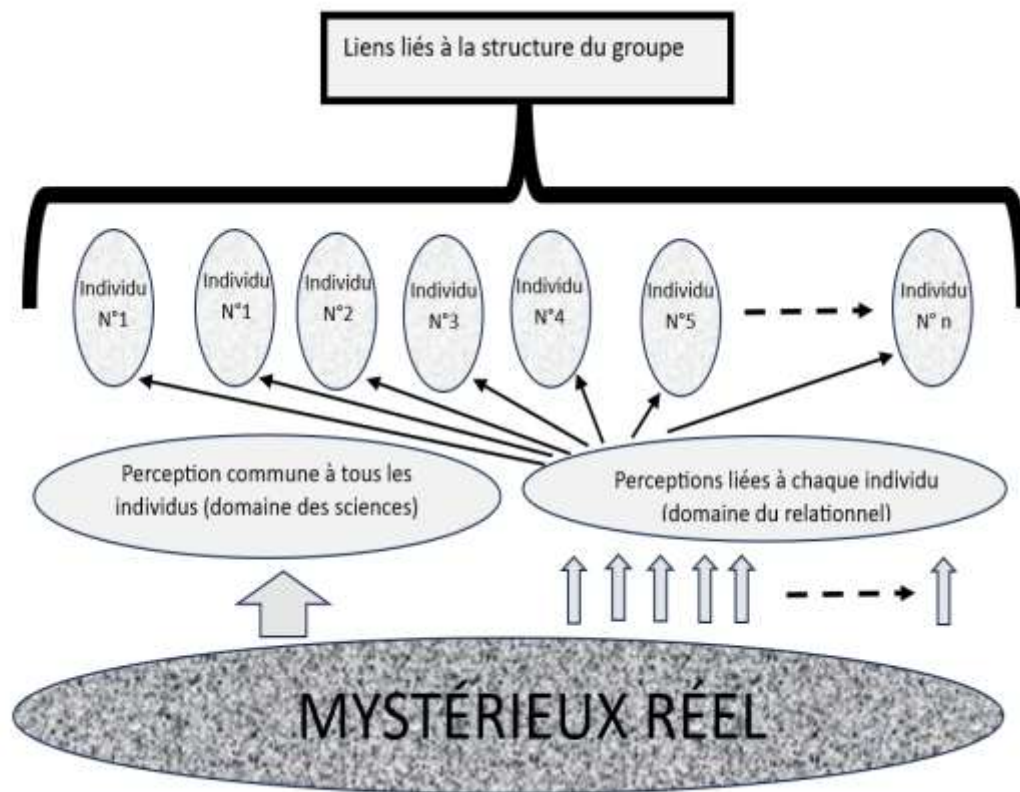
Cette image du Réel nous nous la forgeons à travers l'observation de la répétition d'événements et dans la confrontation avec les autres personnes d'une façon directe (échanges) ou indirecte (lecture, média, enseignement, etc.).

Ceci implique qu'il y a autant d'images du Réel qu'il y a d'individus et que les opinions sont par nature hétérogènes.

Lors d'échanges, cette hétérogénéité des perceptions du Réel demande à être gérée pour permettre la vie en société. Les différents acteurs doivent être capables de s'entendre sur une vision commune du Réel, objet de leurs transactions. Par exemple, dans le commerce de denrées alimentaires, il est nécessaire de connaître le poids de l'objet échangé. En l'absence d'utilisation de balance, l'acte de vente devient vite conflictuel car les opinions sur le poids de la marchandise entre le vendeur et l'acheteur varient énormément. Comme le montre cet exemple la gestion des différentes opinions de chacun a amené la construction « d'objets transitionnels » semblables à la balance. Dans le domaine matériel, ce besoin a progressivement construit la Science.

Dans domaine relationnel, la répétition des situations que l'on utilise dans le domaine matériel pour mesurer le Réel, n'est guère possible. En effet, grâce à la balance, on peut s'entendre facilement sur le poids d'un objet car il sera approximativement le même chaque fois qu'on répètera la pesée. Dans le domaine relationnel ce type de démarche n'a pas grand sens.

D'où le schéma suivant qui fait le point sur ces affirmations :



Pour comprendre le fonctionnement des objets, on est amené à créer des conditions d'expériences reproductibles.

Pour comprendre le fonctionnement des groupes humains on est amené à analyser l'hétérogénéité des opinions.

D'où l'importance de définir un groupe de références (ce groupe de référence existe toujours : il suffit de prendre l'ensemble de l'humanité)

Mais quels groupes et comment ?

Plus la contrainte est grande, plus les opinions des individus vont converger. La création des institutions a permis cette mise en place de contraintes suffisamment fortes pour rapprocher les opinions liées au fonctionnement des services apportés.

Comme on le fait toujours en science, pour analyser le fonctionnement d'une institution on recherchera à repérer les répétitions des opinions des individus du groupe sous contraintes.

Pour cela il faudra récolter les opinions et les analyser ce qui nous amène à rechercher une méthode capable d'atteindre ce but.

Mais comment la créer ? Il fallait se poser la question mais aussi avoir de la chance de tomber sur une expérience qui montrait le chemin.

3. Une enquête qui a tout changé :

L'idée que les questionnaires sont un outil de mesure du relationnel est ancienne. Pour avoir une certaine régularité des réponses à un questionnaire, on imagine facilement qu'il faut que les répondants soient des individus d'un groupe précis. C'est ce que j'ai voulu réaliser en interrogeant les élèves des établissements que je dirigeais. Mais les résultats que j'ai obtenus, s'ils amélioraient la connaissance des publics élèves, ne permettaient pas d'évaluer la qualité du fonctionnement de l'établissement. S'il était possible de mesurer la qualité du fonctionnement des établissements à l'aide de questionnaires, il fallait pouvoir imaginer d'autres types de récolte de données. Tous mes efforts me ramenaient à la question : comment mesurer scientifiquement la qualité de mon management et plus généralement la qualité du management des dirigeants ?

Le hasard a fait, qu'à force de parler de la nécessité de l'évaluation par questionnaires des établissements, le Rectorat d'Amiens, sous la demande du Ministère me commanda la réalisation de l'évaluation de cinq collèges de l'Oise.

Le choix de ces collèges m'était imposé. Cinq collèges furent choisis par le Rectorat. En collaboration avec les principaux de chaque établissement, des élèves et des CPE, j'ai élaboré un questionnaire comportant 123 questions.

Ces questionnaires ont été imprimés et distribués à chaque élève dans chaque collège avec une procédure papier, garantissant un anonymat absolu.

Avec l'aide du Centre Régional de Documentation Pédagogique (CRDP) de Beauvais, les questionnaires ont été numérisés de façon à pouvoir être exploités à l'aide du logiciel d'analyse des données SPHINX.

Rien d'extraordinaire jusque-là tant que les données n'ont pas été exploitées.

En poste comme proviseur d'un gros établissement technique, je n'avais pas le temps de me pencher sur les résultats de l'enquête avant les grandes vacances d'été. C'est ce que je fis dès la fin juillet.

Quand j'ai pu enfin examiner les résultats, ce que j'observais me semblait inimaginable. Pour un nombre important de questions, les distributions (diagrammes en bâtons) générées par les réponses ne différaient pas d'une façon significative d'un collège à l'autre.

Ma première idée a été de remettre en cause ces résultats. Où était l'erreur dans les procédures utilisées ?

Après avoir revisité de toutes les façons possibles les protocoles mis en place, je me rangeais à l'évidence : les résultats étaient incontestables. Il fallait alors donner un sens à mes observations.

Pour répondre à mes obligations administratives j'ai réalisé un rapport très neutre que j'ai remis au Rectorat d'Amiens. Vous trouverez ce rapport très détaillé à l'adresse :

<https://evallib.org/Documentation/5-col-Oise/5-col-Oise.pdf>

Comme je m'y attendais, le document remis ne suscita aucune réaction des autorités académiques. Le Mammouth en a vu d'autre et la poussière ne manquait pas !

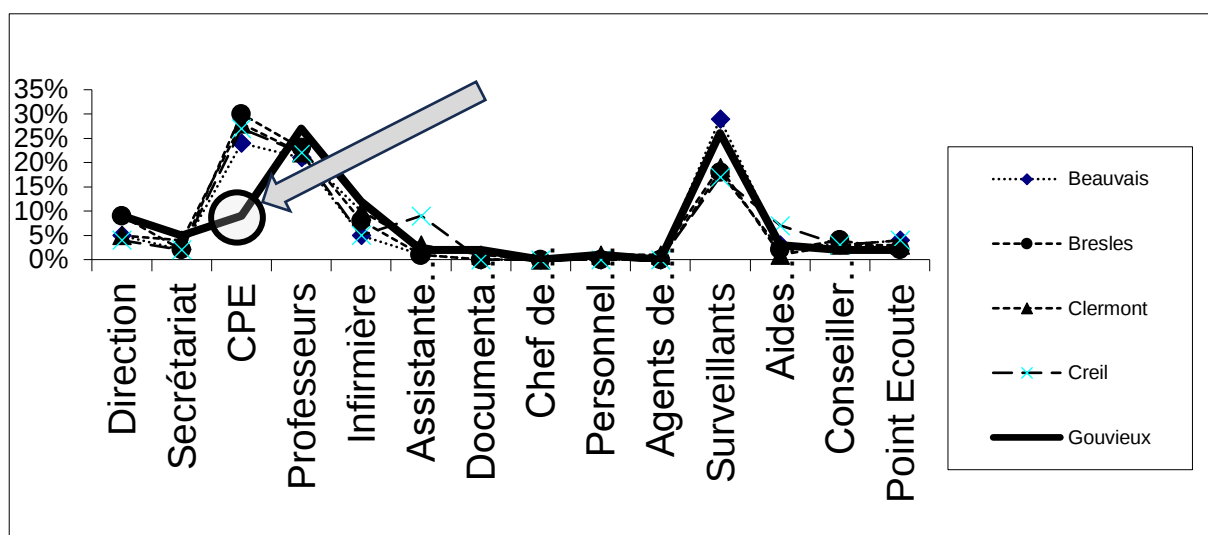
Personnellement je voulais comprendre ce que signifiaient ces résultats et comment on pouvait les utiliser.

Dans un premier temps, j'ai été attiré par un résultat qui, s'il s'avérait significatif, ouvrait des perspectives importantes dans la gestion des établissements scolaires.

Voici la question et les réponses récoltées :

En cas de problèmes, à qui t'adresses-tu au collège ?

	Direction	Secrétariat	CPE	Professeurs	Infirmière	Assistante Sociale	Documentaliste	Chef de travaux	Personnel d'Intendance	Agents de service	Surveillants	Aides Éducateurs	Conseiller d'Orientation	Point Écoute	TOTAL
Beauvais	5%	2%	24%	21%	5%	1%	0%	0%	0%	0%	29%	3%	3%	4%	100%
Bresles	9%	2%	30%	23%	8%	1%	0%	0%	0%	0%	18%	2%	4%	2%	100%
Clermont	5%	4%	28%	22%	10%	3%	1%	0%	1%	1%	19%	1%	3%	3%	100%
Creil	4%	2%	27%	22%	5%	9%	0%	0%	0%	0%	17%	7%	3%	4%	100%
Gouvieux	9%	5%	9%	27%	12%	2%	2%	0%	1%	0%	26%	3%	2%	2%	100%
TOTAL	6%	3%	24%	23%	8%	3%	1%	0%	1%	0%	22%	3%	3%	3%	100%



On remarque que le CPE (Conseiller Principal d'Éducation) de Gouvieux ne répond pas aux besoins d'écoute des élèves. Est-ce vrai ou faux ? Il fallait être sûr.

