

Durand, C. (2021). La Mesure, dans I. Bourgeois et B. Gauthier, Recherche sociale, de la problématique à la collecte des données. PUQ, p. 179-200.

CHAPITRE 9 /

## La mesure

*Claire Durand*

L'action de mesure se situe à la jonction des deux grandes étapes de la recherche, la **formulation de la question de recherche** et la **réponse à cette question**. Ces deux étapes possèdent leur langage propre. Le langage utilisé à l'étape de la formulation de la question est essentiellement **abstrait**. Le chercheur avance des propositions, émet des hypothèses, précise ce qu'il cherche relativement à des relations entre des concepts : il soutient, par exemple, qu'il y a une crise de confiance dans les institutions dans les sociétés démocratiques. Le langage de la vérification est **concret** et se fonde sur l'observation empirique des phénomènes : on peut par exemple examiner sur une longue période de temps les résultats des sondages qui ont porté sur la confiance institutionnelle dans divers pays du monde ayant des régimes politiques différents. On peut alors examiner s'il y a une diminution de la confiance et si celle-ci évolue de façon différente selon les régimes politiques. Les deux langages sont indispensables à l'ensemble de l'opération. La recherche n'est possible que si l'on peut formuler des hypothèses sur la réalité, hypothèses que l'on confronte à l'information recueillie. Ce processus permet, en interprétant les résultats de la confrontation, d'arriver à des conclusions sur la réalité.

Le problème de départ est celui du passage du langage de l'abstraction, qui prévaut dans la formulation de questions et d'hypothèses de recherche, à celui de l'observation ou de la mesure, qui s'impose au moment où on tente d'apporter une réponse à la question, de vérifier les hypothèses avancées. Le problème est similaire à celui qui se pose lorsqu'on veut exprimer une idée, un sentiment. On cherche à utiliser les

mots, les formulations qui traduisent le mieux ce que l'on pense ou ressent. Cette opération est d'autant plus difficile si l'on tente de s'exprimer dans une autre langue que sa langue maternelle. C'est un peu la situation que l'on rencontre lorsqu'on passe du langage de l'abstraction au langage empirique. D'une part, plus clair est ce que l'on pense ou ressent, plus il est facile de le traduire en mots. D'autre part, plus solide est notre connaissance de la langue seconde, la langue empirique, plus il nous est facile de nous exprimer. Toutefois, la langue empirique demeurera toujours une langue seconde.

Ainsi en est-il dans le domaine de la recherche. Il s'agit de trouver pour chacun des concepts retenus un ou plusieurs équivalents empiriques qui constituent la traduction des constructions abstraites de l'esprit dans le langage de l'observation, tout en sachant qu'il n'existe pas de solution parfaite.

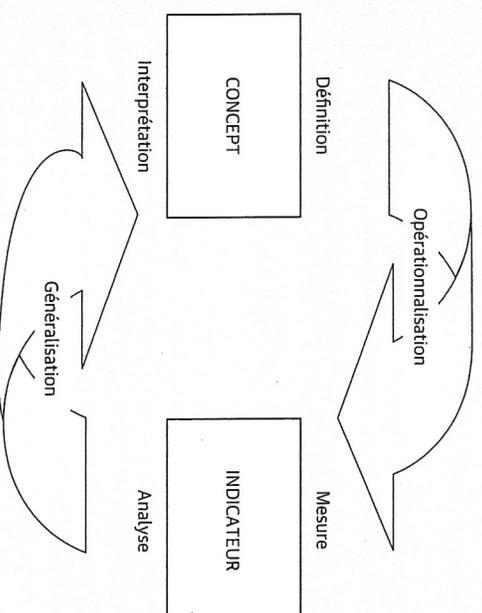
Prenons, par exemple, le concept de participation politique. Un chercheur peut s'intéresser aux facteurs qui influencent la participation et énoncer un certain nombre d'hypothèses à cet égard. La participation politique est une construction de l'esprit à laquelle le chercheur fait appel pour comprendre le réel. Il s'agit d'une abstraction : elle ne se voit pas, ne s'entend pas, ne se sent pas, ne se touche pas. Par contre, le chercheur a accès à des phénomènes qu'il peut interpréter comme étant des signes empiriques de la participation telle qu'il la conçoit et la définit. Un individu vote ou ne vote pas, assiste ou non à une assemblée publique, prend part ou non à une manifestation. Ce sont autant d'indicateurs possibles du concept de participation.

On peut également regarder comment Statistique Canada traduit un concept comme le chômage. Trois indicateurs sont utilisés : le fait d'être sans emploi, le fait d'être disponible pour travailler et le fait de chercher un emploi. Ainsi, un chômeur est défini non pas simplement comme une personne sans emploi, mais également comme une personne qui peut et qui veut travailler.

Le passage de la théorie à la vérification exige que l'on mesure des **concepts** au moyen d'**indicateurs**, ce qui demande d'établir un pont entre l'univers de l'abstraction et celui de l'observation et de la mesure (figure 9.1). Les conclusions d'une recherche dépendent étroitement des décisions qui sont prises à l'étape du choix des indicateurs. Ces conclusions tiennent seulement si les données empiriques – les

indicateurs – reflètent adéquatement les constructions théoriques proposées – les concepts. La sélection des indicateurs représente donc une opération lourde de conséquences.

FIGURE 9.1 / Les deux mouvements de traduction associés à l'opération de mesure



### 1 / Du concept à l'indicateur et à la variable

La question de recherche peut être centrée sur un seul concept dont on étudie l'évolution dans le temps ou la variation dans l'espace. Il en est ainsi des études qui tentent de vérifier si la criminalité, la participation politique, la consommation des médias ou la confiance dans les institutions augmentent ou diminuent dans le temps ou sont plus élevées dans une société que dans une autre.

La question peut aussi porter sur les liens possibles entre un certain nombre de concepts. On tentera de vérifier, par exemple, si l'augmentation du chômage est liée à une aggravation de la criminalité, si la confiance dans les institutions est liée à l'idéologie du parti au pouvoir ou de celle de la personne répondant à l'enquête, ou encore si la participation politique

est liée au niveau de scolarité. Les hypothèses peuvent être plus complexes, mais elles renvoient toujours à des concepts et on a besoin de signes concrets pour pouvoir les confirmer ou les invalider.

