

# ECHELLES DE LIKERT OU METHODE DES CLASSEMENTS ADDITIONNES

Marc Demeuse<sup>1</sup>

## 1. Création et utilisation pratique de l'échelle

Cette méthode est sans doute la plus intuitive qui soit (Mc Iver & Carmines, 1981, p. 22). Elle n'est pas basée, comme celle décrite par Thurstone, sur une étape préalable de calcul des scores des items, puis une seconde étape présentant des réponses dichotomiques lors de l'utilisation de l'échelle.

En quelque sorte, Likert « télescope » les deux étapes en une seule. Il soumet directement aux sujets dont il souhaite mesurer l'attitude, des propositions à juger selon une échelle prédéterminée, du type: *Tout à fait d'accord - D'accord - Sans avis - En désaccord - Tout à fait en désaccord*. Le nombre des modalités possibles peut varier (généralement de 3 à 7), mais correspond le plus souvent à un nombre impair de manière à ménager une modalité relativement neutre ou non polarisée au centre de l'échelle. Le nombre de modalités proposées est fonction à la fois de la précision souhaitée, mais aussi de la capacité présumée des sujets à effectuer des discriminations fines, un trop grand nombre de catégories n'entraînant pas nécessairement une augmentation de la précision.

Chaque catégorie reçoit une note a priori (par exemple, 5 pour *Tout à fait d'accord*, 4 pour *D'accord*, ...et 1 pour *Tout à fait en désaccord*). Pour les items dont l'acceptation est contraire à l'attitude générale que l'on veut mesurer au moyen de cette échelle, les valeurs sont ordonnées dans l'autre sens: 1 pour *Tout à fait d'accord* et 5 pour *Tout à fait en désaccord*.<sup>2</sup>

Il n'y a aucune obligation d'attribuer un score élevé aux propositions favorables à la dimension mesurée et un score faible aux propositions défavorables plutôt que l'inverse. Il faut cependant conserver la même logique dans l'ensemble de l'échelle. Une fois qu'un sujet a répondu, on calcule son score sur l'échelle en additionnant simplement les points de chaque modalité choisie. Cette solution postule bien sûr que chaque item possède le même poids et que la valeur liée à une modalité donnée peut être simplement additionnée à celle d'une autre. Par exemple, être tout à fait d'accord avec les propositions « La peine de mort doit être appliquée dans toutes les affaires impliquant des crimes de sang » et « Tous les fraudeurs devraient être sanctionnés par une amende proportionnelle à l'ampleur de la fraude » crédite le sujet d'un nombre identique de points sur l'échelle de « sévérité des peines à infliger ».

Un prétest permet en général de s'assurer que les items entretiennent bien un rapport avec l'attitude que l'on prétend mesurer. Il s'agit le plus souvent de procéder à une analyse factorielle de manière à vérifier la relative unidimensionnalité de l'échelle (grande cohérence entre les réponses aux différents items) et d'éliminer les items atypiques.

La procédure proposée par Likert est, comme le signale Oppenheim (1972, p. 133), plus légère que celle de Thurstone et les résultats obtenus par les deux méthodes, s'ils présentent les mêmes problèmes, sont relativement bien corrélés. C'est sans doute ce qui a rendu ce type

<sup>1</sup> Avec la collaboration de Philippe Wanlin.

<sup>2</sup> Si nous prenons l'exemple d'une échelle d'attitude relative au divorce, nous aurions, par exemple, l'item « les conditions actuelles du divorce ne sont pas aussi déshonorantes qu'il apparaît ». Choisir la catégorie « tout à fait d'accord » créditerait le sujet de 5 points, tandis que la catégorie « tout à fait en désaccord » le créditerait d'1 point. Pour la proposition « le divorce est scandaleux », on inverse l'attribution des points : « tout à fait en désaccord » = 5 points et « tout à fait d'accord » = 1 point.

d'échelles populaire. Cette méthode constitue une simplification de celle de la méthode de Thurstone.

D'une manière générale donc, après avoir rassemblé une série de propositions relatives à l'échelle à construire, ces items sont soumis à un petit échantillon de sujets. Ceux-ci sont invités, pour chacun d'eux, à émettre un jugement, c'est à dire à marquer leur degré relatif d'accord en indiquant une croix dans la colonne qui convient en regard de chacune des propositions, comme dans l'exemple proposé dans le tableau 1.