Je me suis renseigné très discrètement sur la personne concernée. Ce professionnel était connu pour son incapacité notoire et avait été placé là par l'administration : le collège recrutait sur un secteur favorisé, réputé à priori comme calme. Sans avoir à déplacer un inspecteur, grâce à ce résultat on pouvait identifier les personnels déficients. Pour l'administration, c'était forcément une « bombe » dans les pratiques administratives. Mais cela ne m'intéressait que d'une façon anecdotique. Ma recherche avait deux objectifs :

- ⇒ Être sûr que les données récoltées étaient fiables.
- ⇒ Monter que cet outil était capable de mesurer la qualité du fonctionnement d'une institution.

Les arguments qui m'étaient donnés pour mettre en cause la validité des résultats était du type : les élèves ne répondent pas honnêtement aux questions, ils ne sont pas capables de les comprendre, etc.

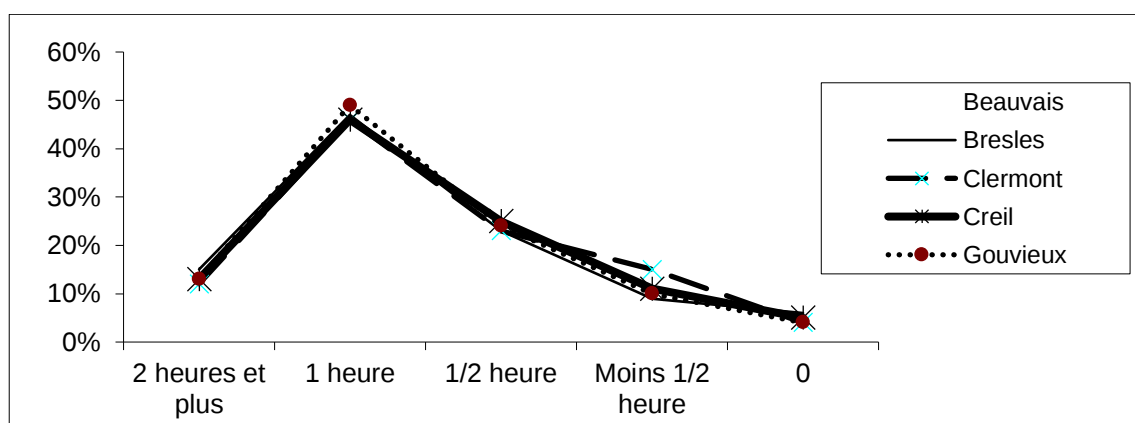
Ces objections ne tiennent pas car si c'était vrai, les distributions ne seraient pas aussi stables d'un établissement à l'autre. D'autre part ce qui nous intéresse ici, ce n'est pas la valeur absolue des réponses mais la comparaison entre les différentes distributions de chaque établissement.

D'où la possibilité d'affirmer que les résultats sont utilisables tels quels, même s'ils sont par nature relatifs.

La capacité des questionnaires à évaluer la qualité du fonctionnement d'un établissement est plus délicate à mettre en valeur. Un résultat m'a dès le début interpellé : les réponses sur le travail à la maison.

Chaque soir que consacres-tu à ton travail scolaire ?

	2 heures et plus	1 heure	1/2 heure	Moins 1/2 heure	0	TOTAL
Beauvais	11%	46%	28%	11%	4%	100%
Bresles	15%	47%	23%	9%	6%	100%
Clermont	12%	46%	23%	15%	4%	100%
Creil	13%	46%	25%	11%	5%	100%
Gouvieux	13%	49%	24%	10%	4%	100%
TOTAL	13%	47%	25%	11%	4%	100%



Ces résultats sont stupéfiants car ils ne sont pas liés au milieu social ou très peu, comme le montre le profil des publics de ces collèges au moment de l'enquête.

	Situation du collège	Origine sociale
Beauvais	Collège de centre-ville	Classes moyennes supérieures
Bresles	Collège très rural	Milieu agricole
Clermont	Petite ville avec gros centre de psychiatre	Milieu hospitalier
Creil	Cités HLM	Milieu très populaire à forte proportion d'enfants d'émigrés
Gouvieux	Zone très résidentielle	Classe supérieure

Ces résultats sont très troublants et révèlent une réalité humaine surprenante. Mais ils sont incontestables. Je suis arrivé à la conclusion, et ceci sur plusieurs années et après de nombreuses enquêtes que l'explication de ces résultats avait deux origines :

- La contrainte sur les individus de l'organisation du collège
- Les comportements des élèves sous les contraintes du collège

On est amené à admettre que les individus, dans des groupes sous contraintes, se positionnent en distribution d'une façon quasi identique d'un groupe à l'autre et que les écarts observés sont révélateurs des particularités locales.

La conséquence de cette observation est grande et complexe :

- Grande car c'est un outil d'évaluation des institutions nouveau et redoutable
- Complexe car le choix des questions significatives se pose et demande une méthodologie

Cette expérience montre que l'on peut construire une science du relationnel à condition de mettre en place les outils adaptés. D'où la nécessité de mettre en place une méthodologie. C'est ce que nous allons développer dans la suite de ce papier.

4. [Généraliser les conditions d'expérience de la première enquête](#)

Pour reproduire les conditions d'expérience de l'enquête que nous venons d'utiliser, il est nécessaire de disposer de **plusieurs évaluations sur des groupes d'individus soumis aux mêmes contraintes**. Ce qui permet d'avoir un **panel de données** comparables à exploiter. C'est le cas des institutions comme les établissements scolaires, les universités, les centres pénitenciers, les hôpitaux, les EHPAD(s), les grosses entreprises, etc.

Ce préalable étant avéré, la construction et la mise en place du questionnaire restent délicates. Les destinataires des questionnaires sont naturellement les usagers (ici les élèves ou les parents) et les professionnels (ici les professeurs et les personnels de service ou d'administration).

Nous proposons dans la suite de ce document une méthodologie d'évaluer le fonctionnement d'une institution à travers des questionnaires et dont les questions ont la **propriété de créer par leurs réponses des distributions proches**.

5. [Le cadre des évaluations : dimension objective et dimension relationnelle](#)

Quand on réalise une évaluation deux éléments interviennent dans la qualité du fonctionnement. La dimension objective des conditions de travail (bâtiments, postes de personnels, lieux d'implantation, etc.) et la dimension relationnelle liée aux groupes des usagers et des personnels. Si l'analyse des données objectives est la base de toute évaluation, la dimension affective qui dépend des opinions des acteurs n'en est pas moins fondamentale. En effet si les relations humaines sont très dégradées, le fonctionnement d'un groupe peut être catastrophique alors que les conditions objectives sont bonnes.

- ⇒ L'analyse des conditions objectives se réalise à l'aide d'études documentaires et de visites in situ. Pratiques obligatoires et connues sur lesquelles nous ne nous appesantirons pas.
- ⇒ L'évaluation de la qualité du relationnel se réalise en envoyant des inspecteurs ou en réalisant des questionnaires

Si on choisit le questionnaire, pour évaluer le relationnel, plusieurs conditions sont à respecter :

- ⇒ Pour la construction du questionnaire :
 - La mise en place d'un groupe de pilotage.
 - La définition des missions pour construire les questions échelles.
 - La liste des services pour construire les questions à choix multiple.
 - Une procédure de récolte de données garantissant l'anonymat.
- ⇒ Pour l'analyse des données :
 - Disposer d'un panel de données sur des groupes comparables
 - Analyser les tris à plats (distributions)
 - Choisir des couples de données pertinentes pour les croiser
 - Réaliser des ACP (Analyse en Composantes Principales) et des régressions avec des logiciels adaptés.

6. [L'analyse qui précède la construction des questionnaires](#)

Toutes les institutions ou les services ont en commun un objectif qui se décline en mission et qui se traduit par l'organisation de services. Ces deux dimensions structurent la création des évaluations internes ou externes. D'où deux types de questions, les échelles pour la mesure des missions et les questions à choix multiples pour la mesure des qualités ou des dysfonctionnements des services.

La mesure de la qualité de la réalisation des missions relève du ressenti des usagers et des personnels. Il s'agit de donner une image des opinions des différents acteurs. Les échelles à 6 degrés sont particulièrement adaptées

7. [La forme des questions utiles](#)

Précisons les types de questions nécessaires au bon déroulement d'une évaluation en les illustrant par des exemples de l'étude des « cinq collèges de l'Oise ».

Nous avons classé les questions en trois grandes catégories :

1) Les questions échelles liées aux missions

⇒ Les questions liées à la satisfaction du service rendu par le collège par exemple :

- Es-tu à l'aise avec les adultes du collège ?
- Comment juges-tu les conditions de vie au collège ?
- Comment juges-tu les relations entre élèves dans le collège ?
- Par leurs attitudes négatives, combien d'élèves gênent-ils le fonctionnement normal de la classe
- Penses-tu que les professeurs font leur possible pour ta réussite scolaire ?
- Trouves-tu que les surveillants sont ?
-

⇒ Les questions qui décrivent le vécu du groupe « élèves » par exemple :

- Comment juges-tu tes capacités en Français ?
- Quel souvenir gardes-tu de l'école élémentaire ?
- Es-tu à l'aise avec tes amis ?
-

2) Les questions à choix multiples liées aux services rendus ou attendus

- En cas de problèmes, à qui t'adresses-tu au collège ?
- Qu'attends-tu le plus des adultes du collège ?
- Quelles devraient être à ton avis les préoccupations des adultes vis-à-vis des élèves du collège ?
-

3) Les questions qui définissent une partition dans le groupe des élèves (les usagers) par exemple :

- Nombre de frères et de sœurs ?
- Où habites-tu ?
- Dans quel type de maison habites-tu ?
- Profession de ton père ?
- Profession de ta mère ?
-

Ces trois types ont un rôle très différent dans l'analyse des données et nous donnent une direction pour créer **les bonnes pratiques professionnelles de l'évaluation des institutions**.

8. [La nature des indicateurs générés par des questions](#)

Les résultats des questionnaires sont **des indicateurs** qui à chaque individu font correspondre ses réponses.

Si les questions sont à choix unique, il y a autant d'indicateurs que de questions.

Si les questions sont à choix multiples, il y a par question autant d'indicateurs que de choix possibles.

Les questions à choix uniques sont de deux types : les échelles et les partitions. Les échelles sont des questions du type ci-dessous :

Comment jugez-vous la qualité de l'accueil des usagers au quotidien ?

Excellent		Bon		Assez bon		Médiocre		Mauvais		Très mauvais
-----------	--	-----	--	-----------	--	----------	--	---------	--	--------------

Ou sous une autre forme équivalente :

Comment jugez-vous la qualité de l'accueil des usagers au quotidien ?

+++		++		+		-		--		---
-----	--	----	--	---	--	---	--	----	--	-----

C'est à dire des questions demandant des réponses uniques graduées. Elles sont classées en base de données sous forme de distributions.

Les partitions sont des questions qui partagent le groupe questionné en sous-groupes comme les classes d'âge, les fonctions, etc. Elles ne sont pas ordonnées.