La mesure est le processus par lequel on passe des concepts abstraits à des indicateurs concrets. Elle est définie comme l'ensemble des opérations effectuées à l'aide d'un ou de plusieurs instruments de mise en forme de l'information qui permettent de classer un objet dans une catégorie pour une caractéristique donnée. On utilise aussi le terme *mesure* pour désigner le résultat de l'opération. On dira que l'exercice du droit de vote, la présence à une manifestation ou à une assemblée publique d'un parti politique sont des mesures - ou des indicateurs - de la participation politique. Ce processus n'est pas réservé aux individus; on pourrait également classer les pays en fonction du degré de participation électorale, de l'adhésion de la population à la démocratie ou de son niveau de confiance moyen envers les institutions.

Nous avons parlé jusqu'ici exclusivement de « concept » et d'« indicateur ». Le terme **variable** est aussi couramment utilisé en recherche, surtout dans les analyses quantitatives. Il s'emploie par opposition au terme *constante* pour indiquer **toute caractéristique susceptible de varier entre les unités**. Le concept est associé à une théorie et la variable, à une hypothèse. On parlera d'« hypothèses portant sur les relations entre des variables ». C'est ici qu'interviennent les notions de « **variable indépendante** » et de « **variable dépendante** ». Prenons l'hypothèse selon laquelle la participation politique est associée au niveau de scolarité. Dans cette situation, le niveau de scolarité est la variable indépendante parce qu'on suppose qu'il est **antérieur** à la participation politique, et cette dernière est considérée comme la variable dépendante parce qu'on suppose qu'elle découle du niveau de scolarité. Dans cet exemple, les variables sont constituées à partir d'un seul indicateur. Une variable peut également être construite en combinant des indicateurs au moyen de diverses opérations logiques ou mathématiques. Une variable peut être le résultat d'une somme ou d'une moyenne d'indicateurs, par exemple.

## 1 / L'objectif : la classification

L'objectif poursuivi dans le processus de mesure est de classer les unités par rapport à un concept donné tels le degré de participation politique, l'occupation, l'éthnicité, l'opinion sur l'avortement, la satisfaction au travail sur le plan individuel ou, pour ce qui est d'un pays, le degré de

démocratie. On peut difficilement mesurer la participation politique si l'on n'a pas une conception précise de ce qu'elle comprend et ne comprend pas. Une définition claire des concepts utilisés est essentielle pour choisir les indicateurs appropriés. Cette opération n'est pas aussi facile que l'on pourrait le penser.

D'une part, **un concept peut avoir plusieurs significations**. Il en est ainsi pour la plupart des concepts liés à la nationalité ou à l'éthnicité. Le concept de « Québécois », par exemple, selon le contexte dans lequel il est utilisé, désigne soit l'ensemble des citoyens canadiens résidant au Québec, soit seulement les Québécois considérés d'« origine » issus de plusieurs générations de Québécois francophones. Cette illustration vaut pour la plupart des nationalités. À qui réfère-t-on quand on parle d'un Français, d'un Belge, d'un Sénégalais ?

D'autre part, **un concept peut être difficile à définir**. Ainsi en est-il du concept d'« Autochtone ». Certaines définitions incluent toute personne ayant des origines autochtones, alors que d'autres restreignent la définition aux individus qui s'identifient comme tels ou même exclusivement aux personnes résidant dans des communautés autochtones. Des catégories comme « Noir » ou « Blanc » font référence à la couleur de la peau, et des catégories comme « Francophone » ou « Anglophone », à la langue. Ce sont des regroupements très hétérogènes sur le plan des origines nationales, de la couleur de la peau, de la ou des langues maternelles et parlées à la maison. L'augmentation de la diversité sociale est concurrencée par une forte augmentation des personnes dont les parents sont d'origines différentes. L'origine ethnique n'est donc pas nécessairement unique. La plupart des recensements permettent maintenant plusieurs identifications pour répondre à cette évolution de la société.

Enfin, **l'évolution de la société peut entraîner des redéfinitions** qui exigent de modifier les indicateurs utilisés. La redéfinition légale du mariage dans plusieurs pays a amené leurs recensements à permettre de se déclarer « marié à un conjoint de même sexe », ce qui était impossible jusqu'à tout récemment. Les discussions sont en cours sur la définition même de « sexe ». Devrait-on permettre une catégorie « autre » ? Devrait-on préciser les caractéristiques des personnes qui peuvent se déclarer de sexe « autre » ?

L'indicateur mesure un seul aspect d'un concept. Pour mesurer un concept, il est souvent nécessaire de recourir à plusieurs indicateurs associées à divers aspects du concept à mesurer. L'éthnicité, par exemple, peut

être mesurée par une combinaison d'indicateurs relatifs au pays de naissance, à la langue maternelle et parfois aussi à la religion, cela pour l'individu enquêté qui fournit la réponse et pour ses parents.

L'indicateur ou les indicateurs visent à placer les objets d'étude dans des catégories qui correspondent aux différentes situations où ils peuvent se retrouver par rapport à la caractéristique retenue. Si l'on s'intéresse à la participation politique, par exemple, on pourra distinguer des degrés de participation élevés, modérés, faibles ou inexistant. La classification demande de construire des catégories et de déterminer les critères d'assignation à chaque catégorie. L'objectif est d'inclure chaque unité - individu, groupe d'individus, municipalité, pays, etc. - dans l'une des «boîtes» ainsi constituées, de déterminer dans quelle catégorie - ou modalité - telle unité doit être placée par rapport à un indicateur.

À titre d'exemple, on peut définir le groupe ethnique comme un ensemble d'individus qui partagent des caractéristiques communes se rapportant à leur lieu de naissance, à celui de chacun de leurs parents ainsi qu'à leur langue ou même à leur religion dans certains cas. Les critères de classification sont basés sur des indicateurs tels que le lieu de naissance, ceux du père et de la mère, la langue maternelle principale (définie comme la première langue parlée et encore comprise), la langue d'usage (définie comme la langue parlée le plus souvent à la maison). Les cas pour lesquels les lieux de naissance de la personne et de chacun de ses parents sont semblables ne posent pas problème. Par contre, lorsque ces indicateurs sont différents, il faut recourir à des critères de classement supplémentaires telle la langue maternelle ou d'usage. Ainsi, les personnes nées en Asie dont au moins un des parents est né en Asie, celles qui sont nées au Québec dont les deux parents sont nés en Asie ainsi que celles qui sont nées au Québec qui déclarent comme langue maternelle ou comme langue d'usage une langue asiatique sont classées dans le groupe asiatique. Cette attribution exclut du groupe asiatique une personne née en Asie adoptée par des parents québécois.