**Tableau 1** - Réponses d'un sujet à une échelle de type Likert sur l'attitude face aux enfants. Traduit et adapté de l'exemple proposé par Oppenheim (1972, p. 135)

	Tout à fait d'accord	D'accord	Sans avis	En désaccord	Tout à fait en désaccord
(1) Les enfants rapprochent le mari et la femme l'un de l'autre.	5	4	8	2	1
(2) Il est amusant de montrer aux enfants comment faire quelque chose.	5	8	3	2	1
(3) Il est nécessaire de veiller à faire sortir les enfants de certains de leurs travers naturels	1	2	3	4	8
(4) Une mère avec son jeune enfant échappe difficilement à la conversation et à la compagnie des adultes.	8	2	3	4	5
(5) En y réfléchissant, un enfant est plus une bénédiction qu'un fardeau.	5	8	3	2	1
(6) Il est souvent difficile de garder son calme avec un enfant.	1	2	3	8	5
(7) M'occuper des enfants me demande réellement beaucoup trop d'efforts.	1	2	3	8	5
(8) Si nous en avons les moyens, nous préférierions envoyer nos enfants en pension.	5	4	8	2	1
(9) Quand les choses sont difficiles, les enfants sont souvent une grande source de courage et d'inspiration.	5	4	3	8	1
(10) Si je devais recommencer ma vie, je souhaiterais encore avoir des enfants.	8	4	3	2	1

Contrairement à la procédure précédente, Likert ne recourt pas à des juges pour établir les valeurs à accorder aux différentes modalités des questions. C'est directement les réponses des sujets qui sont utilisées, mais les scores sont fixés a priori par l'auteur de l'échelle qui décide également de l'orientation de l'item.

Certains items sont parfois très clairement positifs (par exemple, l'item 2) ou négatifs (comme l'item 3) par rapport à l'attitude à mesurer, comme ici, dans le cas de l'attitude face aux enfants. D'autres sont plus difficiles à juger, comme l'item 8. Il a cependant été nécessaire de prendre une décision de manière à établir un score. En cas d'erreur (prendre un item pour favorable alors qu'il est défavorable, ou l'inverse), les corrélations entre les scores obtenus pour chaque item et l'ensemble de l'échelle nous permettent de corriger a posteriori le sens attribué à un item douteux.

Dans l'exemple, qui ne comporte que 10 items, le score obtenu par l'individu qui a coché cette grille peut se calculer comme indiqué dans le tableau 2.

**Tableau 2** - Notes obtenues par le sujet qui a rempli le questionnaire du tableau 1

item 1 =	3 points
item 2 =	4 points
item 3 =	5 points
item 4 =	1 points
item 5 =	4 points
item 6 =	4 points
item 7 =	4 points
item 8 =	3 points
item 9 =	2 points
item 10 =	5 points
Total =	35 points

Dans l'exemple, le score théorique le plus faible est de 10 points (car 10 items x 1 point) et le plus élevé, de 50 (10 items x 5 points). Une analyse des corrélations entre les scores obtenus à chaque item et le score total, pour un ensemble suffisamment large d'individus permet d'éprouver la consistance interne de l'échelle. Cela ne démontre pas qu'elle mesure parfaitement ce qu'elle prétend (validité), mais qu'elle est relativement cohérente par rapport à elle-même (fidélité).

D'un point de vue pratique, cette consistance interne s'éprouve par des techniques telles que celles qui ont été envisagées dans le chapitre relatif à la mesure de la fidélité (*formule de Spearman-Brown* à partir d'une comparaison dite « Split half », c'est-à-dire établie sur deux parties de l'échelle), *KR20* et, à présent, *alpha de Cronbach* (1952).

Les items qui sont corrélés négativement (mais fortement) avec le score total, comme nous l'avons vu, sont généralement recodés tandis que ceux qui obtiennent une corrélation faible sont éliminés (Murphy & Likert, 1938). C'est la raison pour laquelle un ensemble relativement plus large d'items que ceux qui seront finalement intégrés à l'échelle, sont testés lors de cette première phase.

Une validation externe peut également être entreprise, soit en s'assurant que les individus porteurs de scores élevés appartiennent à une population particulière (malades mentaux, consommateurs particuliers, membres d'associations militantes, ...) en rapport avec l'échelle alors que des scores beaucoup plus bas sont observés dans la population en général.

Un autre type de validation peut être réalisé en mettant en relation les scores avec une mesure ultérieure (par exemple, un vote réellement effectué par les sujets, les notes scolaires obtenues en fin d'année scolaire, ...) (validité prédictive). Une dernière manière de valider une échelle de ce type peut aussi consister à comparer ce nouvel instrument avec d'autres échelles, plus anciennes, et bien documentées (validité concurrente).

Une fois ce travail effectué, l'échelle peut alors être employée de manière effective. Le calcul des scores individuels est entrepris exactement de la même manière que lors de la phase de mise au point sur les items qui ont été conservés dans l'échelle.