Les questions à choix multiples sont très commodées pour identifier la qualité des services. Elles sont aussi stockées dans la base de données comme autant d'indicateurs que de possibilités données par les choix multiples.

9. [Panel de données : force de l'évaluation externe](#)

Si on ne dispose que de données sur son propre établissement, on est dans l'impossibilité de savoir si les résultats obtenus révèlent un fonctionnement satisfaisant ou non.

Voici un exemple pour illustrer cette affirmation :

Les plats servis à une cantine sont toujours dépréciés alors que le même plat servi dans un hôtel quatre étoiles va être apprécié. Ainsi le changement de structure, à question et objet évalué identique change le résultat de l'évaluation. La structure a donc une importance primordiale sur les réponses données. Il en résulte que la somme des réponses positives à la question donnée n'apporte aucune information significative.

Pour sortir de ce genre de situation il est nécessaire de disposer des résultats sur plusieurs structures comparables. Dans notre exemple, pour connaître la qualité du service apporté à une cantine, il faut les comparer à ceux d'autres cantines.

C'est pourquoi, pour **juger** avec rigueur le fonctionnement d'une institution, il est nécessaire de disposer d'un panel d'institutions comparables. Cela permettra de définir des taux attendus et de les comparer aux réponses de telle ou telle institution comparable.

Ce type d'évaluation qui nécessite plusieurs institutions pour constituer une base de données se nomme **évaluation externe**.

En l'absence de données issues d'un panel quel est l'intérêt d'une évaluation interne faisant appel à des questionnaires ? La question se pose.

Dans ce cadre, l'évaluation quitte le champ de la mesure de la qualité pour devenir un outil transitionnel de management. C'est le champ d'action des **évaluations internes** qui rentre dans la démarche qualité d'amélioration continue du fonctionnement.

10. Panel de données : les qualités particulières des indicateurs relationnels

Nous allons voir pourquoi nous ne pouvons pas nous contenter pour mesurer la qualité du fonctionnement d'une institution d'une seule valeur comme on s'en contente dans d'autres domaines comme par exemple lorsqu'on mesure une masse.

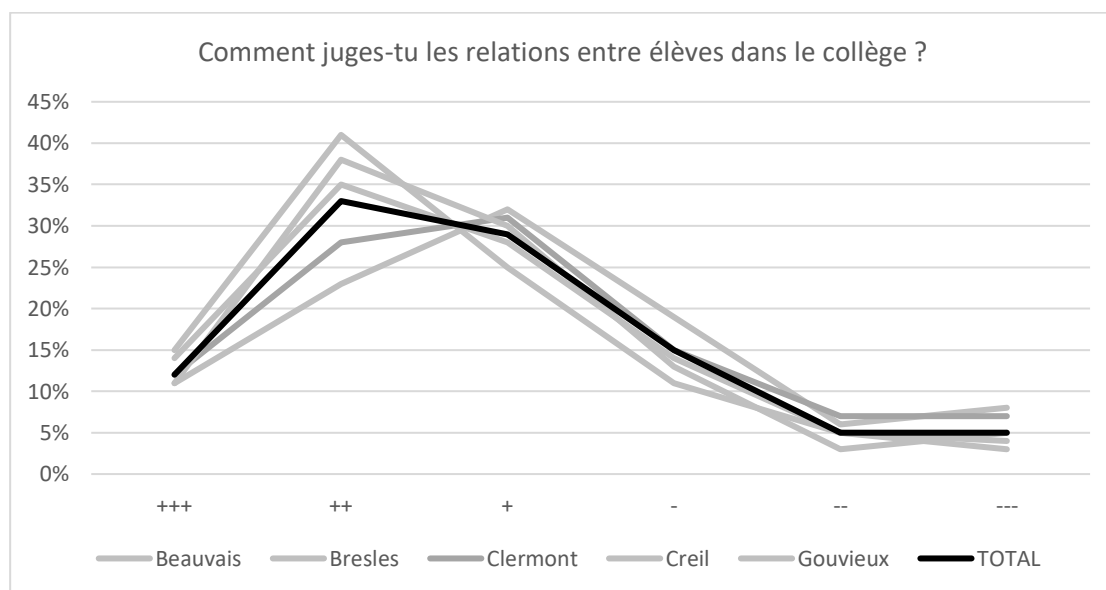
Pour cela prenons un exemple donné dans le tableau suivant issu de données réelles d'une enquête sur cinq collèges de l'Oise. (<https://evallib.org/jules/jules.html>)

*Question posée aux élèves : « **Comment juges-tu les relations entre élèves dans le collège ?** »*

Les réponses obtenues regroupées dans un tableau sous forme de pourcentages sont :

	+++	++	+	-	--	---	TOTAL
Beauvais	14%	35%	28%	14%	5%	4%	100%
Bresles	12%	23%	32%	19%	6%	8%	100%
Clermont	12%	28%	31%	15%	7%	7%	100%
Creil	15%	41%	25%	11%	5%	3%	100%
Gouvieux	11%	38%	30%	13%	3%	5%	100%
TOTAL (courbe de l'ensemble de la population de tous les collèges)	13%	33%	29%	15%	5%	5%	100%

Et les graphes des distributions correspondantes sont alors :



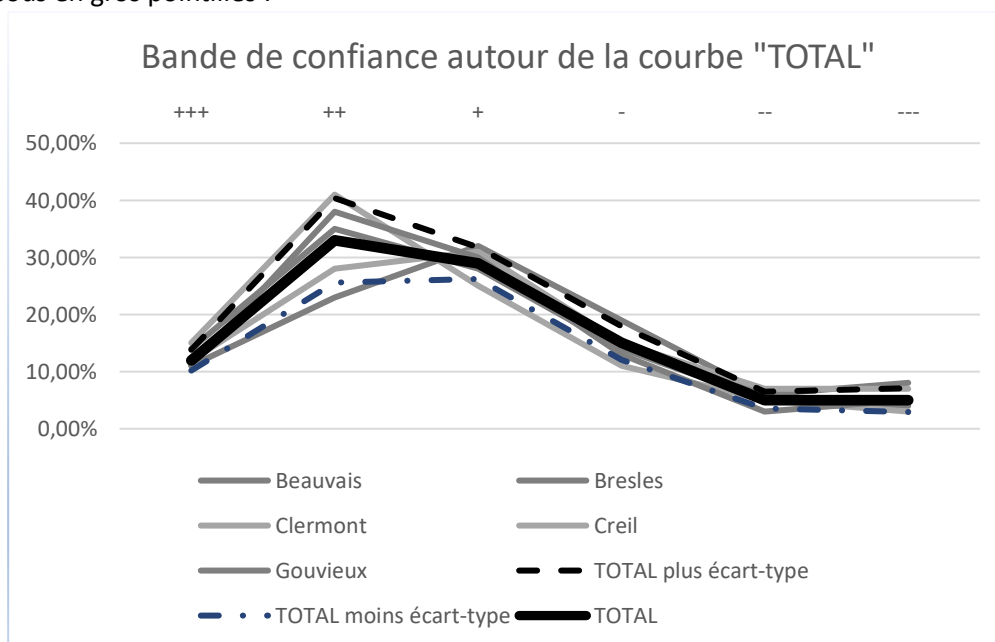
On observe une forme générale donnée par la courbe « TOTAL » que nous allons exploiter. Cette courbe « TOTAL » représente l'ensemble des données de tous les collèges réunis. Elle peut se calculer aussi en utilisant la moyenne (pondérée) des toutes les courbes.

On remarque aussi que chaque courbe « suit » la courbe « TOTAL ». Ce qui définit une forme générale des données.

Pour préciser cette notion de « forme » on est amené à calculer les variations des courbes autour de la courbe « TOTAL » à l'aide de l'écart-type (qui est une sorte d'écart moyen) comme nous le faisons ci-dessous :

	+++	++	+	-	--	---	TOTAL
Beauvais	14,00%	35,00%	28,00%	14,00%	5,00%	4,00%	100%
Bresles	11,00%	23,00%	32,00%	19,00%	6,00%	8,00%	100%
Clermont	12,00%	28,00%	31,00%	15,00%	7,00%	7,00%	100%
Creil	15,00%	41,00%	25,00%	11,00%	5,00%	3,00%	100%
Gouvieux	11,00%	38,00%	30,00%	13,00%	3,00%	5,00%	100%
TOTAL (courbe de l'ensemble de la population de tous les collèges)	12,00%	33,00%	29,00%	15,00%	5,00%	5,00%	100%
Écart-type (sort d'écart moyen)	1,82%	7,38%	2,77%	2,97%	1,48%	2,07%	
TOTAL plus écart-type	13,82%	40,38%	31,77%	17,97%	6,48%	7,07%	
TOTAL moins écart-type	10,18%	25,62%	26,23%	12,03%	3,52%	2,93%	

Calcul qui permet de représenter une bande de confiance autour de la courbe « TOTAL » comme ci-dessous en gros pointillés :



Dans la pratique, quand on est dans cette situation et que le nombre d'institutions augmente, la courbe « TOTAL » se stabilise et devient une sorte de modèle de la forme des réponses que nous nommerons « Taux attendu par l'évaluation ». D'où la définition :

➤ **Définition du « Taux attendu par l'évaluation » :**

On appelle « Taux attendu par l'évaluation » d'un ensemble de distributions conditionnelles, liées à un panel d'institutions pour une question échelle donnée, la limite (si elle existe) de la moyenne pondérée de toutes les distributions quand le nombre d'institutions augmente.

Il faut remarquer que cette définition n'est pas suffisante pour décrire la situation qui nous intéresse. En effet cette définition a toujours un sens même si les courbes liées à chaque institution (distributions conditionnelles) « ne suivent pas la forme générale de la courbe TOTAL ». Sur le plan pratique on calcule la « bande Écart-Type » comme nous l'avons fait ci-dessus et considérerons que si l'Écart-type est toujours inférieur à 10% (ici il est inférieur à 7,38%), alors nous sommes dans la situation qui nous intéresse et que nous nommerons « **Indicateur Relationnel** ». D'où la définition suivante :

Définition des indicateurs relationnels :

Nous désignerons par **indicateur relationnel** un indicateur défini sur **plusieurs structures comparables** (organisation et contraintes) et dont les distributions, structure par structure, (distributions conditionnelles) sont **proches de la courbe « TOTAL »**

Le mot « proche » étant subjectif comme nous l'avons fait remarquer.

Alors se pose la question

Les situations, où la notion de « Taux attendu » et « d'Indicateur Relationnel » sont-elles fréquemment rencontrées ? Et si oui pour quelles raisons ?

A la première question la réponse est oui. Cela peut paraître curieux mais c'est l'expérience qui le prouve. Regardez pour vous en persuader le document « Audit assisté par questionnaires de cinq collèges de l'Oise » à l'adresse : <http://evallib.org/Documentation/5-col-Oise/5-col-Oise.pdf>.

Plus de 80% des questions ont des résultats qui justifient l'utilisation de la notion d' « Indicateur relationnel et de Taux attendu par l'évaluation ». Oui mais pourquoi « ça marche » ?

La réponse n'est pas simple, mais le constat est là. Il est compréhensible que dans la réponse d'un individu rentrent en jeu deux dimensions : le caractère de l'individu et la pression qui lui est exercée par la structure. Le comportement de chaque individu est modifié quand il est dans sa sphère de travail ou lorsqu'il va demander un service. Mais cela n'explique rien. Nous n'avons pas d'explication rigoureuse de ce phénomène mais nous tenterons quelques éléments d'explication dans le paragraphe suivant.