Cet exemple illustre le processus d'opérationnalisation du concept. Plusieurs indicateurs peuvent être nécessaires pour classer les personnes relativement à un concept. La fonction des indicateurs est de permettre de traduire le mieux possible le concept que l'on désire mesurer dans l'univers empirique.

On distingue trois types de catégorisation :

- Le premier est la **catégorisation nominale**, où les catégories sont simplement juxtaposées les unes aux autres. Cette opération permet de différencier les individus en fonction de critères « qualitatifs », sans qu'il y ait un ordre entre les catégories. Le sexe, le lieu de naissance et la langue maternelle en sont des exemples.
- Le deuxième type est la **catégorisation ordinale**, dans laquelle les catégories sont ordonnées, hiérarchisées les unes par rapport aux autres. Le degré de participation politique d'un individu ou de démocratie d'un pays appartient à ce type. L'information est plus riche. On peut non seulement distinguer les unités les unes des autres, mais il est aussi possible de les ranger des moins participatifs aux plus participatifs dans le cas des individus, des plus autoritaires aux plus démocratiques dans le cas des pays. La mesure des attitudes utilise fréquemment des échelles de ce type. À titre d'exemple, on mesurera le degré de satisfaction à l'aide d'une échelle de réponse du type « très satisfait, assez satisfait, peu satisfait, pas du tout satisfait » appliquée à un certain nombre d'éléments. Dans ce genre de catégorisation, on sait qu'une personne très satisfaite l'est plus qu'une personne assez satisfaite, mais on ne sait pas jusqu'à quel point. L'ordre entre les catégories est connu, mais non la distance entre elles.
- Le troisième type de catégorisation, la **catégorisation numérique**, est encore plus précis. L'âge, mesuré en années, le revenu, exprimé dans une unité monétaire, et le taux de chômage, formulé en proportion du nombre de personnes au chômage dans la population active, en sont des exemples. Dans ces cas, on peut non seulement ranger les objets d'étude les uns par rapport aux autres, mais on peut également connaître exactement l'écart qui les sépare les uns des autres. On pourra dire d'un individu que son revenu est deux fois plus élevé que celui d'un autre, ou d'un pays que son taux de chômage est de deux points de pourcentage inférieur à celui d'un autre pays. Ce type de catégorisation requiert une unité standardisée (une monnaie de référence comme l'euro ou le dollar américain) ou comparable (une entente sur le concept et sa mesure dans le cas du taux de chômage).

Le chercheur a intérêt à faire appel à la catégorisation la plus riche, qui se prête à des traitements statistiques plus affinés et variés. Lorsqu'on veut classer les individus selon l'âge, il est préférable de poser une question

sur l'année de naissance (de façon à ce que l'âge soit établi en nombre d'années) plutôt que de se fier à des catégories moins précises (18 à 24 ans, 25 à 34 ans, etc.) que l'on ne peut pas modifier. Il est d'autant plus important de procéder ainsi lorsqu'on veut effectuer des analyses de tendance et qu'on doit être en mesure de modifier les catégories d'âge pour tenir compte de l'évolution dans le temps. C'est la nature des concepts retenus qui dicte le degré de catégorisation. Certains concepts sont nécessairement qualitatifs et ne se prêtent pas à la catégorisation numérique ou même ordinale. Il s'agit en somme de choisir le type de catégorisation qui convient au concept que l'on veut mesurer.

Toutes les classifications ont en commun que les catégories utilisées devraient être collectivement **exhaustives** et mutuellement **exclusives**.

- Cela signifie que **tout objet d'étude doit pouvoir être placé dans une catégorie et que la liste des possibilités est complète**. Si l'on s'intéresse à l'occupation, par exemple, les catégories doivent faire référence à toutes les occupations possibles, comme occuper un emploi, être aux études, au chômage, à la retraite, tenir maison ou être invalide.

- Il faut également que **tout objet puisse être assigné à une seule catégorie et qu'il n'y ait aucun recoupement possible entre les catégories**. À titre d'exemple, certaines personnes occupent un emploi tout en étant aux études, de sorte qu'elles pourraient être classées dans deux catégories. Pour contourner cette difficulté, on utilisera le concept d'«occupation principale», ce qui permet de classer chaque individu dans une seule catégorie. On pourra également ajouter une question supplémentaire sur l'occupation secondaire. Dans d'autres cas, comme la langue maternelle, la principale langue parlée à la maison ou l'origine ethnique, il est indispensable de prévoir la possibilité de réponses multiples.

### / Le moyen : les instruments et les opérations

Comme l'objectif de la mesure est de placer chaque objet étudié dans une catégorie, il faut déterminer des règles précises d'assignation aux diverses catégories. À partir de quel critère désignera-t-on une personne comme ayant une participation politique élevée ou modérée? En fonction de quelle information - le prix de quels biens - dira-t-on que le taux d'inflation a été de 1,4% au cours des 12 derniers mois? À partir de combien

de cours suivis dira-t-on qu'une personne étudie à temps plein? La classification demande des règles qui sont traduites dans un ensemble d'opérations empiriques.

Revenons à l'exemple du chômage. À quoi reconnait-on qu'un individu est au chômage? Selon la définition que Statistique Canada utilise - comme beaucoup d'organismes statistiques nationaux -, un individu qui, au moment de l'enquête mensuelle, affirme qu'il était sans travail au cours de la semaine précédente, qu'il n'y avait aucune raison qui l'empêchait d'occuper un emploi et qu'il a cherché du travail au cours des quatre dernières semaines<sup>1</sup> est considéré en chômage. En somme, Statistique Canada mesure le chômage à partir d'un ensemble de réponses données à un questionnaire administré à un échantillon de personnes qui seront classées dans cette catégorie selon les réponses qu'elles ont fournies. Ces questions découlent de la conception que l'on se fait du chômage: il ne suffit pas de savoir si une personne est sans emploi. On doit aussi déterminer si elle «veut» travailler, ce qui amène à relever des signes de «bonne volonté». On doit également pouvoir classer les personnes qui occupent un emploi à temps partiel, mais qui désirent travailler à temps plein. C'est à l'aide de telles opérations que l'on classe les objets d'étude dans des catégories.

