

ECHELLES DE THURSTONE OU ECHELLES D'INTERVALLES (APPARAISSANT) EGAUX

Marc Demeuse¹

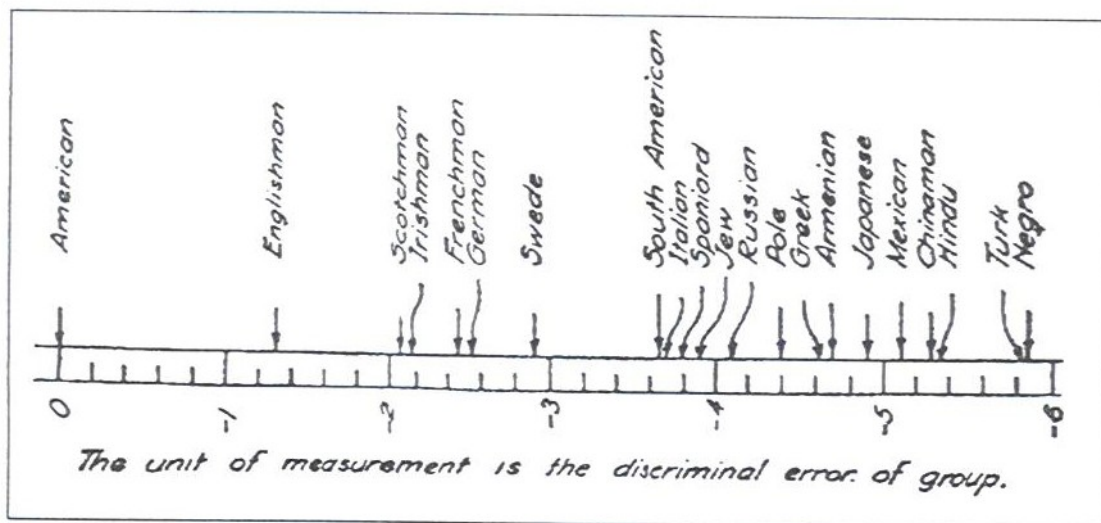
1. Description et principes

L'échelle postulée par Thurstone est une échelle « psychologique » non précisée, contrairement à l'échelle de Fechner qui postule une réelle échelle de sensation. En particulier, l'approche de Thurstone ne nécessite nullement d'utiliser des stimuli possédant une métrique physique. Thurstone (1959) s'intéressait en particulier à l'obtention d'échelles d'« excellence de l'écriture », de la « gravité des crimes », de la « préférence pour les nationalités » (encart 1). En psychologie sociale, un grand nombre d'échelles d'attitude sont construites sur cette base. (Bonnet, 1986, p. 136)

Ce type d'échelles repose sur deux étapes distinctes :

- la mise au point de l'échelle ;
- son utilisation pratique.

Encart 1 – Un exemple pratique d'échelle mise au point par Thurstone (1928b) : l'échelle des nationalités préférées des américains (d'après Martin, 1997, p. 89).



2. Mise au point de l'échelle

En fait, Thurstone (encart 2) tente de transposer les concepts et les méthodes de la psychophysique afin qu'ils puissent être utilisés dans d'autres domaines, comme celui de la mesure des attitudes. Dans ce domaine, il est cependant beaucoup plus difficile de classer les stimuli sur une base « évidente » d'intensité, du plus petit au plus grand ou du plus favorable au moins favorable, en fixant par ailleurs la taille de l'intervalle qui sépare deux propositions sur ce continuum. En effet, comme le fait remarquer Alexandre (1971, p. 27), les stimuli ne peuvent être rangés en grandeur croissante ou décroissante que par référence au jugement des personnes qui s'efforcent d'en percevoir les différences.

¹ Avec la collaboration de Philippe Wanlin.