➤ Quelques éléments d'explication sur le comportement des indicateurs relationnels

Dans une institution les usagers ou les professionnels sont, explicitement ou non, recrutés ou admis sur des critères précis. Lorsqu'un individu répond à une question (Q) sa réponse est liée à sa personnalité et aux conditions de vie au sein de l'institution. (*)

Si la perception de la qualité de vie dans l'institution est la même pour tous, les personnalités des répondants ne sont pas liées à une institution en particulier. Il semblerait que, pour une question donnée, quelles que soient les institutions comparables, les répondants ont des sensibilités globalement identiques (les distributions sont proches).

Pour comprendre ce phénomène **Supposons** que nous soyons dans le cas où les individus (les usagers ou les professionnels) de chaque institution sont une représentation fidèle d'un échantillon de la population et désignons par (I) la variable qui à chaque individu fait correspondance l'établissement où il demeure. Considérons la variable contrainte (C) qui à chaque individu fait correspondre sa sensibilité.

Alors : les variables C et I seraient **indépendantes**. Toujours dans ce cas totalement théorique, les distributions conditionnelles de chaque institution seraient **confondues**.

Dans la pratique les variables C et I sur certaines questions sont « presque » indépendantes ce qui explique les résultats observés. C'est généralement le cas des questions échelles décrivant les missions de l'institution concernée.

(*) Nous reviendrons plus tard, en annexe, sur cette question lors de l'analyse mathématique des données.

11. Panel de données : la possibilité de créer des plateformes d'évaluations fiables

La lourdeur des évaluations suggère la recherche d'automatisation pour créer des plateformes. La question qui se pose est la fiabilité d'une telle démarche. Nous nous sommes « attaqués » à ce problème dans une application que vous trouverez à l'adresse <https://evallib.fr>.

Quelles sont les savoir-faire que nous avons retirés de cette expérience ?

⇒ La création des questionnaires :

La création des questionnaires se fait comme nous l'avons expliquée en impliquant des usagers et des personnels dans un groupe de pilotage. L'expérience montre qu'au bout de cinq ou six évaluations les questions et leur nombre se stabilisent.

⇒ La récolte des données :

Elles sont de deux types :

- Les données objectives, c'est-à-dire les caractéristiques de l'institution (surface des bâtiments, nombre de personnels, nombre d'usagers, types de services rendus, etc.
- Les données subjectives correspondantes aux réponses des questionnaires. Elles peuvent se faire sur internet et demandent un accompagnement. Elles peuvent aussi se faire sur papier à condition de disposer d'un lecteur par scanner. Dans tous les cas une réunion d'information auprès des usagers et des personnels est nécessaire pour éviter les rejets et les blocages.

⇒ L'analyse des données :

C'est la partie la plus délicate du fonctionnement de la plateforme. La base de données créée par l'ensemble des réponses aux questionnaires s'enrichit au fur et à mesure de l'augmentation du nombre d'institutions évaluées. Les distributions liées à l'analyse des missions se stabilisent définissant des taux attendus. Jusque-là tout va bien. Mais comment répondre à l'administration qui demande quels sont les points forts et les points faibles de telle ou telle institution ?

Pour y répondre, il faut au moins par mission être capable de savoir si elle est bien assurée ou non. C'est-à-dire classer le fonctionnement de chaque institution sur chaque mission (chaque question) en trois ou quatre catégories. Ce qui pose deux problèmes :

- Pour classer des objets, il faut qu'ils soient mesurés d'une façon unidimensionnelle (par un nombre) ce qui n'est pas le cas des missions qui sont mesurées par des distributions. D'où une impasse
- Être capable à chaque demande de l'administration de faire une réponse automatique

D'où la question : Comment peut-on réduire les données issues de distributions qui sont des fonctions (six valeurs par questions) en une seule valeur et est-ce qu'on en a le droit ?

⇒ La réduction des données

Nous nous plaçons ici dans le cas d'un ensemble d'institutions comparables et d'une question-échelle portant sur une mission. Lorsqu'on est dans ce cas chaque distribution liée à cette question comporte une partie positive et une partie négative. Alors l'idée la plus simple est de remplacer les données de chaque distribution par la somme des valeurs positives. Nous dirons pour simplifier la compréhension, que c'est la valeur positive (de la distribution).

C'est simple mais réducteur. La question est donc de se demander si cette réduction n'est pas abusive. La réponse est mitigée.

D'une part, la propriété N°2 de l'annexe N° 6 permet d'affirmer que valeur positive conserve la propriété des taux attendus. Ce qui veut dire que pour une question donnée la valeur positive de la moyenne pondérée sera la moyenne pondérée des valeurs positives.

Ce n'est pas suffisant car dans le cas général deux distributions ayant la même valeur positive peuvent avoir des profils très différents ce qui amène à être très prudent.

Cependant, il faut aussi prendre en compte le fait que dans la pratique, les distributions liées à des questions portant sur des missions sont en probabilités dans une « bande de confiance (voir paragraphe 14). De ce fait les parties positives de l'ensemble des distributions liées à une mission sont contraintes et ne peuvent pas se comporter d'une façon totalement aléatoire. D'où la stratégie suivante pour une question donnée liée à une mission :

- ⇒ Prendre pour chaque distribution les parties positives
- ⇒ Disposer ainsi d'un ensemble de valeurs réelles ordonnées comprises entre 0 (0%) et 1 (100%)
- ⇒ Partager les données ainsi récoltées en quatre groupes d'effectifs égaux (quartiles)
- ⇒ Y associer une valeur de la façon suivante
 - Le groupe faible : Quartile 0 -----→ Fonctionnement à regarder d'urgence
 - Le groupe le moyen faible : Quartile 1 -----→ Fonctionnement à surveiller
 - Le groupe le moyen fort : Quartile 2 -----→ Fonctionnement satisfaisant
 - Le groupe fort : Quartile 2 -----→ Fonctionnement très satisfaisant
- ⇒ Définir la bande de confiance (taux attendu plus ou moins l'écart type)
- ⇒ Sélectionner les distributions qui ne sont pas dans la bande de confiance pour faire une analyse détaillée

Cet ensemble d'outils permet d'automatiser les évaluations externes au sein de plateformes.

12. Absence de panel de données : l'exploitation des évaluations internes

Disposer d'un panel de données permet de juger la qualité d'une institution ou d'un service mais suppose, au préalable, d'effectuer un ensemble d'évaluations de groupes sous contraintes comparables.

Dans la pratique ce n'est pas souvent le cas : on est souvent amené à utiliser l'évaluation pour analyser son propre fonctionnement.

Ne pouvant juger la qualité du fonctionnement l'objectif les **évaluations internes** se limite à l'analyse de la **qualité du relationnel** et la **recherche axes de progrès**. Plus que pour les évaluations externes l'association des acteurs est de rigueur.

Si l'objectif est différent les évaluations externe et internes procèdent dans leur mise en place de méthodologies comparables. Voici dans les grandes lignes les différentes étapes d'une évaluation interne

- ⇒ Trouver un professionnel de l'évaluation ou une formation pour les dirigeants
- ⇒ Avertir les usagers et les personnels de l'imminence et de l'objectif de l'évaluation
- ⇒ Donner une liste des étapes de l'évaluation future
- ⇒ Créer un groupe de pilotage composé d'usagers et de personnels dont l'objectif est de participer à la création, la diffusion, la récolte de données et la diffusion des résultats
- ⇒ Structurer la construction du questionnaire usager et de celui du personnel en :
 - Identifiant les missions principales (questions échelles)
 - Identifiant les services (questions multiples)
- ⇒ Distribuer le questionnaire et contrôler sa diffusion et les retours

- ⇒ Analyser les données à l'aide d'un logiciel du type Sphinx
- ⇒ Restituer les résultats de la façon la plus factuelle à l'aide de tris à plat
- ⇒ Mettre en place des analyses croisées pour rechercher les axes de progrès

13. Précision sur le rôle des normes dans les évaluations

Lors d'une évaluation les intervenants doivent vérifier un certain nombre de faits ou de situations. Par exemple, la régularité des pesées des résidents ou la présence de défibrillateurs dans un EHPAD ainsi que le nombre d'extincteurs en bon état. Ce sont les normes obligatoires pour bon nombre d'institutions. C'est en fait des indicateurs qui représentent des pré requis du bon fonctionnement d'une institution ou d'un service. Ils sont sensés donner le cap à atteindre pour garantir la qualité des services apportés aux usagers. Souvent décriés, quand ils sont surabondants, ils n'en restent pas moins très utiles. Mais le respect des normes ne garantit en rien la qualité du fonctionnement. La qualité du relationnel reste déterminant dans la qualité du fonctionnement des institutions.

14. Les évaluations des individus basées sur des notations

L'évaluation des individus par d'autres individus relèvent d'un autre processus que celui de l'évaluation du fonctionnement des institutions. Les objectifs et les méthodes recherchés dans ce type d'évaluations sont multiples. Les plus courants sont la progression des connaissances des élèves, le recrutement d'un profil, etc. Le but de ce type d'évaluation est d'affecter à chaque individu d'un groupe une ou plusieurs valeurs à l'aide d'un protocole établi à l'avance. Souvent on récolte des données numériques à partir d'un groupe d'évaluateurs (les correcteurs pour l'enseignement) et on réduit les valeurs récoltées par des règles de calcul les plus adaptées possibles à l'objectif recherché (souvent de moyennes). Les méthodes utilisées pour transformer les données et les rendre utilisables pour atteindre l'objectif initial, ont un grand impact sur le fonctionnement des groupes concernés. Pour plus d'information, voir l'annexe 4 : « L'évaluation de l'individu et le principe de Peter » et le texte de la rubrique : « Rétablir l'ascenseur social et l'autorité des professeurs » sur la page d'accueil du site <https://evallib.org>.

15. Les différents types de groupes concernés par l'évaluation du relationnel

Pour pouvoir disposer d'indicateurs porteurs d'informations ayant sens il est nécessaire d'évaluer des groupes sous contraintes. Citons :

- ✓ Les institutions (établissements scolaires, hôpitaux, EHPADs, etc..)
- ✓ Les services d'une institution
- ✓ Les classes d'un établissement scolaire
- ✓ La formation continue dans les entreprises (ex. : évaluation continue chez Peugeot réalisée par l'entreprise Sphinx)
- ✓ L'évaluation de voyages organisés
- ✓ Etc.

Si on désire être capable de juger le fonctionnement on devra mettre en place un protocole d'évaluation et réaliser 7 à 10 évaluations avant de pouvoir avancer des jugements. Le plus

efficace sera de mettre en place un questionnaire usagers et personnels avec des groupes de pilotage.

Si on ne peut pas disposer d'un panel de données issues d'enquêtes antérieures on devra se contenter d'une évaluation interne.

16. Les différents champs de compétences nécessaire aux évaluations

Comme nous l'avons dit en introduction, l'évaluation s'est longtemps confondue avec la notation ou avec des rapports d'inspection. Ce n'est qu'une vision restreinte qu'il faut dépasser. L'évaluation du fonctionnement de groupes sous contraintes fait appel à des champs plus larges que le simple acte de notation ou celui de l'inspection. Au minimum, lors d'une évaluation les champs suivants sont à solliciter :

- ✓ Champ éthique
- ✓ Champ administratif
- ✓ Champ juridique
- ✓ Champ social
- ✓ Champ de la psychologie
- ✓ Champ statistique
- ✓ Champ informatique
- ✓

a. Champ éthique et la mise en place d'un groupe de pilotage

L'évaluation du fonctionnement d'une institution, dans la mesure où elle prend en compte l'opinion des usagers (ou des clients) et des personnels peut être très vite ressentie comme une intrusion dans la vie des personnes.