La construction des indicateurs fait appel à des **instruments de mise en forme** de l'information. Ces instruments incluent le **questionnaire**, mais également **l'observation directe** et **l'analyse de divers types de contenus** - entrevues, discours, articles de journaux, contenus de médias sociaux, contenu visuel, données institutionnelles chiffrées ou textuelles, etc. Ces instruments sont fondés sur l'observation de comportements, de documents ou de réponses à des questions. La construction des indicateurs exige de choisir d'abord un de ces instruments ou une combinaison d'entre eux, mais cela n'est pas suffisant. L'opérationnalisation renvoie également au «mode d'emploi» de l'instrument, à la grille d'observation ou d'analyse, à la formulation des questions ainsi qu'à l'interprétation des résultats obtenus.

1 STATISTIQUE CANADA (2020). Enquête sur la population active. <http://www23.statcan.gc.ca/fmhd/p25v\_f.pdf?function=getSurveys&SDS=3701>, consulté le 6 octobre 2020.

On parle d'opérationnalisation ou de mesure pour faire référence au processus qui permet de traduire un concept en indicateurs empiriques. Toute opérationnalisation doit permettre de reproduire de façon exacte les opérations effectuées de façon à ce qu'un autre chercheur puisse les vérifier et les répéter. Les instruments utilisés pour procéder à la classification et les opérations qui sont effectuées doivent donc être précises. Il ne suffit pas de dire que les données proviennent d'un sondage, il faut indiquer la procédure d'échantillonnage qui a été utilisée, le mode d'administration du questionnaire, la période où les données ont été collectées et les résultats de cette collecte. De la même manière, une analyse de contenu demande de préciser les concepts utilisés, leur définition et leur opérationnalisation et une observation directe, la grille d'observation. Il faut informer de façon transparente de toutes les opérations faites pour passer du concept à la mesure.

Si chaque chercheur créait son propre thermomètre et sa propre échelle pour mesurer la température, il deviendrait extrêmement difficile de parler de température. C'est pourtant une situation fréquente en sciences sociales, où les mêmes concepts sont souvent mesurés de manière différente. Il est donc essentiel que la procédure suivie soit précisée.

Cette exigence découle de la conception que la communauté scientifique se fait de la connaissance. Toute recherche procédée par découpage et ne peut éclairer, dans le meilleur des cas, qu'une partie de la réalité à un moment donné. De plus, comme les risques d'erreurs sont importants, les chercheurs ne peuvent avoir une confiance totale dans les résultats que s'ils sont corroborés par d'autres études. D'où la nécessité de répéter une recherche pour en vérifier les conclusions. Cela exige un échange optimal d'informations entre les chercheurs, de façon à distinguer les résultats les plus fiables et à contribuer ainsi à l'accumulation des connaissances ; c'est le principe de la **transparence**. Si la procédure employée par le chercheur n'est pas suffisamment bien rapportée pour qu'on puisse la reproduire, il y a une lacune sérieuse qui amène à mettre en doute les conclusions de la recherche.

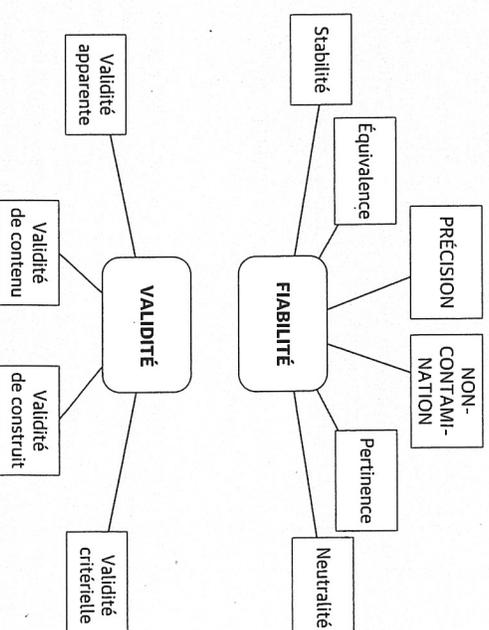
En résumé, la première étape permet de définir le concept que l'on veut mesurer de façon précise. La deuxième étape consiste à élaborer ou à sélectionner un ou des indicateurs qui permettront de construire une mesure susceptible de traduire le concept d'intérêt dans le langage empirique. Pour que la valeur de ce processus puisse être appréciée, certains critères doivent être respectés.

## 2 / Les critères d'appréciation

L'opération de mesure se fait par approximation, de sorte qu'on peut difficilement se prononcer de façon définitive sur la qualité d'une mesure. Une telle indétermination ne signifie pas que la sélection des indicateurs relève de l'arbitraire. La communauté scientifique s'est entendue sur des critères d'appréciation. Chacun peut être considéré comme une condition nécessaire, mais non suffisante. C'est seulement s'ils semblent tous respectés que l'on pourra conclure que l'indicateur apparaît satisfaisant.

Les critères généraux d'appréciation sont la **fiabilité** et la **validité**. La fiabilité - ou *fidélité* dans certains textes - a trait à la qualité de la mesure elle-même, alors que la validité porte sur la qualité de la traduction du concept en mesure empirique (figure 9.2).

FIGURE 9.2 / L'appréciation des indicateurs : concepts centraux et concepts associés



### La fiabilité

**La fiabilité est un indice de la qualité de la mesure « en soi ».** Le principe est simple : l'indicateur doit mesurer une caractéristique précise d'un objet et rien d'autre. De plus, chaque mesure faite à partir des mêmes opérations devrait donner un résultat identique, pour autant que l'objet soit similaire ou demeure inchangé.

La **précision** de l'instrument demande la **stabilité** et l'**équivalence** de la mesure.