Encart 2 – Louis Leon Thurstone (d'après Martin, 1997)

Thurstone (1887-1955) est certainement l'acteur central du rapprochement de la psychophysique et la théorie des tests à travers deux contributions essentielles : sa recherche d'une unité de mesure et d'un zéro vrai en psychologie et sa « loi du jugement comparatif ». C'est sans doute à sa formation initiale d'ingénieur¹ et à l'intérêt qu'il porte par la suite à la psychologie, suivant ses premiers cours auprès de Madison Bentley et Edward Bradford Titchener² à l'Université Cornell qu'on devra ses développements. Il allie en effet de très solides connaissances mathématiques, une formation en physique, la fréquentation de professeurs issus de la psychophysique et la rencontre, à travers le *Psychological Committee*, des figures de proue de la mesure de l'intelligence aux Etats-Unis, comme Goddard, Terman, Thorndike et Yerkes.

Une façon d'approcher ce processus consiste à demander aux sujets de comparer l'ensemble des stimuli par paires de manière à établir des relations d'ordre entre chacun d'eux. C'est ce que nous avons vu lors de la description de la méthode de comparaisons pairées, dans la partie relative à la psychophysique. Une autre solution, celle que nous allons détailler ici, consiste d'abord à faire classer les stimuli par des juges. Ainsi, on demande à un nombre important de juges, pouvant aller de quelques dizaines à plusieurs centaines, selon la précision désirée, de classer généralement en 7 à 11 catégories ordonnées les différents stimuli ou items selon qu'ils sont favorables ou défavorables par rapport à l'objet vis-à-vis duquel on met au point l'échelle. La position de chaque stimulus ou item sur l'échelle correspond à la valeur prise par la moyenne ou, plus fréquemment, le médian des réponses des juges. Mais nous reviendrons sur ce point plus loin. Le problème le plus difficile consiste à estimer la taille des intervalles entre ceux-ci.

Après avoir réalisé cette première étape qui vise essentiellement à attribuer une valeur métrique à chaque item, l'instrument est alors opérationnel et peut être utilisé pratiquement pour situer les sujets sur ce même continuum en fonction des items qu'ils sont susceptibles d'accepter ou de rejeter.

De manière à mieux faire comprendre la méthode, nous allons proposer deux exemples, l'un dans le domaine de la psychophysique, l'autre dans le domaine des échelles d'attitude.

3. Un exemple d'échelle de Thurstone ayant un substrat physique mesurable

3.1. Création de l'échelle

Nous allons d'abord considérer un exemple pour lequel existe un continuum physique: la couleur. Imaginons que nous ne disposons pas d'un photomètre, instrument relativement coûteux, mais d'une grande série de petits témoins colorés de même saturation, mais de teintes différentes³. Nous souhaitons réaliser une échelle colorimétrique allant du rouge au bleu de manière à pouvoir « qualifier » de façon univoque et précise des pièces de tissus colorés. Nous ne connaissons pas a priori l'écart qui existe entre chaque témoin.

Comme Thurstone, on pourrait charger un large groupe de personnes (ayant une vision des couleurs non altérée !) de classer séparément les différents témoins selon la dimension "teinte" en demandant qu'ils placent chaque témoin dans l'une des 11 catégories, numérotées de 1 à 11, allant de « absolument rouge » (catégorie n° 1) à « absolument bleu » (catégorie n° 11), en essayant de couvrir l'ensemble du spectre.

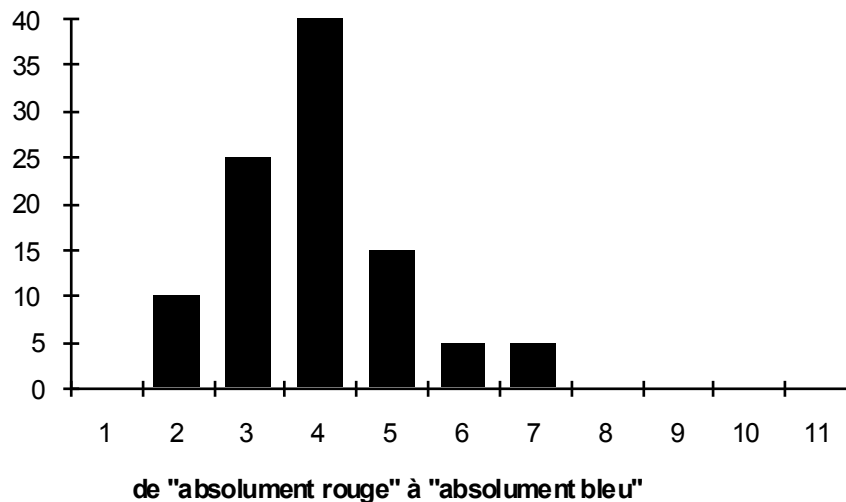
¹ Il travaillera au sein du laboratoire de Thomas Edison.