## 2. Critique de l'échelle

L'objection la plus souvent formulée à l'encontre de ce type d'échelle concerne son unidimensionnalité. Alexandre (1971) écrit que *le groupe de sujets utilisé dans le pré-questionnaire ne fait pas nécessairement partie de la population étudiée au cours de l'enquête proprement dite* (p. 35). Il ajoute que *le calcul des corrélations n'a de sens que si que si la distribution des scores est normale, ce qui n'est pas souvent le cas* (p.35). Il continue en soutenant qu'à cette critique sur le continuum, s'en ajoute une autre sur l'unité de mesure elle-même. Ainsi, *la notation des catégories et l'addition des scores, ne suffit pas, en effet,*

pour affirmer que les unités de mesure sur le continuum psychologique soient égales entre elles. Elles peuvent avoir pour effet de donner un même score à des sujets, alors que leur attitude est profondément différente (p. 35). En d'autres termes, cette méthode, plus rapide et aisée à mener que celle proposée par Thurstone, ne garantit cependant pas l'égalité des intervalles et ne tient pas compte du caractère plus ou moins « informatif » de chaque item. Par ailleurs, un grand nombre de combinaisons de réponses permettent d'obtenir un score identique (Anderson, 1990, p. 335).

Néanmoins, leurs nombreuses applications et la grande fidélité qu'elles manifestent généralement constituent des atouts importants pour les échelles de type « Likert ». De plus, ces échelles permettent de structurer les opinions entre elles et d'évaluer la position des sujets en fonction de cette structure. Le coût lié à la création de telles échelles est également un atout très important. Malheureusement, le développement des échelles, sans prétest sérieux, constitue un risque important, notamment en regard de la consistance interne de l'échelle. C'est alors *a posteriori* qu'un travail très important d'analyse et de sélection d'items doit prendre place. Cette solution, en apparence économique, est souvent adoptée notamment par les étudiants lors de la réalisation de leurs travaux de fin d'études, sans nécessairement être armés pour mener toutes les analyses. Un autre problème lié à l'absence de prétest est le coût important que cette manière d'investiguer occasionne cette fois au niveau de la passation auprès des sujets : des items très nombreux doivent être soumis de manière à disposer d'un volume suffisant d'information utile, au risque, dans le cas contraire, de ne pouvoir disposer d'une échelle aux qualités psychométriques suffisantes.

### 3. En résumé

La méthode élaborée par Likert est, en quelque sorte, une fusion entre l'étape de création de l'échelle et celle de son utilisation, même si on ne peut qu'encourager les usagers à mettre correctement au point celle-ci avant de la soumettre à leur échantillon. Ainsi, on souhaite mesurer l'attitude de divers individus en leur présentant des affirmations préalablement incorporées à l'échelle. On leur demande de se prononcer, pour chaque affirmation, selon un système de modalités - généralement de 3 à 7 - ordonnées. Ces modalités de réponse aux items sont créditées d'un score fixé a priori par l'auteur de l'échelle. Celui-ci n'est pas nécessairement connu par les sujets qui peuvent s'exprimer sur une échelle verbale, voire même une échelle graphique continue, bien que cela soit beaucoup plus rare. On calcule le score de chaque sujet sur l'échelle en additionnant les points de chaque modalité choisie pour chacune des affirmations. Certaines procédures statistiques permettent de garantir la fidélité et la validité de l'échelle et de rejeter éventuellement certains items :

- L'analyse factorielle permet de s'assurer que les items entretiennent entre eux de forts liens. Elle permet donc de vérifier l'unidimensionnalité de l'ensemble de l'échelle ou de sous-échelles au sein de celle-ci et d'éliminer les items ambigus, polyvalents tout en gardant ceux qui fournissent un maximum d'information sur l'attitude étudiée.
- De la même manière, l'analyse des corrélations entre les scores obtenus à chaque affirmation et le score total permet de contrôler si l'échelle est relativement cohérente avec elle-même (fidélité par consistance interne). Certains items mal codés, c'est-à-dire présentant une corrélation forte mais négative, sont recodés, tandis que ceux qui ne sont pas ou faiblement corrélés sont éliminés. Des formules générales, comme celle de Spearman-Brown ou l'alpha de Cronbach permettent de calculer la fidélité de l'échelle à travers l'ensemble des items.

- En vérifiant si les individus ayant un score élevé appartiennent à une population présentant en général l'attitude mesurée, on peut s'assurer de la validité externe de l'échelle. On peut également mettre la validité à l'épreuve en mettant en relation les scores observés avec une mesure ultérieure (validité prédictive) ou en comparant l'échelle créée avec d'autres échelles éprouvées par le passé (validité concurrente).

## Bibliographie

Alexandre, V. (1971). *Les échelles d'attitude*. France: Editions universitaires.

Anderson, L.W. (1990). Guttman Scales, in H.J. WALBERG and G. HAERTEL (Eds), *The International Encyclopedia of Educational Evaluation*. Pergamon.

Laurencelle, L. (1998). *Théorie et techniques de la mesure instrumentale*. Sainte-Foy : Presses de l'Université du Québec.

Mc Iver, J.P. and Carmines, E.G. (1981). *Unidimensional scaling*. London: Sage Publications, Coll. Sage University papers, Quantitative Applications in the social Sciences, No. 24.

Murphy, G. and Likert, R. (1938). *Public Opinion and the Individual*. New York : Harper and Row.

Oppenheim, A. N. (1972). *Questionnaire Design and Attitude Measurement*, Heinemann.