Pour éviter toute situation de rejet et par conséquent de nullité des résultats, il faut prendre plusieurs précautions. Dans la phase de création de l'outil d'évaluation la mise en place d'un groupe de pilotage est indispensable. Groupe de pilotage associant usagers, personnel et administration. Si un questionnaire est envisagé la garantie de l'anonymat des réponses est fondamentale.

b. Champ administratif

Dans le processus d'évaluation le rôle de l'administration est central. Elle met en place les protocoles gérant la chronologie de l'évaluation, vérifie le bon déroulement des opérations, diffuse les résultats à l'ensemble de la structure concernée, explique les prises de décisions consécutives à l'évaluation. Il résulte de toutes ces obligations qui incombent au personnel d'une administration, qu'une sensibilisation aux difficultés possibles doit être réalisée avant toute action d'évaluation.

c. Champ juridique

Lors d'une évaluation, la conformité des documents mis en place par l'administration (protocoles, questionnaires, etc.) doit être validée par les services juridiques : si des tensions sociales existent, les risques de procès ne sont jamais négligeables.

d. Champ social

L'introduction d'une démarche d'évaluation du fonctionnement dans une institution n'est pas un acte neutre. L'accompagnement tout au long du processus permet d'éviter des crispations ou des rejets dans la prise d'information. La communication en amont est indispensable. Le déroulement de l'évaluation doit être expliqué : protocole de mise en place, garantie de l'anonymat lors de la récolte des données, traitement des données et modalité de publication des résultats. L'intervention de spécialistes de l'évaluation externe à l'institution ou à l'entreprise, travaillant en

collaboration avec l'administration, est certainement la méthode la plus sûre pour garantir le bon déroulement du processus d'évaluation.

e. Champ de la psychologie

La présentation de l'évaluation a besoin de prendre en compte les sensibilités des personnes interrogées et du milieu auquel elle s'applique. La soumission des procédures et la rédaction des questions à l'approbation de psychologues est souhaitable : une formulation « trop scientifique » peut heurter certaines sensibilités. Des modules de sensibilisation à cette problématique, animés par des psychologues professionnels, sont à intégrer au cursus des futurs évaluateurs professionnels.

f. Champ statistique

Une fois la récolte des données réalisée, l'analyse des résultats se pose. Les simples tris à plat (par exemple les diagrammes en bâtons) sont souvent une source de renseignements très riche mais insuffisante pour prendre des décisions et proposer une politique d'amélioration. Les savoir-faire des statisticiens à ce stade sont incontournables. Mais leur rôle ne s'arrête pas là. Pour que les données soient exploitables, il faut que le choix et la forme des indicateurs (questions) soient conformes aux besoins d'analyse. Un minimum de formation en statistiques pour les professionnels de l'évaluation est souhaitable.

17. Réussir la mise en place pratique des évaluations de terrain

La facilité apportée par les plateformes informatiques pousse les administrations à confier à des prestataires le soin de construire les outils d'évaluation des institutions. Cette pratique risque de générer des blocages et des rejets par les acteurs de terrain, si elle n'est pas consolidée par une implication des usagers et des personnels. La mise en place d'un groupe de pilotage permet d'échapper à ce type de situation. Le rôle de ce groupe sera :

- De créer ou d'adapter les questionnaires ?
- D'intégrer à ce questionnaire les questions imposées par l'administration centrale
- De diffuser et récolter les données (sous forme papier ou informatique) ?
- Prendre connaissance des résultats et les diffuser ?

Tout ceci nécessite l'accompagnement d'un professionnel de l'évaluation

Il en découle que la mise en place de formations d'étudiants aux pratiques de l'évaluation des institutions, au sein des universités, répondraient aux besoins actuels et futurs.

18. L'analyse des données recueillies à l'aide des questionnaires

Les données recueillies à l'aide de questionnaires se rangent dans deux tableaux (bases de données), un des usagers et un autre des personnels. Les lignes représentent les réponses des individus et les colonnes les indicateurs générés. Le traitement des données dépend de l'objectif que l'on recherche.

Il y a trois grandes directions qui supposent une approche différente.

- Les analyses élémentaires du type tri à plat

Ces analyses sont souvent suffisantes pour une exploitation en local. Publier les réponses brutes des questionnaires et les illustrer par des diagrammes est un outil de management très efficace. Il permet d'entamer des réflexions sur les points forts et les points faibles du fonctionnement entre la direction, les personnels et les associations d'usagers. De nombreux logiciels existent pour cela. Par exemple le logiciel SPHINX est l'outil idéal pour réaliser ce travail préparatoire. La limite de ces analyses élémentaires se trouve dans l'interprétation des données : certains résultats ne

permettent pas d'estimer la qualité du travail fait par l'institution car ils sont par nature toujours négatifs quel que soit l'établissement évalué.

- La cartographie des réponses données par les usagers ou les personnels.

C'est la démarche classique d'analyse des données (Type ACP). Ce type d'analyse est basé sur la notion de corrélation, c'est-à-dire de liens entre indicateurs. Là aussi l'appel à un logiciel du type SPHINX convient parfaitement. Dans cette situation il n'est pas nécessaire de disposer d'un panel d'établissements. Les cartes obtenues peuvent permettre d'identifier les individus isolés (dans le groupe d'usagers ou de professionnels) et d'en tirer une réflexion sur des axes possibles de progrès.

Si ce genre d'étude est intéressant, elle ne permet pas de juger si le fonctionnement d'une institution est satisfaisant au regard de l'attendu de l'administration centrale comme nous l'avons déjà mentionné. On a besoin pour cela de données issues d'un panel d'institutions.

- Le positionnement du fonctionnement de l'institution évaluée par rapport aux attentes de l'administration centrale.

Ici le but est différent, on désire répondre à la question : l'institution fonctionne-t-elle d'une façon satisfaisante ? On entend par satisfaisante la mesure de la qualité du fonctionnement par rapport à un standard défini par l'Administration Centrale. On utilise pour cela les notions d'indicateurs relationnels, de taux attendus et de quartiles.

19. La diffusion des résultats et l'après évaluation

Trop d'évaluations restent dans les tiroirs des bureaux de la direction. C'est une erreur qui ne profite ni à la direction, ni aux personnels, ni aux usagers. L'affichage au vu et au su de tout le monde des résultats sous forme de tri à plat est le minimum souhaitable. L'exploitation des données par le groupe de pilotage appuyé par un professionnel de l'évaluation facilitera la production d'une synthèse courte mettant en valeur les points forts et les points faibles. La communication aux conseils d'Administration des résultats et des axes de progrès visés par la direction est un complément très appréciable.

Annexes

1. Annexe 1 : Les différentes formes d'évaluation du relationnel

Le fonctionnement d'établissements ou entreprises n'est pas le seul à pouvoir être évalué à l'aide de la méthode que nous avons proposée.

L'évaluation du fonctionnement de sous-groupes ou de groupes constitués temporaires peut aussi se réaliser d'une façon similaire. Citons comme exemples :

- ✓ Les services d'une institution
- ✓ Les classes d'un établissement scolaire
- ✓ La formation continue dans les entreprises (ex. : évaluation continue chez Peugeot réalisée par l'entreprise Sphinx)
- ✓ L'évaluation de voyages organisés
- ✓ Etc.

Mais il y a une autre évaluation qui fonctionne différemment : c'est l'évaluation de l'individu. Cette évaluation mérite que nous nous y attardions pour deux raisons :

- ✓ L'évaluation de l'individu est la plus ancienne évaluation du relationnel dont l'origine remonte à la plus haute antiquité
- ✓ La caractérisation des compétences d'un individu est fortement liée à sa place dans le groupe

2. Annexe 2 : Évaluation interne et évaluation externe : un peu d'histoire

L'idée d'évaluer le fonctionnement des établissements rendant un service public est très ancienne. Dans un premier temps l'évaluation des institutions a été confiée à des corps d'inspection. Ce système trop coûteux et souvent contesté a évolué dans deux directions complémentaires : les évaluations internes ou auto-évaluations et les évaluations externes réalisées par des organismes agréés.

C'est dans le domaine de la santé que cette réflexion a été menée le plus loin par les services de l'État. La LOI no 2002-2 du 2 janvier 2002 rénovant l'action sociale et médico-sociale inscrit, *pour l'ensemble des ESSMS (Établissements et Services Sociaux et Médico-Sociaux) l'obligation d'évaluation leurs activités et de la qualité des prestations qu'ils délivrent (article L. 312-8 du code de l'action sociale et des familles).*

L'évaluation interne ou auto-évaluation est inscrite à l'article L.312-8 al.1 et L'évaluation externe à l'article L.312-8 al 2 à 4.

Une dernière réforme cadrée par Le décret n° 2021-1476 du 12 novembre 2021 relatif au rythme des évaluations de la qualité des ESSMS et son décret modificatif n° 2022-695 du 26 avril 2022 définissent le nouveau cadre réglementaire applicable aux nouvelles évaluations réalisées sur la base du référentiel de la HAS. L'évaluation interne n'est plus une obligation réglementaire et les évaluations externes sont réalisées par des organismes accrédités et cadrées par un référentiel et un logiciel appelé SYNAE.

Cette dernière réforme est la conséquence de difficultés rencontrées par les ESSMS à mettre en place les évaluations internes et externes jugées trop complexes et trop chronophages.

Pourtant, dans le passé, quand l'évaluation interne a pu se faire sérieusement dans un établissement, l'impact sur le fonctionnement a été généralement positif. Cependant la difficulté

majeure de l'évaluation interne a toujours été la lourdeur administrative. Sa réalisation mobilisait la direction plus de trois semaines en continu à cause du temps nécessaire pour l'organisation des réunions de concertation et la rédaction du rapport. Malgré ces efforts, les professionnels subissent toujours l'évaluation plutôt que d'en bénéficier.

La société et plus particulièrement la classe politique a bien intégré la nécessité pour les institutions de réaliser des évaluations internes et externes, mais les outils manquent pour rendre ces démarches légères sur le plan administratif et par corollaire efficaces. Des solutions existent à travers l'intelligence artificielle qui pourraient concilier le besoin d'évaluation périodique des établissements avec la charge de travail administratif. (*)

(*) Nous avons créé une plateforme (<https://evallib.fr>) capable de réaliser des évaluations internes et externes conformes et non chronophages.

3. Annexe 3 : Un cas « d'école » : la notation, une auto-évaluation particulière

Les établissements scolaires ou universitaires sont des structures particulières au regard de l'évaluation. La mission du groupe des personnels (enseignant et administration) est double : la transmission des savoirs et l'évaluation des acquis. Sans la prise en compte de cette seconde mission les établissements scolaires et universitaires ne fonctionnent pas. L'efficacité de la notation dans le bon fonctionnement des établissements est très liée à la composition du groupe des apprenants. L'expérience montre que plus le groupe est hétérogène plus la notation perd son sens. Cela amène à trouver des outils pour analyser l'hétérogénéité du groupe élèves (ou étudiants). La recherche de ces outils est malheureusement loin de la culture universitaire qui s'intéresse essentiellement à la création des élites.