² Edward Bradford Titchener a été l'élève de Wundt en Allemagne où il a travaillé dans le laboratoire de psychophysique.

³ Concrètement, ces témoins sont de petits cartons colorés.

Chaque témoin recevra pour chaque juge le numéro de la case qu'il occupe chez ce juge. Comme le classement des différents juges peut fluctuer, on pourra attribuer à chaque témoin la valeur moyenne (ou médiane) obtenue pour l'ensemble des juges et calculer la probabilité qu'un témoin donné obtienne les différentes valeurs allant de 1 à 11. La figure 1 représente, pour un témoin donné, la répartition des réponses de tous les juges.

Figure 1 - Exemple d'histogramme relatif à la distribution des réponses d'un groupe de juges pour un témoin coloré particulier.



Comme nous souhaitons quantifier la distance qui sépare deux témoins, de manière à constituer une échelle métrique, nous allons utiliser la distribution normale réduite pour normer le continuum. Cette distribution théorique, outre qu'elle intervient habituellement en psychophysique pour modéliser les probabilités de réponses de détection des sujets, a aussi l'avantage d'être parfaitement tabulée¹.

Le recours à des probabilités théoriques nous permettra, connaissant la probabilité qu'un témoin appartienne à chaque case, de déterminer la « largeur » des différentes cases. Cette transformation d'une échelle ordinaire comportant différentes modalités de « rouge » à « bleu » en une échelle d'intervalles, suppose cependant que nous acceptions que les jugements des sujets se distribuent normalement autour d'une valeur vraie, ce qui constitue un postulat classique en psychophysique.

¹ Nous disposons actuellement d'ordinateurs et de calculettes, ce qui n'était pas le cas, de manière courante avant le début des années 60. Le recours à des tables était donc indispensable. La distribution normale réduite constituait l'une des distributions bien tabulée et qui, de plus, correspondait à un modèle classique de la répartition des erreurs de détections en psychophysique, ce qui explique son utilisation ici (cf table en annexe).

Tableau 1 - *Distribution des réponses de l'ensemble des juges pour un témoin donné*

Classes	Fréquences relative	Fréquences relatives cumulées
1	0,00	0,00
2	0,10	0,10
3	0,25	0,35
4	0,40	0,75
5	0,15	0,90
6	0,05	0,95
7	0,05	1,00
8	0,00	1,00
9	0,00	1,00
10	0,00	1,00
11	0,00	1,00

A partir du tableau 1, qui reproduit, de manière numérique cette fois, les informations contenues dans la figure 1, nous allons estimer la largeur de chacune des classes. Pour ce faire, nous allons calculer les fréquences cumulées en partant de la classe 1 (aucun juge n'a placé le témoin dans cette case, la fréquence absolue, comme la fréquence cumulée, sont nulle), jusqu'à la 11^e classe (la fréquence absolue est nulle, mais puisqu'il s'agit d'un choix forcé, la fréquence relative cumulée est nécessairement égale à 1,00 ou 100%, si nous nous exprimons en pourcentage).

En considérant les fréquences relatives cumulées par rapport à celles d'une table normale réduite (annexe 1), on peut faire correspondre à chacune d'elles une position en terme d'écart à la moyenne qui, quant à elle, vaut 0 par définition de la distribution normale réduite). Ainsi, par simple lecture de la table, on constitue le tableau 3 en ajoutant une colonne au tableau précédent.