Prenons l'exemple de la satisfaction au travail. Pour la mesurer, on utilisera habituellement une série d'énoncés pour lesquels on demandera aux répondants d'indiquer leur degré d'accord. On dira de cet instrument qu'il est fiable si, en l'absence de changements majeurs, le niveau de satisfaction mesuré chez les mêmes personnes dans une organisation est constant. On parlera de **stabilité** de la mesure dans le temps si deux mesures prises à des moments différents sont similaires. La constance s'applique aussi en comparant les instruments : on parlera alors d'**équivalence**. Deux séries d'énoncés différents visant à mesurer la satisfaction au travail peuvent être vues comme des échantillons de l'ensemble des énoncés qui pourraient être utilisés. Des séries d'énoncés qui ont une même visée doivent donner des résultats similaires, sinon on ne peut pas vraiment s'y fier. Aucune des opérations ne permet de savoir si les énoncés mesurent vraiment la satisfaction au travail ; on sait seulement qu'ils mesurent « quelque chose » de façon constante lorsque certaines conditions sont respectées.

Un indicateur précis est compris par tous et compris de la même manière par tous. En ce qui concerne le questionnaire ou l'entrevue, la précision demande d'utiliser un niveau de langage approprié au sujet et des termes qui ont la même signification pour tous. Pour ce qui est de la construction d'une grille d'observation, il faut définir très précisément ce qui doit être observé de manière à ce que deux personnes qui observent la même situation arrivent à des résultats similaires. Dans l'analyse de contenus, les codes et leur signification doivent être définis de façon à ne pas laisser aux juges trop de latitude dans l'interprétation.

L'indicateur informe sur l'état où se trouve un objet d'étude par rapport à une caractéristique donnée à l'aide d'instruments de mesure. Or ces instruments introduisent une dimension nouvelle. L'interviewer et

son questionnaire, l'individu qui observe et sa grille d'observation ne font pas partie de la vie quotidienne. Leur seule présence peut entraîner des effets particuliers qui sont distincts de ceux que l'on veut mesurer. On parle alors de « **contamination de l'instrument** ». Ce qui est observé est différent de ce qui l'aurait été en l'absence de l'instrument.

Le cas classique de contamination est une étude effectuée dans les années 1930 pour mesurer l'effet des conditions de travail sur la productivité du personnel d'une usine<sup>2</sup>. Pendant plus d'un an, les chercheurs ont modifié l'horaire, les pauses, l'éclairage et le système de rémunération d'un petit groupe. Peu importe les conditions, ce petit groupe s'est avéré plus productif que les autres membres du personnel de l'usine. Les chercheurs ont constaté que leur productivité était influencée par le fait qu'ils se savaient observés. L'observation créait un milieu artificiel dans lequel ils se comportaient de façon « anormale ».

Les effets de contamination sont difficiles à mesurer, de sorte qu'il est difficile d'en estimer les conséquences. Deux stratégies s'offrent au chercheur qui veut minimiser la contamination :

- La première stratégie consiste à choisir l'instrument le moins contaminant possible. Sur ce plan, l'analyse de contenu présente un grand intérêt, le document étudié n'étant pas influencé par la présence de l'analyste. Il en est ainsi de certaines formes d'observation « non visibles ». Les sondages auto-administrés sont également moins sujets à contamination que les entrevues. Cette stratégie est cependant d'une application limitée. La sélection d'un instrument est d'abord dictée par d'autres considérations fondamentales comme la question de recherche, la nature des concepts à mesurer et l'importance de pouvoir généraliser les résultats à l'ensemble d'une population.
- La deuxième stratégie consiste à limiter la contamination en prêtant attention à l'opérationnalisation elle-même. Si on utilise l'entrevue, qu'elle soit ou non standardisée, on doit former les personnes qui mènent les interviews pour qu'elles n'influencent pas les réponses. Elles ne doivent pas donner l'impression, par des attitudes, des remarques, etc., qu'une réponse serait préférable à

<sup>2</sup> ROETHLISBERGER, F. J. et W. J. DICKSON (1939). *Management and the Worker*. Harvard University Press.

une autre. Si le chercheur participe à la vie d'un groupe, il ne procédera à la collecte de données qu'après une période d'intégration au groupe, au moment où sa présence est devenue moins perceptible et ne modifiera plus le comportement habituel.

Dans la même perspective, on doit faire la collecte d'informations au cours de périodes « normales », habituelles, et non dans des circonstances exceptionnelles, à moins que l'enquête porte précisément sur la situation exceptionnelle. Ainsi, une recherche menée pendant la pandémie de COVID-19 ne permet pas de conclure sur l'état de santé de la population ou le taux de chômage « en temps normal », mais bien durant la pandémie. On peut aussi penser que la température qui a cours - vague de froid, canicule - au moment d'une enquête sur la perception du réchauffement climatique peut influencer les conclusions de l'enquête.

La volonté de ne pas contaminer la mesure incite à s'assurer de la **pertinence** des questions posées. Poser des questions d'opinion à des personnes qui ne connaissent rien au sujet sur lequel on demande leur avis ou qui n'y ont jamais réfléchi entraîne des réponses peu fiables. Une opinion non informée est plus sensible à la manière dont la question est posée et au moment où elle est posée. Elle peut aussi entraîner la non-réponse ou même l'abandon du sondage.

La volonté de ne pas contaminer la mesure incite également à s'assurer de la **neutralité** des questions. Une question qui laisse entendre qu'un point de vue est plus acceptable qu'un autre biaisera les résultats et produira un indicateur non fiable. Une grille d'observation qui tiendrait compte uniquement des phénomènes qui confirment les hypothèses de recherche ou la codification d'un discours qui garderait seulement ce qui est conforme à ces hypothèses entacherait inévitablement la fiabilité des mesures, puisqu'un autre analyste qui répèterait la même opération n'arriverait pas au même résultat. De même, une question de sondage qui proposerait des modalités non équilibrées - en privilégiant par exemple plus de choix positifs que négatifs - ou qui tenterait de privilégier un point de vue en incitant à la pitié ou à la solidarité avec les personnes du même groupe social ne satisfait pas au critère de neutralité.

Le critère de fiabilité est fondamental. Quand un indicateur n'est pas fiable, on ne sait plus trop ce qu'il mesure. Il est rare qu'un indicateur donne des résultats parfaitement constants, ce qui oblige à accepter une certaine dose d'« infidélité ». Toutefois, si un minimum de constance et d'équivalence ne peut être atteint, il est préférable de rejeter l'indicateur.