Dans un autre document intitulé : « Rétablir l'ascenseur social et l'autorité des professeurs au collège » (sur <http://evallib.org/jules/jules.html>) nous montrons l'importance de l'analyse de l'hétérogénéité du groupe élève. Les outils d'analyse que nous avons créés nous ont permis de proposer une organisation qui prendrait en compte l'intégralité des besoins des collégiens. Mais ces outils ne sont pas totalement efficaces quand on s'adresse à des étudiants qui ont déjà des conditions de vie d'adultes. Les d'outils d'analyse de l'hétérogénéité des publics étudiants fait défaut et demanderait qu'on s'y attarde pour permettre aux universités de répondre au plus près aux besoins des étudiants.

Cela étant dit, il ne faut pas sous-estimer l'évaluation de la qualité du fonctionnement qui est nécessaire et se pose comme dans toute institution.

4. Annexe 4 : L'évaluation de l'individu et le principe de Peter

La notation des qualités ou des performances d'un individu est la forme d'évaluation la plus naturelle et la plus ancienne. Elle consiste à émettre une opinion sur un individu à l'aide d'une procédure plus ou moins sophistiquée (correction d'un écrit, passage d'un oral, rapport de stages, entretiens d'embauche, etc.). L'objectif de cette démarche est de classer les individus lors d'un recrutement ou d'un positionnement dans un groupe.

Plusieurs critères sont généralement utilisés qui fournissent un ensemble de notes. Ces notes sont transformées en une seule valeur globale pour permettre un classement des individus.

Mais lorsque on choisit une méthode pour réduire plusieurs notes en une seule note, en fonction de l'objectif recherché et du groupe auquel l'individu est rattaché cette démarche n'est pas toujours justifiée.

Par exemple, lorsque Jules Ferry avait défini la valeur de l'élève par la moyenne de ses notes et classé ainsi les élèves, sa démarche était justifiée au sein du système éducatif de la troisième République. C'est-à-dire lorsque le groupe d'élèves managé par cette technique était une élite. Mais comme nous l'expliquerons dans le document « Rétablir l'ascenseur social et l'autorité des professeurs au collège » (evallib.org/jules/jules.html), cette même technique n'a plus été opérationnelle quand le groupe est devenu très hétérogène après la massification de l'éducation au sein du collège (Réforme Haby).

La commodité apportée dans la gestion des groupes par la réduction de plusieurs notes à l'aide de moyennes peut laisser penser que c'est la seule technique possible, tant elle fut utilisée dans le passé. Nous allons voir sur un exemple que ce n'est pas une fatalité.

Connaissez-vous la loi empirique énoncée par le pédagogue canadien Laurence Peter en 1970 dans son livre paru en France en 1970 : « le Principe de Peter » ? :

“Dans une hiérarchie tout employé a tendance à s'élever au niveau de son incompetence.”

On trouve sur internet de nombreuses explications de sens de ce principe comme celle-ci :

« Un chef ne va pas promouvoir quelqu'un de meilleur que lui ou qui va lui faire de l'ombre. Question d'ego. Au moment de choisir l'élue, « un chef va se demander qui lui sera le plus loyal, le plus utile, le plus redevable. Il cherchera quelqu'un qui ne lui plantera pas un couteau dans le dos ». L'incompétent tombe à pic : ce n'est ni un concurrent, ni quelqu'un d'assez indépendant. Il rassure et conforte le chef dans son propre pouvoir. En le promouvant, ce dernier garde le contrôle. »

Au-delà de l'aspect subversif de ce principe comment expliquer ce phénomène observé par Peter ?

En quoi rentre-t-il dans la problématique de l'évaluation d'un individu ?

Si on s'écarte des situations de favoritisme toujours présentes dans un groupe, la promotion d'un individu dans une institution se réalise généralement à travers l'évaluation de plusieurs compétences. Compétences qui se déclinent sous la forme de critères.

Or plus une situation fait appel à des compétences de management d'équipes, plus le nombre de critères est important. Le manager va se retrouver de fait à comprendre tous les enjeux de ses managers, que ce soit sur le plan technique ou humain.

Posons ce principe pour compléter celui de Peter :

« La valeur d'un manager est nulle si, dans un seul domaine de compétence des équipes qu'il dirige, sa propre compétence est nulle »

Comment traduire techniquement et concrètement ces principes ?

Voici une piste sur des exemples imaginaires :

Supposons que dans une entreprise trois candidats postulent pour un poste de responsable faisant intervenir 3 qualités techniques et la qualité relationnelle.

Type de qualité	Mesure de la qualité N°1	Mesure de la qualité N°2	Mesure de la qualité N°3
Qualité technique 1	80%	75%	75%
Qualité technique 2	95%	45%	12%
Qualité technique 3	92%	82%	82%

Qualité relationnelle	10%	85%	95%
-----------------------	-----	-----	-----

Pour un responsable, le choix est évident. Le candidat N°2 doit être choisi car il est suffisamment compétent dans le domaine relationnel et apte à comprendre les problèmes techniques. Le premier candidat a de fortes chances de générer des conflits : il est trop sensible à la dimension technique du fonctionnement. Le dernier candidat, en valorisant la dimension affective, risque de prendre des décisions qui mettront en péril le bon fonctionnement de l'entreprise.

Il y a-t-il une formule qui arrive à la même conclusion ?

On est sur un choix d'exclusion pour une incapacité dans un domaine donné. Prendre le minimum (noté min) est une première solution :

Valeur (candidat N°1) = $\min(80 ; 95 ; 92 ; 10) = 10$

Valeur (candidat N°2) = $\min(75 ; 45 ; 82 ; 85) = 45$

Valeur (candidat N°3) = $\min(75 ; 12 ; 82 ; 95) = 12$

Le choix ne serait pas aussi facile si le candidat N°1 avait pour valeur de sa qualité relationnelle une valeur proche de la valeur globale du N°2. Par exemple 47%.

Les valeurs obtenues seraient alors

Valeur (candidat N°1) = $\min(80 ; 95 ; 92 ; 47) = 47$

Valeur (candidat N°2) = $\min(75 ; 45 ; 82 ; 85) = 45$

Valeur (candidat N°3) = $\min(75 ; 12 ; 82 ; 95) = 12$

On pourrait hésiter entre les profils 1 et 2 car l'importance accordée au relationnel doit primer sur les autres critères. Plusieurs idées peuvent être mises en œuvre pour modifier les choix. Par exemple on décide qu'un candidat doit être éliminé s'il n'a pas au moins une valorisation sur le plan relationnel de plus de 80%. Pour traduire cette décision on « fabrique » un nombre Élimination (x) = 0 si $x < 80$ et Élimination (x) = 1 si $x \geq 80$.

Ce qui s'écrira, pour les spécialistes avec la formule suivante : $\min(a ; b ; c ; d) \times \text{Élimination}(x)$

D'où le tableau suivant :

Valeur (candidat N°1) = $\min(80 ; 95 ; 92 ; 47) \times 0 = 0$ (élimination pour incompétence au poste)

Valeur (candidat N°2) = $\min(75 ; 45 ; 82 ; 85) \times 1 = 45$

Valeur (candidat N°3) = $\min(75 ; 12 ; 82 ; 95) \times 1 = 12$

L'intérêt de cette réflexion sur le principe de Peter est de démontrer que les décisions administratives ne peuvent pas toujours se baser sur des moyennes fussent-elle pondérées.

Toutes ces explications ne sont que des exemples pour convaincre qu'il n'existe pas, pour transformer plusieurs critères en un seul, uniquement l'utilisation de la moyenne. Chaque situation humaine doit faire l'objet d'analyse pour **adapter les outils de calcul aux besoins du terrain**. On montera que la réduction de plusieurs notes en une seule n'est pas toujours la bonne stratégie à utiliser pour prendre des décisions sur des individus appartenant à un groupe : l'utilisation simultanée de plusieurs indicateurs s'impose souvent.

5. Annexe 5 : L'évaluation à la rescousse de la pédagogie des transmissions des savoirs

Les enseignants ont longtemps retiré leur légitimité de la transmission de leur savoir. Le savoir ne se trouvait que dans peu de livres ou dans la tête des enseignants. Le développement d'Internet a changé cet état des choses. L'enseignant a perdu une part de son aura. Plus grave encore les fake news font que la vérité est contestée. Les élèves ont des difficultés à accepter un enseignement basé uniquement sur l'autorité du professeur.

Il devient nécessaire que les jeunes apprennent à différencier le vrai du faux. Pour arriver à ce résultat, toutes les techniques sont bonnes. Mais la tâche est de plus en plus difficile sur les sujets

déliés. Des professeurs se censurent dans les cours d'Histoire, de Littérature, de Biologie, etc. Les échanges des élèves entre eux et avec les professeurs sont de plus en plus difficiles. Le respect de l'opinion de chacun est problématique.

Comment répondre à cette nouvelle situation ?

Plutôt que de « faire l'autruche » en continuant un enseignement s'appuyant sur l'autorité, il est nécessaire de reconstruire les bases des savoirs avec les élèves. C'est déjà fait dans les petites classes (primaire et collège) dans les disciplines scientifiques où des expériences élémentaires font prendre conscience aux élèves de la différence entre la sensation première et la réalité expérimentale.

Dans le domaine littéraire et historique la situation est autre. Le vrai du faux ne peut être démêlé par des expériences élémentaires. La perception de la réalité est souvent liée aux opinions et à l'expérience. Pour que chacun prenne conscience de la diversité des opinions, les échanges entre élèves sont nécessaires. Le groupe classe a besoin de construire une image fiable, admise par tous et non conflictuelle des écarts entre les visions de chaque individu.

Une bonne solution pour répondre à ce besoin consiste à faire appel à des auto-questionnaires. Cela permet dès le plus jeune âge de renvoyer au groupe classe une image de la diversité des opinions des élèves sans créer de conflit. Chaque étape de ce processus d'auto-évaluation améliore la conscience de chacun :

- ⇒ La définition précise du sujet
- ⇒ La construction collective du questionnaire
- ⇒ Le dépouillement des réponses
- ⇒ L'analyse des données

La mise en place d'auto-évaluation assistée par questionnaire, améliorerait la connaissance du réel et par ricochet la transmission des savoirs et l'autorité des enseignants.

6. Annexe 6 : Quelques mathématiques qui justifient le traitement des données

La pratique de l'évaluation des institutions amène à modifier les données par différentes méthodes calculatoires sans savoir si ces transformations sont justifiées. Or des calculs qui semblent logiques peuvent amener les administrations à prendre des décisions erronées et au final très graves. Ici nous allons essayer de mathématiser les données issues des évaluations pour justifier ou pour rejeter certaines pratiques courantes comme la réduction de distributions issues d'une question à leurs pourcentages de valeurs positives. Par exemple à partir du tableau suivant déjà utilisé :

Question posée aux élèves : « **Comment juges-tu les relations entre élèves dans le collège ?** »

	+++	++	+	-	--	---	TOTAL
Beauvais	14%	35%	28%	14%	5%	4%	100%
Bresles	12%	23%	32%	19%	6%	8%	100%
Clermont	12%	28%	31%	15%	7%	7%	100%
Creil	15%	41%	25%	11%	5%	3%	100%
Gouvieux	11%	38%	30%	13%	3%	5%	100%
TOTAL (courbe de l'ensemble de la population de tous les collèges)	13%	33%	29%	15%	5%	5%	100%

On crée un tableau des réponses positives en faisant la somme des trois premières colonnes :

Pourcentage des réponses favorables	
Beauvais	77%
Bresles	67%
Clermont	71%
Creil	81%
Gouvieux	79%
TOTAL	75%

Et on commence à raisonner sur le dernier tableau sans se poser la question de la légitimité de cette démarche et des risques liés à la perte de données. Les résultats ici exposés sont réservés à des spécialistes.