Tableau 2 – Transformation des fréquences cumulées en écarts à la moyenne à partir de la lecture d'une table normale réduite

Classes	Fréquences relative	Fréquences relatives cumulées	Écarts à la moyenne
1	0,00	0,00	---
2	0,10	0,10	-1,32
3	0,25	0,35	-0,415
4	0,40	0,75	0,675
5	0,15	0,90	1,28
6	0,05	0,95	1,645
7	0,05	1,00	---
8	0,00	1,00	---
9	0,00	1,00	---
10	0,00	1,00	---
11	0,00	1,00	---

Il est impossible d'attribuer une valeur exacte aux fréquences cumulées nulles ou égales à 1. En effet, ces deux fréquences ne correspondent à aucune valeur de la table, celle-ci tendant vers $-\infty$ pour une fréquence nulle et vers $+\infty$ pour une valeur égale à l'unité.

A partir de ces valeurs, nous calculons ensuite les écarts entre chaque paire de classes contiguës. La distance entre le centre de la classe 2 et celui de la classe 3 est donc de $-0,415 - (-1,32)$, soit 0,905 alors celle qui sépare le centre de la classe 3 du centre de la classe 4 est de $0,675 - (-0,415)$, soit 1,09.

En répétant cette opération pour tous les témoins, il est possible de calculer ensuite la taille moyenne de chaque distance entre deux centres de classes adjacentes. On obtient de cette manière une métrique basée sur la différence qui sépare deux centres de classes. Une autre façon d'exprimer cette idée est de dire que chaque classe possède à présent une certaine « largeur », en convenant que la limite qui sépare deux classes voisines se situe à égale distance de chacun des deux centres de classes. La taille d'une classe vaut donc la moitié de la distance entre le centre de la classe précédente et le centre de la classe considérée plus la moitié de la distance entre ce centre de classe et le centre de la classe voisine. De proche en proche, sauf pour les deux classes extrême¹, on peut établir la largeur de chaque classe. De manière à donner une origine à l'échelle, on fixera arbitrairement la valeur du centre de classe de la première classe à 0. En totalisant toutes les distances entre les centres de classe qui le précèdent et un centre de classe en particulier, on obtiendra alors la valeur de ce centre de classe.

Les différents témoins qui ont permis d'établir les valeurs des centres de classes recevront soit la valeur correspondant au centre de la classe modale des choix des juges, soit ils seront positionnés en calculant leur valeur moyenne ou médiane au départ des choix des juges. Ainsi, le témoin qui aura obtenu une moyenne de 5,666, en additionnant tous les choix

¹ La première classe, comme la dernière, de possèdent pas de limite, soit inférieure, dans le premier cas, soit supérieure, dans le second. Cette absence de limites résulte de la nature asymptotique de la distribution normale qui est utilisée.

effectués par les juges et en divisant par le nombre de ceux-ci, sera-t-il positionné entre le centre de la classe 5 et celui de la classe 6, à une distance du centre de la seconde équivalent à 1/3 de la distance qui sépare ces deux centres de classe. Lorsque tous les témoins auront été positionnés correctement sur l'échelle ainsi constituée, il sera possible d'utiliser l'instrument pour attribuer à un morceau de tissu teint sa valeur colorée.

3.2. Utilisation pratique de l'échelle

Pour attribuer une valeur à une pièce de tissu provenant de l'atelier de teinture, il suffira de présenter à un individu l'ensemble des témoins qui ont servi à établir l'échelle et de lui demander auxquels de ceux-ci la pièce examinée est identique. On calculera alors la valeur moyenne des témoins qui auront été jugés équivalents à la pièce de tissu et on attribuera cette valeur à l'étoffe. L'opération sera reproduite sur l'ensemble des tissus de manière, par exemple, à indexer les pièces produites dans un catalogue.