2.2 /

## La validité

Un indicateur est **valide** lorsqu'il **représente adéquatement** un concept. Un indicateur ne peut être valide que s'il est fiable et donc précis et non contaminé. Mais ces conditions ne sont pas suffisantes. D'autres distortions peuvent se produire au moment de la traduction du concept.

Supposons qu'une chargée de recherche désire mesurer le niveau de bonheur des gens, qu'elle choisit comme indicateur la fréquence du sourire dans une population donnée et qu'elle décide d'utiliser l'observation directe pour recueillir l'information. Elle pourrait élaborer une grille d'observation précise, en tester la fiabilité et s'assurer que les personnes responsables des observations seront pratiquement invisibles. Même si toutes ces conditions sont satisfaites, elle ne pourra pas prétendre hors de tout doute qu'il s'agit d'un indicateur valide du bonheur : le sourire fait à un moment donné, dans des circonstances précises, est-il un signe de bonheur ? Par contre, si elle choisit plutôt de demander aux gens à quel point ils se sentent heureux, elle devra présupposer que leurs réponses reflètent véritablement leur sentiment.

Le passage du concept à l'indicateur demande de respecter certains critères. On parlera de **validité apparente** lorsque l'indicateur apparaît valide « à sa face même », de **validité de contenu** lorsque l'ensemble du « domaine » du concept mesuré est couvert, de **validité de construit** lorsque le concept est mesuré de façon appropriée et enfin, de **validité liée au critère** lorsque la mesure d'un concept est associée à d'autres mesures auxquelles le concept est censé être lié.

La **validité apparente** est la validité conceptuelle de la mesure. Un indicateur doit paraître valide comme la « justice doit paraître avoir été rendue ». Le fait de se promener dans un magasin le matin n'apparaît pas comme un bon indicateur de chômage, puisque les horaires de travail ne sont pas les mêmes pour tous et que bien des raisons peuvent expliquer cette présence, y compris le fait d'être employé dans les environs et d'être en pause syndicale.

La **validité de contenu** est satisfaisante lorsque tous les aspects du concept que l'on désire mesurer sont couverts. Ainsi, une mesure de la satisfaction au travail doit couvrir l'ensemble des aspects du travail, intrinsèques, extrinsèques et sociaux. Une mesure de la participation politique doit inclure plusieurs types d'activités politiques et non

seulement la participation au vote. Une excellente connaissance du sujet est un préalable incontournable pour en arriver à construire un instrument ayant une bonne validité de contenu.

La **validité de construit** a trait à la qualité de l'élaboration et du choix des indicateurs. La **validité convergente et discriminante**, qui veut qu'un indicateur mesure un et un seul concept, est un test d'une bonne validité de construit. Elle repose sur l'idée que, d'une part, les divers indicateurs d'un même concept seront liés entre eux et que, d'autre part, ces mêmes indicateurs seront peu corrélés à d'autres indicateurs censés mesurer des concepts différents.

La **validité liée au critère** est respectée lorsque la mesure d'un concept dont on postule qu'il est lié à un autre concept - le critère - l'est effectivement. On la dit **concurrente** lorsque le critère est mesuré en même temps, et **prédictive** lorsque le critère est mesuré plus tard.

Un exemple illustre les divers types de validité. L'élaboration d'un nouveau test de maîtrise du français demanderait de poser des questions 1) manifestement reliées à la connaissance du français (validité apparente), 2) qui couvrent les principales dimensions de la connaissance du français et non un nombre restreint de dimensions (validité de contenu) et 3) qui sont liées positivement entre elles et moins liées aux réponses à un test de mathématiques (validité de construit, convergente et discriminante). Enfin, le score total devrait être corrélé aux résultats scolaires en français au même moment, puisque ceux-ci constituent une autre mesure des connaissances en français (validité concurrente). Il devrait également permettre de prédire les résultats scolaires futurs (validité prédictive).

Il revient au chercheur de démontrer que les opérations choisies traduisent adéquatement le concept qu'il prétend mesurer. L'appréciation de la validité d'un indicateur est en partie qualitative et subjective. La tâche de la personne responsable de la recherche consiste à réduire la subjectivité au minimum à partir de ses connaissances du phénomène étudié et des instruments utilisés. Parce que l'élément subjectif demeure, il peut y avoir des désaccords sur la validité d'un indicateur. Les débats sont nombreux, parfois virulents, et ils sont rarement tranchés de façon définitive. Dans la plupart des cas, une opérationnalisation est retenue dans la majorité des études parce qu'elle semble mieux traduire le concept de départ jusqu'à ce que de nouvelles informations remettent ce choix en question.

2.3 /

### La vérification empirique des critères : la fiabilité

Une étape importante de la structuration de la recherche consiste à vérifier la fiabilité des indicateurs, selon les deux aspects fondamentaux : la stabilité dans le temps et l'équivalence dans l'espace. La procédure d'évaluation classique de la stabilité dans le temps est connue sous le nom de « **test-retest** ». La même mesure est reprise à des moments différents auprès des mêmes unités et les résultats sont comparés. Cette procédure pose des problèmes particuliers en sciences sociales. Prenons le cas d'un questionnaire qui porterait sur la consommation des médias. On pourrait vérifier sa fiabilité en faisant passer le même questionnaire après un intervalle de quelques semaines ou mois. Toutefois, on ne peut pas être certain que les personnes interviewées n'ont pas modifié leur comportement entre les deux mesures. Il est possible de réduire l'intervalle entre les deux mesures pour remédier à ce problème, mais ce faisant, on crée une autre difficulté. Si l'intervalle est trop court, les individus se rappelleront les réponses données la première fois et pourront tenter de répéter celles qu'ils croient avoir fournies. Il n'existe aucune solution vraiment satisfaisante à ce problème. En conséquence, il est rare en recherche sociale que la stabilité des indicateurs soit testée par la méthode test-retest, sauf pour des caractéristiques qui, en principe, changent peu dans le temps.

On doit porter plus attention au critère d'équivalence. La procédure habituelle consiste à demander à des chercheurs différents d'effectuer les mêmes opérations et de comparer les résultats.