1. Démonstration N°1

Question posée : Quel est le sens mathématique du taux attendu quand on mesure la qualité du fonctionnement d'un panel N d'institutions à l'aide d'une question ?

2. Notations utilisées :

On se donne une famille d'ensembles $\{G_i\}_{i \in [1, N]}$

$G_i \cap G_j = \emptyset$ si $i \neq j$ pour tout i et j éléments de $[1, N]$.

On définit $G = \bigcup_{i \in [1, N]} G_i$.

Soit H un ensemble de valeurs discrètes.

(Par exemple : $H = \{1, 2, 3, 4, 5, 6\}$).

Si $A \subset G$ et $B \subset G_i$ et on définit $P(A) = \text{card}(A) / \text{card}(G)$ et $P_{G_i}(B) = \text{card}(B) / \text{card}(G_i)$, où $\text{card}(E)$ désigne le nombre d'éléments de E .

Pour chaque G_i on suppose avoir une variable aléatoire :

$X_i: \omega \in G_i \longrightarrow X_i(\omega) \in H$

On définit la variable X de G dans H par :

si $\omega \in G_i$, $X(\omega) = X_i(\omega)$

On note :

$X_i = a$ l'ensemble des $\omega \in G_i$ tels que $X_i(\omega) = a$ et

$X = a$ l'ensemble des $\omega \in G$ tels que $X(\omega) = a$

La distribution F_i de X_i est alors définie par :

$F_i: a \in H \longrightarrow F_i(a) = P_{G_i}(X_i = a)$

On définit d'une façon analogue la distribution F de G par :

$F: a \in H \longrightarrow F(a) = P(X = a)$

On remarque que les F_i sont les distributions conditionnelles de F .

Par définition, F représente le taux attendu. Définition qui est justifiée par le lemme suivant. On remarquera au passage que le taux attendu est ici une fonction !

3. Lemme 1:

Alors la distribution F est la moyenne pondérée par $P(G_i)$ $i \in [1, N]$ des distributions $\{F_i\}_{i \in [1, N]}$

Il suffit de montrer que $\sum_{i \in [1, N]} P(G_i) \cdot F_i(a) = F(a)$ pour tout $a \in H$.

Les $\{G_i\}_{i \in [1, N]}$ forment une partition de G et les $\{X_i = a\}_{i \in [1, N]}$ une partition de $X=a$ d'où :

$P(G) = \sum_{i \in [1, N]} P(G_i)$

et

$$P(\{X = a\}) = \sum_{i \in [1, N]} P(\{X_i = a\})_{i \in [1, N]}$$

Preuve :

$$\begin{aligned} \sum_{i \in [1, N]} P(G_i) \cdot F_i(a) &= \sum_{i \in [1, N]} \frac{\text{card}(G_i)}{\text{card}(G)} \cdot P_{G_i}(X_i = a) = \\ &= \sum_{i \in [1, N]} \frac{\text{card}(G_i)}{\text{card}(G)} \cdot \frac{\text{card}(X_i = a)}{\text{card}(G_i)} = \sum_{i \in [1, N]} \frac{\text{card}(X_i = a)}{\text{card}(G)} \\ &= \sum_{i \in [1, N]} P(X_i = a) = P(X = a) = F(a) \end{aligned}$$

4. Démonstration N°2

Le but de cette démonstration est de regarder si la notion de taux attendu persiste quand on transforme les données en les regroupant (par exemple on rassemble les données positives en une seule valeur).

5. Préambule

Avec les notations ci-dessus

Soient deux valeurs distinctes, $a \in H$ et $b \in H$, on dit qu'on polarise a et b en c relativement à X_i (resp. X) quand on définit un ensemble T et une variable X'_i (resp. X') comme suit :

- i. Pour l'ensemble T sa définition est :
 - On définit H_o comme le complémentaire de $\{a, b\}$ dans H
 - Soit c un nombre, on définit T par $T = H_o \cup \{c\}$ (c remplace a et b)
- ii. Pour la variable X'_i (resp. X') leurs définitions sont : si $\omega \in T$ avec $X_i(\omega) \neq a$ et $X_i(\omega) \neq b$ alors $X'_i(\omega) = X_i(\omega)$, si non $X'_i(\omega) = c$. (respectivement pour X , où $X'(\omega) = X(\omega)$ sauf pour la valeur c où $X'(\omega) = c$).
- iii. De plus on suppose que l'ensemble $X'_i = c$ est la réunion des ensembles disjoints $X_i = a$ et $X_i = b$. C'est-à-dire : $X'_i = c = X_i = a \cup X_i = b$. (resp. pour X).

Comme $X_i = a$ et $X_i = b$, sont disjoints :

$$P(X'_i = c) = P(X = a) + P(X = b).$$

6. Remarque :

Pour tout i , soit K_i la distribution de X'_i et K celle de X' .

Alors : $K_i(c) = F_i(a) + F_i(b)$ et $K(c) = F(a) + F(b)$.

Preuve :

$$K_i(c) = P_{G_i}(X'_i = c) = P_{G_i}(X_i = a) + P_{G_i}(X_i = b) = F_i(a) + F_i(b) \text{ (resp. pour } X, K(c) = F(a) + F(b)).$$

On remarque aussi que si $x \notin \{a, b\}$, $K_i(x) = F_i(x)$

(resp. $K(x) = F(x)$).

7. Lemme 2 :

La polarisation des distributions du groupe G , reste la moyenne pondérée des polarisations de tous les groupes G_i

Il suffit de montrer pour la valeur c que :

$$\sum_{i \in [1, N]} P(G_i) \cdot K_i(c) = K(c)$$

En effet :

$$\sum_{i \in [1, N]} P(G_i) \cdot K_i(c) = \sum_{i \in [1, N]} P(G_i) \cdot (F_i(a) + F_i(b))$$

$$= \sum_{i \in [1, N]} P(G_i) \cdot F_i(a) + \sum_{i \in [1, N]} P(G_i) \cdot F_i(b) = F(a) + F(b) = K(c)$$

8. Démonstration N°3

Le but de cette démonstration est de définir la notion de profil en croisant deux variables et de vérifier que le profil global est la moyenne pondérée de chaque profil de base. Pour cela, il faut d'abord définir la notion de « case » liée à deux variables.

9. Préambule : tableaux croisés et taux attendus

On utilise les notations ci-dessous :

Soient X et Y deux variables définies sur G et à valeur dans H et leurs restrictions $\{X_i\}_{i \in [1, N]}$ et $\{Y_i\}_{i \in [1, N]}$ sur chaque G_i .

On définit les cases de $G \times G$ liées à X et Y par :

➤ Si $a \in H$ et $b \in H$ alors :

$$\text{case}(a, b) = \{ (x, y) \in G \times G \mid X(x) = a \text{ et } Y(y) = b \}$$

et pour tout i les cases de $G_i \times G_i$ liées à X_i et Y_i par :

➤ Si $a \in H$ et $b \in H$ alors :

$$\text{case}_i(a, b) = \{ (x, y) \in G_i \times G_i \mid X_i(x) = a \text{ et } Y_i(y) = b \}$$

10. Remarques :

Pour toute paire de couple $(a, b) \neq (c, d)$ de $H \times H$ on a :

$$\text{case}(a, b) \cap \text{case}(c, d) = \emptyset \text{ avec } a, b, c, d \text{ éléments de } H.$$

$$(\text{resp. } \text{case}_i(a, b) \cap \text{case}_i(c, d) = \emptyset).$$

On a aussi :

$$\bigcup_{(a, b) \in H \times H} \text{case}(a, b) = G \times G$$

$$(\text{resp. } \bigcup_{(a, b) \in H \times H} \text{case}_i(a, b) = G_i \times G_i),$$

ce qui permet de dire que l'ensemble des cases de $G \times G$ forment une partition de $G \times G$ (resp. des $G_i \times G_i$).

D'autre part, les $\text{case}_i(a, b)$ pour $i \in [1, N]$ forment une partition de la case (a, b) .

11. Lemme 3

La probabilité de la case (a, b) est la moyenne pondérée par $P(G_i)$ des $\text{case}_i(a, b)$ quand i parcourt $[1, N]$

Preuve :

Il faut montrer que :

$$P(\text{case}(a, b)) = \sum_{i \in [1, N]} P(G_i) \times P_{G_i}(\text{case}_i(a, b))$$

On a d'une part:

$$P(\text{case}(a, b)) = \frac{\text{card}(\text{case}(a, b))}{\text{card } G}$$

Et d'autre part

$$\begin{aligned} \sum_{i \in [1, N]} P(G_i) \times P_{G_i}(\text{case}_i(a, b)) &= \sum_{i \in [1, N]} \frac{\text{card}(G_i)}{\text{card}(G)} \times \frac{\text{card}(\text{case}_i(a, b))}{\text{card}(G_i)} = \\ &= \frac{1}{\text{card } G} \sum_{i \in [1, N]} \text{card}(\text{case}_i(a, b)) \end{aligned}$$

Comme les $\text{case}_i(a, b)$ pour $i \in [1, N]$ forment une partition de la case (a, b) .

$$\begin{aligned} &= \frac{1}{\text{card } G} \sum_{i \in [1, N]} \text{card}(\text{case}_i(a, b)) = \\ &= \frac{1}{\text{card}(G)} \text{card}(\bigcup_{i \in [1, N]} \text{case}_i(a, b)) = \\ &= \frac{1}{\text{card } G} \text{card}(\text{case}(a, b)) = \frac{\text{card}(\text{case}(a, b))}{\text{card } G} = \end{aligned}$$

$P(\text{case}(a,b))$

Ce qui termine la démonstration

12. Lemme 4 : polarisation de cases

On suppose qu'on regroupe un nombre S fini de cases différentes de $G \times G$ qui forment un ensemble R avec

$$R = \bigcup_{s \in S} \text{case}(s).$$

Les ensemble $R_i = R \cap G_i$ sont un ensemble de cases disjointes de $G_i \times G_i$ tel que pour tout $i \in [1, N]$:

$$R_i = \bigcup_{s \in S} \text{case}_i(s) \text{ et } \bigcup_{i \in [1, N]} \bigcup_{s \in S} \text{case}_i(s) = R$$

Alors on a :

$$P(R) = \sum_{i \in [1, N]} P(G_i) \cdot P_{G_i}(R_i)$$

Il suffit de décomposer R en cases qui sont forcément disjointes ainsi que chaque R_i et d'appliquer le lemme ci-dessus et les règles des probabilités sur la réunion d'ensembles disjoints.