4. La construction d'échelles d'attitudes sans substrat physique.

Dans le domaine des attitudes, la méthode, développée par Thurstone comprend également les deux phases, comme nous venons de le voir, la première consistant à attribuer à des témoins une valeur métrique sur une échelle, la seconde à calculer la valeur individuelle d'un objet particulier qu'on souhaite « qualifier ».

Afin d'éviter la méthode des comparaisons par paires (Green, 1954; Edwards, 1957a) où chaque item est comparé successivement à tous les autres, ce qui représente un nombre considérable de comparaisons¹ et constitue une tâche longue et fastidieuse, Thurstone recourt également à la méthode des juges dans le domaine des échelles d'attitude. Il demande aux juges de distribuer, comme nous venons de l'illustrer avec l'échelle colorimétrique, sur un continuum matérialisé par 11 (parfois 7) intervalles l'ensemble des propositions qu'ils ont devant eux. Cette opération est effectuée en principe une fois pour toutes et sert à créer l'échelle proprement dite.

Alexandre (1971, pp 26-27) écrit que *le principal fondement est l'existence de l'attitude définie comme un continuum psychique de sorte que l'accord, le désaccord ou même l'indifférence à l'égard d'une question donnée font partie du même ensemble et expriment les différences non de nature mais de degré qui se répartissent, comme les nombres, le long d'une série arithmétique. L'objectif du constructeur de l'échelle est de créer, dès lors, un continuum de stimuli qui reproduise le continuum psychique de sorte que la valeur de chaque item rende compte de celle du degré d'attitude correspondant. Plus nombreux sont les degrés signifiés, meilleure est, par conséquent, l'analyse. Mais, dans la réalité, on distingue ordinairement 11 ou 7 intervalles, ce qui permet d'avoir une région neutre et un continuum suffisamment grand pour avoir la précision que l'on désire.*

Pratiquement, sept ou onze emplacements adjacents sont déterminés sur une table et les juges doivent chacun, sans se concerter, déposer les différentes propositions, imprimées sur des petits cartons, dans l'un de ces emplacements. Ils doivent placer celles qu'ils estiment "les plus..." à l'extrême gauche, "les moins..." à l'extrême droite et les autres entre ces deux extrêmes sans aucune obligation quant à la répartition de ceux-ci² (Edwards, 1957b).

¹ Nombre de comparaisons = **Erreur !**, n étant égal au nombre d'items.

² Cependant, les juges qui placent 25 % ou plus d'items dans la même case sont généralement éliminés (car ils sont supposés manquer de discrimination). Par ailleurs, la disposition elle-même des intervalles joue un certain rôle d'attraction, les cases n° 2, 6 et 10 étant le plus souvent choisies à cause d'un effet de tassement (end effect).

Pour que soit évaluée la distance entre les items et pas seulement l'ordre, Farnsworth¹, imagine de préciser que la case n° 6 est la région neutre et qu'il faut s'efforcer de déterminer la distance en fonction de ce point neutre de sorte que chaque proposition puisse être affectée d'un signe plus ou d'un signe moins et d'une intensité exprimée par la place de l'intervalle entre ce point neutre et l'extrémité².

Le nombre de juges fixé par Thurstone est assez important (200 à 300). Il ne donne pas de consigne particulière à propos de leur compétence sinon qu'ils doivent savoir lire et comprendre le sens des items. Il ne s'agit donc pas au sens strict d'experts du domaine concerné.

Cette méthode ne garantit bien sûr pas l'égalité absolue des intervalles. De façon triviale, on ne peut pas considérer que l'écart entre la classe N°5 et la classe N°6 est forcément équivalent à celui qui sépare deux autres classes adjacentes (disons les classes N°7 et N°8, par exemple). La méthode pose également le problème de l'origine (présence d'un zéro vrai). Thurstone s'est particulièrement intéressé à l'identification d'un zéro vrai qui permettrait de passer d'une échelle d'intervalles égaux à une échelle de rapports (encart 3).

Encart 3 – A la recherche du « zéro absolu ».