- Dans le cas d'un questionnaire, on peut comparer les réponses obtenues par différents interviewers. Ces interviewers n'ont pas interrogé les mêmes personnes mais, dans la mesure où les sujets ont été répartis aléatoirement entre les interviewers, on considère les différents sous-groupes comme équivalents. Si la distribution des réponses est très différente selon les interviewers, on conclura à un manque de fiabilité de l'instrument, associé à l'effet de l'interviewer. On doit se demander si la formulation de la question est assez précise et si les directives aux interviewers sont claires et bien comprises. Les firmes de sondages ont l'habitude de vérifier 10 % des entrevues pour s'assurer de détecter et de corriger cet effet le cas échéant.

- La même procédure s'applique à l'analyse de contenu. Pour estimer la proportion de nouvelles internationales dans la presse écrite, par exemple, on soumettra un échantillon du matériel à analyser à deux ou trois juges et on comparera leurs résultats. On parle

alors d'accord inter-juges. S'il y a des écarts significatifs entre les juges, il faudra vérifier que les règles d'interprétation de ce qui constitue une nouvelle internationale sont suffisamment précises et bien comprises.

- Il en est de même pour l'observation directe. Si l'on veut connaître la proportion des gens qui sourient dans le métro, on a intérêt à confier la tâche à plusieurs personnes et à comparer leurs comptations pour des moments et des lieux équivalents. Si les comptations divergent sensiblement, c'est que la grille d'observation n'est pas suffisamment détaillée, soit dans la description de ce qui constitue un sourire, soit dans les consignes données sur l'échantillonnage ou la durée de l'observation.

Ces exemples montrent l'avantage d'avoir recours à plusieurs personnes pour obtenir l'information souhaitée. C'est souvent la seule façon de vérifier la fiabilité des indicateurs, surtout si le concept d'intérêt est mesuré par un seul indicateur.

Le **prétest des instruments** est une étape essentielle qui devrait permettre de détecter les problèmes et de les corriger avant de procéder à la collecte des informations proprement dite. Il consiste à vérifier les instruments et le processus de collecte prévu sur une partie du matériel ou de l'échantillon avant de procéder à l'enquête principale. On doit, par exemple, prétester les questionnaires d'entrevue auprès de personnes qui feraient normalement partie de la population sondée, ce qui permet de voir si les questions sont comprises facilement et de la même manière par tous. Si ce n'est pas le cas, les personnes sondées auront tendance à demander de répéter les questions ou à ne pas y répondre, et certaines réponses ne correspondront pas à ce que l'on cherche à savoir. De la même manière, on testera les codes prévus pour une analyse de contenu auprès d'un échantillon préliminaire de textes ou une grille d'observation pour une période restreinte avant de procéder à la recherche proprement dite.

La situation est différente, mais le principe est le même lorsque l'on a recours à plusieurs indicateurs - et donc à plusieurs observations - pour mesurer un concept, une situation fréquente lorsqu'on mesure les attitudes. La fiabilité est alors vérifiée en ayant recours à la théorie classique des tests qui énonce que le score observé - la réponse d'une personne à une question - est une combinaison d'une partie représentant le « vrai » score - la « vraie » valeur du positionnement sur une échelle - et d'une partie d'erreur aléatoire attribuable aux aspects propres à chaque question. Le test de fiabilité le plus couramment utilisé est l'**alpha de Cronbach**, qui mesure la consistance interne, c'est-à-dire jusqu'à quel

point chacun des indicateurs - chacune des questions ou chacun des items, selon la terminologie de la psychométrie - constitue une mesure équivalente d'un même concept. Ce coefficient est en quelque sorte l'estimation de la corrélation que l'on obtiendrait entre deux séries d'indicateurs d'un même concept comprenant le même nombre d'éléments. Il indique jusqu'à quel point on peut se fier à l'ensemble d'indicateurs retenu comme mesure du concept d'intérêt.

Le coefficient alpha est la borne inférieure de la fiabilité réelle : celle-ci ne peut être inférieure à la valeur de l'indice alpha. Cette valeur augmente avec le nombre d'indicateurs, à condition que la corrélation moyenne entre ces derniers ne diminue pas avec l'ajout de nouveaux indicateurs. Il est donc possible de vérifier empiriquement la fiabilité lorsqu'on a recours à plusieurs indicateurs pour mesurer un même concept.

Ainsi, une mesure de la satisfaction intrinsèque au travail comprendra des questions relatives au degré de satisfaction concernant divers aspects relatifs au contenu du travail : intérêt des tâches, autonomie, défis, valorisation, utilisation des compétences, etc. Le degré de corrélation moyen entre les réponses à ces questions mesuré par le coefficient alpha de Cronbach donne un indice de la fiabilité. Celle-ci est considérée comme excellente si le coefficient alpha atteint plus de 0,85 et elle apparaît problématique s'il se situe au-dessous de 0,70. Un indice alpha élevé signifie que l'ensemble des éléments choisis pour représenter le contenu du travail est cohérent et qu'il constitue une bonne mesure de la satisfaction intrinsèque. D'autres indications permettront de voir si certains indicateurs réduisent la fiabilité et devraient être retirés. Si la fiabilité est acceptable, il est possible de construire une variable composite - une échelle - en additionnant les réponses aux divers indicateurs du concept.

Il est courant que la vérification de la consistance interne, en particulier dans le cadre de la mesure des attitudes, se fasse concurremment avec le contrôle de la validité convergente et discriminante.

## 2.4 / La vérification empirique des critères: la validité

Les validités apparente et de contenu se prêtent à un jugement qualitatif et subjectif. Par contre, lorsque l'on mesure un concept à l'aide de plusieurs indicateurs, on peut vérifier la validité de construit, convergente

et discriminante, par la concordance des mesures. De même, la validité liée au critère peut être vérifiée en examinant les relations entre les indicateurs d'un concept et d'autres variables mesurées de façon concurrente ou postérieure.

Lorsque plusieurs indicateurs d'un même concept sont utilisés, il est possible de vérifier que chaque indicateur représente bien une mesure du concept d'intérêt et qu'il constitue une mesure de ce seul concept dans les cas où la validité des mesures de plusieurs concepts est vérifiée concurrentement. Cette vérification se fait généralement à l'aide de l'analyse factorielle, exploratoire ou confirmatoire, une procédure qui vise à réduire un nombre important d'informations à un ensemble restreint de «composantes» ou de facteurs. On parle de **variables latentes** pour désigner ces facteurs qui existent sur le plan conceptuel seulement et qui ne sont pas mesurés directement.