Voici le calcul détaillé :

$$\begin{aligned} \sum_{i \in [1, N]} P(G_i) P_{G_i}(R_i) &= \sum_{i \in [1, N]} P(G_i) P_{G_i} \left(\bigcup_{s \in S} \text{case}_i(s) \right) = \\ &= \sum_{i \in [1, N]} P(G_i) \sum_{s \in S} P_{G_i}(\text{case}_i(s)) = \\ &= \sum_{s \in S} \sum_{i \in [1, N]} P(G_i) P_{G_i}(\text{case}_i(s)) \end{aligned}$$

D'après le lemme ci-dessus :

$$\sum_{i \in [1, N]} P(G_i) P_{G_i}(\text{case}_i(s)) = P(\text{case}(s))$$

D'où :

$$\begin{aligned} &\sum_{s \in S} \sum_{i \in [1, N]} P(G_i) P_{G_i}(\text{case}_i(s)) \\ &= \sum_{s \in S} P(\text{case}(s)) = P \left(\bigcup_{s \in S} \text{case}(s) \right) = P(R) \end{aligned}$$

13. Démonstration N°4

Le but de cette annexe est de montrer qu'il est dangereux de manipuler des résultats au niveau global même si au niveau local, tout semble se présenter d'une façon idéale. C'est pourquoi, un contre-exemple est présenté avant de montrer un résultat conforme à ce qu'on attend sur le plan humain.

Avec les notations ci-dessus on suppose que pour tout i , X_i et Y_i sont indépendants ce qui permet de dire qu'une décision prise sur les bases des résultats des indicateurs $\{X_i\}_{i \in N}$ n'impactera pas les résultats des indicateurs $\{Y_i\}_{i \in N}$.

Alors on se pose la question : quand est-il de l'indépendance de X et Y ?

Malheureusement, en général les indicateurs X et Y ne sont pas indépendants.

Pour prouver cela, il suffit de trouver un contre-exemple.

14. Contre-exemple

On suppose que $N = 2$ et que $H = \{a, b\}$, où a et b sont des réels distincts.

On construit deux tableaux croisés T_1 cases associées à G_1 et T_2 cases associées à G_2 qui représentent des variables X_1 et Y_1 (resp. X_2 et Y_2) indépendantes. Puis un troisième tableau T , réunion des deux premiers, de façon à avoir une représentation du tableau croisé de X et Y . Le but du contre-exemple est de montrer que le tableau croisé T , ne représente pas des variables de X et Y indépendantes.

Les trois tableaux suivants représentent le nom des croisements de X et Y , sur le groupe G_1 , G_2 et G .

	T_1			T_2	
$X_1 = 2$	case ₁ (1, 2)	case ₁ (2, 2)	$X_2 = 2$	case ₂ (1, 2)	case ₂ (2, 2)
$X_1 = 1$	case ₁ (1, 1)	case ₁ (2, 1)	$X_2 = 1$	case ₂ (1, 1)	case ₂ (2, 1)
	$Y_1 = 1$	$Y_1 = 2$		$Y_2 = 1$	$Y_2 = 2$

	T	
$X = 2$	case (1, 2)	case (2, 2)
$X = 1$	case (1, 1)	case (2, 1)
	$Y = 1$	$Y = 2$

On suppose que les valeurs prises par X_1 , X_2 et Y_1 , Y_2 sont telles que les cases se remplissent comme ci-dessous :

	T_1			T_2	
50	0	50	0	0	0
0	0	0	50	50	0
	0	50		50	0

Le tableau T se remplit alors de la façon suivante :

	T	
50	0	50
50	50	0
	50	50

On calcule les pourcentages par groupe :

	T_1			T_2	
100%	0%	100%	0%	0%	0%
0%	0%	0%	100%	100%	0%
	0%	100%		100%	0%

	T	
50%	0%	50%
50%	50%	0%
	50%	50%

X_1 et Y_1 (resp. X_2 et Y_2) sont indépendantes comme on le vérifie facilement. Qu'en est-il de X et Y ? Si c'était le cas chaque case du dernier tableau devraient être les produits des profils. Par exemple la case (1,1) devrait prendre la valeur $50\% \cdot 50\% = 25\%$, comme c'est fait dans le tableau théorique ci-dessous :

	T théorique	
50%	25%	25%
50%	25%	25%
	50%	50%

Ce qui est évidemment faux.

L'existence de contre-exemples est donc prouvée.

15. Définitions préalables :

On pose :

$$\alpha_i(a) = P_i(X_i = a) - P(X = a)$$

Et

$$\beta_i(b) = P_i(Y_i = b) - P(Y = b)$$

Les écarts entre la distribution conditionnelle et la distribution globale.

16. Lemme 5 :

Avec les notations ci-dessus on a :

$$\sum_{i \in [1, N]} P(\Omega_i) \alpha_i(a) = 0$$

C'est-à-dire la somme des écarts pondérés est nulle.

Preuve :

Remarquons que la famille $\{X_i = a\}_{i \in [1, N]}$ forme une partition de l'ensemble $X=a$ d'où :

On a :

$$P(X = a) = \sum_{i \in [1, N]} P(X_i = a) = \sum_{i \in [1, N]} \frac{\text{card}(X_i = a)}{\text{card } \Omega} = \sum_{i \in [1, N]} \frac{\text{card}(X_i = a)}{\text{card } \Omega_i} \frac{\text{card } \Omega_i}{\text{card } \Omega} = \sum_{i \in [1, N]} P(\Omega_i) P_i(X_i = a)$$

Or :

$$P_i(X_i = a) = P(X = a) + \alpha_i(a)$$

Donc :

$$\begin{aligned} P(X = a) &= \sum_{i \in [1, N]} P(\Omega_i) P(X = a) + \sum_{i \in [1, N]} P(\Omega_i) \alpha_i(a) = \\ P(X = a) \sum_{i \in [1, N]} P(\Omega_i) + \sum_{i \in [1, N]} \alpha_i(a) P(\Omega_i) &= \\ P(X = a) \cdot 1 + \sum_{i \in [1, N]} \alpha_i(a) P(\Omega_i) &= \\ P(X = a) + \sum_{i \in [1, N]} \alpha_i(a) P(\Omega_i) \end{aligned}$$

D'où :

$$\sum_{i \in [1, N]} \alpha_i(a) P(\Omega_i) = 0$$

17. Théorème de conservation « approximative » de l'indépendance

Avec les mêmes notations que ci-dessus, on suppose que pour tout i les variables X_i et Y_i sont indépendantes. Alors on a :

$$P(X = a \cap Y = b) =$$

$$P(X=a) P(Y = b) + \sum_{i \in [1, N]} \alpha_i(a) \beta_i(b) P(\Omega_i)$$

Démonstration :

Pour tout i on a :

$$P_i(X_i=a \cap Y_i=b) = P_i(X_i=a) P_i(Y_i=b)$$

On remarque que :

$$P_i(X_i=a) = P(X=a) + \alpha_i(a) \text{ et que } P_i(Y_i=b) = P(Y=b) + \beta_i(b)$$

D'où en développant

$$(i) \quad P_i(X_i=a \cap Y_i=b)$$

On obtient les quatre termes suivants :

$$(ii) \quad P_i(X_i=a) P_i(Y_i=b) = [P(X=a) + \alpha_i(a)] [P(Y=b) + \beta_i(b)]$$

$$= P(X=a) \cdot P(Y=b) + P(Y=b) \cdot \alpha_i(a) + P(X=a) \cdot \beta_i(b) + \alpha_i(a) \cdot \beta_i(b)$$

En multipliant les termes de la ligne ci-dessus (ii) par $P(\Omega_i)$ et en faisant la somme sur i on obtient les termes suivants :

Terme 1 :

$$\sum_{i \in [1, N]} P(\Omega_i) \cdot P(X = a) \cdot P(Y = b) = P(X=a) \cdot P(Y=b)$$

Terme 2

$$\sum_{i \in [1, N]} [P(Y = b) \cdot P(\Omega_i) \alpha_i(a)] =$$

$$P(Y = b) \sum_{i \in [1, N]} [P(\Omega_i) \cdot \alpha_i(a)] = 0$$

d'après le lemme précédent

Terme 3

$$\sum_{i \in [1, N]} [P(X = a) \cdot P(\Omega_i) \cdot \beta_i(b)] = 0$$

Enfin Terme 4 :

$$\sum_{i \in [1, N]} \alpha_i(a) \cdot \beta_i(b) \cdot P(\Omega_i)$$

D'où le résultat :

$$\sum_{i \in [1, N]} P(\Omega_i) \cdot P_i(X_i = a \cap Y_i = b) =$$

$$P(X=a) \cdot P(Y=b) + \sum_{i \in [1, N]} \alpha_i(a) \cdot \beta_i(b) P(\Omega_i)$$

D'autre part en multipliant le terme (i) par $P(\Omega_i)$ et en faisant la somme sur i on obtient

$$\begin{aligned} \sum_{i \in [1, N]} P_i(X_i = a \cap Y_i = b) \cdot P(\Omega_i) &= \sum_{i \in [1, N]} \frac{\text{card}(X_i=a \cap Y_i=b)}{\text{card}(\Omega_i)} \cdot \frac{\text{card}(\Omega_i)}{\text{card}(\Omega)} = \\ \frac{1}{\text{card}(\Omega)} \sum_{i \in [1, N]} \text{card}(X_i = a \cap Y_i = b) &= \frac{\text{card}(X = a \cap Y = b)}{\text{card}(\Omega)} = P(X = a \cap Y = b) \end{aligned}$$

D'où le résultat cherché :

$$\begin{aligned} P(X = a \cap Y = b) &= \\ P(X=a) P(Y = b) + \sum_{i \in [1, N]} \alpha_i(a) \cdot \beta_i(b) P(\Omega_i) \end{aligned}$$

18. Remarque

Dans la pratique on sélectionne les questions (variables) dont les distributions conditionnelles $F_{i \in [1, N]}$ sont proches de la distribution globale F . De ce fait, les termes $|\alpha_i(a)|$ et $|\beta_i(b)|$ sont

généralement inférieurs à 10%, ce qui entraîne que le terme $|\sum_{i \in [1, N]} \alpha_i(a) \beta_i(b) P(\Omega_i)|$ est inférieur à 1% car :

$$\left| \sum_{i \in [1, N]} \alpha_i(a) \beta_i(b) P(\Omega_i) \right| \leq \sum_{i \in [1, N]} |\alpha_i(a) \beta_i(b)| P(\Omega_i) \leq \sum_{i \in [1, N]} \frac{1}{100} P(\Omega_i) \leq \frac{1}{100}$$

Les apports de ce papier justifient au moins partiellement beaucoup de pratiques de l'administration. Cependant, il faut rester prudent comme le montrent le contre-exemple ci-dessus.

D'autres questions peuvent se poser qui demanderaient plus de développement. La plus pertinente est la question sur les corrélations entre variables. Question similaire à la question sur la conservation de l'indépendance. Je laisse à ceux qui désireraient continuer ce travail le soin de le développer.

Quatrième de couverture essai-évaluation

Savoir ce qui est vrai ou faux et prendre des décisions a toujours été très difficile. Aujourd'hui, où l'intelligence artificielle est capable de générer parfaitement de fausses informations et de fausses images, la situation est pire. L'estimation du Réel est un challenge bien plus difficile qu'il y a un siècle. Dans l'approche du Réel il y a deux voies qui s'imposent :

- La mesure des réalités concrètes et objectives
- L'évaluation des besoins des groupes humains

La première voie c'est celle de la science, basée sur la répétition de la mesure des événements sous conditions identiques.

La seconde c'est l'évaluation du fonctionnement des groupes sous contrainte à travers la prise en compte des opinions des différents acteurs.

Cet essai porte sur la compréhension et la légitimité de la seconde voie.