La satisfaction au travail porte sur trois dimensions du travail : la satisfaction intrinsèque porte sur la nature même du travail, la satisfaction extrinsèque sur le salaire et les conditions de travail, et la satisfaction sociale sur la qualité des relations interpersonnelles. Pour chacune des dimensions, on pose des questions du type «Êtes-vous très satisfait, assez satisfait, peu satisfait ou pas du tout satisfait de... ? » pour un certain nombre d'éléments. Pour la satisfaction intrinsèque, ces éléments – tâches, défis, etc. – ont été mentionnés plus haut. En procédant ainsi, **on postule qu'une dimension générale de satisfaction intrinsèque existe** et que le positionnement des individus par rapport à cette dimension « explique » et donc « prédit » leurs réponses à chacune des questions. **Si cette hypothèse est avérée**, les réponses à ces questions seront fortement corrélées entre elles (**validité convergente**) et elles seront moins corrélées avec les indicateurs liés à la satisfaction extrinsèque ou sociale (**validité discriminante**), puisque les personnes auront tendance à répondre de la même manière aux questions portant sur une même dimension et différemment aux questions portant sur les autres dimensions.

L'analyse factorielle donne un sommaire des patrons de corrélations entre les indicateurs d'un ensemble donné. Si les regroupements proposés par l'analyse factorielle confirment que les indicateurs de chaque dimension sont liés à la dimension prévue et uniquement à elle, il est ensuite possible de vérifier la fiabilité des regroupements en

utilisant l'alpha de Cronbach et de procéder à la constitution d'échelles additives dans le cas où toutes les indications (validité convergente et discriminante, fiabilité) concordent.

Ainsi, il est possible de s'assurer avec un certain degré de confiance de la validité des mesures sur le plan empirique en procédant aux vérifications appropriées quant à la concordance de ces mesures. Une fois la validité de construit établie et les échelles construites le cas échéant, on pourra vérifier jusqu'à quel point ces échelles sont liées à d'autres mesures auxquelles elles devraient normalement être liées en utilisant les procédures statistiques appropriées (corrélations ou régressions, analyses de variance), ce qui permet de vérifier la validité critique.

## Conclusion

Les données empiriques n'existent pas à l'état pur. Elles sont produites par l'équipe de recherche en fonction de ses objectifs. C'est pourquoi il convient de parler de la **construction des indicateurs**, ce qui amène à reconnaître le rôle actif qui est dévolu aux responsables de la recherche dans la structuration de l'étude. Cette construction consiste en un ensemble d'opérations qui permettent de **traduire un concept** dans le langage de l'observation. Elle procède par approximation et vise à **classer des objets** à l'aide d'un certain nombre d'opérations effectuées à partir d'un ou de plusieurs instruments de mise en forme de l'information.

Pour que la traduction soit la plus adéquate possible, il faut à la fois faire preuve d'imagination - de façon à prendre en considération l'éventail des possibles - et de rigueur - de manière à écarter les indicateurs qui présentent de trop grandes lacunes. Un indicateur n'est satisfaisant que s'il est **fiable**, c'est-à-dire s'il donne des résultats constants dans le temps et l'espace. Pour cela, il devra être **précis**, ce qui **requiert** que la procédure soit suffisamment bien définie pour être reproduite. Il devra également être **non contaminé**, de telle sorte que les résultats ne soient pas subtilement influencés par la procédure utilisée. Ces conditions ne sont cependant pas suffisantes. La **validité** d'un indicateur repose sur la correspondance entre le contenu des opérations effectuées et la représentation que l'on se fait du concept de départ. La vérification de la validité comporte une part plus qualitative, basée sur le jugement du chercheur. Toutefois, il est possible de procéder à des vérifications empiriques de la

validité de construit et de la validité critérielle. Ces opérations revêtent une grande importance, puisqu'elles constituent le seul moyen de s'assurer que les conclusions tirées après avoir vérifié les hypothèses posées sont valables.

Comment l'équipe de recherche procède-t-elle pour être en mesure de mener ce processus? Elle **recense d'abord l'ensemble des indicateurs utilisés** dans les études qui ont utilisé les mêmes concepts. L'objectif de cette étape est d'établir une liste exhaustive des possibilités de mesure. Les indicateurs peuvent renvoyer à différents instruments de recherche, questionnaire, observation, analyse de contenu. On procède ensuite à **l'évaluation des indicateurs recensés** selon les critères énoncés. Cette évaluation se fait à partir des connaissances méthodologiques acquises sur les mérites et les limites des différents instruments de recherche, des évaluations qui ont déjà été faites dans les recherches antérieures et également d'une certaine connaissance du sujet étudié. L'opération vise à éliminer les indicateurs les moins appropriés, ce qui peut amener l'équipe de recherche à ne retenir qu'un indicateur qui apparaît nettement plus faible et valide que les autres. Cependant, on essaiera habituellement de retenir plus d'un « élu » de façon à pouvoir vérifier empiriquement la fiabilité et la validité. En l'absence d'indicateurs satisfaisants, on devra élaborer ses propres indicateurs, un choix qui a l'inconvénient de ne pas permettre la comparaison avec les études antérieures.

Après la collecte des informations, il sera possible de procéder aux analyses permettant la vérification empirique de la validité et de la fiabilité des mesures et de comparer ces informations à celles fournies par d'autres analystes qui ont utilisé les mêmes indicateurs, le cas échéant. L'étape de la mesure aura constitué un moment essentiel du processus qui permettra ensuite de répondre à la question de recherche qui, elle, porte sur les relations entre les concepts.

## Suggestions de lecture

- DURAND, C. (2016). « Surveys and Society », dans C. Wolf, D. Joye, T.W. Smith et Y.C. Fu (dir.), *The SAGE Handbook of Survey Methodology*, p. 58-66.
- GOULD, S.J. (1997). *La mal-mesure de l'homme*, Paris, Odile Jacob.
- MOULIN, S. et J.P. BEAUD (2015). « Quantification et mesure », dans J. Prud'homme, P. Doray et F. Bouchard (dir.), *Sciences, technologies et sociétés de A à Z*, Montréal, Presses de l'Université de Montréal, p. 186-188.
- SEIZ, M. et F. MAILLOCHON (2009). « Quantifier pour objectiver », dans *Le raisonnement statistique en sociologie*, Paris, Presses universitaires de France, p. 93-153.