Thurstone, dans un article de 1928a intitulé « The Absolute Zero in Intelligence Measurement », propose une méthode permettant d'identifier un « zéro vrai » ou « zéro absolu » dans le domaine de la mesure de l'intelligence. Constatant la diminution de la moyenne et de la variance des scores entre sujets d'un même âge au fur et à mesure qu'on se rapproche de l'âge de 3 ou 4 ans en dessous duquel il devient impossible de soumettre les enfants aux tests d'intelligence, il calcule pour sept tests différents la valeur de cette intelligence nulle qu'il situe entre quelques mois avant la naissance et la naissance elle-même. Cette tentative, et surtout la position d'un « zéro vrai », qui rapproche la mesure de l'intelligence du problème de la température, dans le domaine de la physique, n'est pas anecdotique de la part de Thurstone dont la formation initiale est celle d'un ingénieur (Martin, 1997, p. 67 et suivantes) : s'il existe bel et bien un tel zéro, ainsi qu'une unité de mesure, il devient possible de considérer le problème de l'intelligence comme n'importe quel problème de mesure en sciences physiques. Il devient également possible de normer les tests et d'exprimer les résultats obtenus en « valeur absolue » d'intelligence et non plus en fonction d'un test particulier.

De manière à calculer la taille « réelle » des intervalles sur le continuum psychologique, Guilford (1954, pp. 223-...) détaille la méthode dite des « catégories successives » de Thurstone³.

Imaginons que nous souhaitions mesurer l'attitude d'une population face au divorce à partir d'une échelle qui établirait une note pour chaque personne en fonction d'un questionnaire. La première étape consiste à établir l'échelle de mesure de l'attitude puis, de confronter nos sujets à celle-ci.

4.1. Création de l'échelle

Pour des raisons de simplification, nous allons considérer, pendant la première étape que nous demanderons à un échantillon de 100 personnes, de classer une série de propositions relatives au divorce dans l'une des 7 catégories allant de « propositions acceptées par les personnes les plus défavorables au divorce » à « propositions acceptées par les personnes résolument en faveur du divorce » de manière à dégager un continuum de propositions.

¹ Farnsworth est également l'auteur d'un test de vision des couleurs permettant de diagnostiquer les troubles dans ce domaine.

² On se situe alors sur une échelle bipolaire.

³ Comme la mise en œuvre de la méthode est très lourde, nous reprendrons en l'adaptant à un autre domaine, l'exemple proposé par Guilford.

Les items utilisés par Thurstone - nous verrons un peu plus loin l'ensemble de l'échelle qui en comporte 22 - sont du type suivant :

1. Les conditions actuelles du divorce ne sont pas aussi déshonorantes qu'il apparaît.
2. Le divorce est scandaleux.
3. Une personne devrait avoir le droit de se marier ou de divorcer aussi souvent qu'elle le voudrait.

Le tableau 3 présente les résultats obtenus pour un ensemble de 6 des items de cette échelle. Ces items sont baptisés B, C, I, J, K, et N. Nous pourrions entreprendre la démonstration sur les 22 items, mais cela alourdirait quelque peu notre propos¹. Pour chacun d'eux, le tableau indique la répartition des choix des juges de l'échantillon (proportions cumulées).

Tableau 3 - Répartition des choix de 100 juges pour 6 des items de l'échelle d'attitudes face au divorce (fréquence cumulée des choix pour chaque item dans chacune des 7 catégories de jugement.

Item	Catégories successives						
	1	2	3	4	5	6	7
B	0,010	0,030	0,080	0,160	0,410	0,740	1,000
C	0,010	0,010	0,020	0,080	0,260	0,630	1,000
I	0,000	0,030	0,110	0,360	0,630	0,870	1,000
J	0,060	0,140	0,290	0,560	0,790	0,960	1,000
K	0,010	0,042	0,146	0,490	0,750	0,938	1,000
N	0,010	0,041	0,071	0,235	0,551	0,837	1,000

Puisque les distributions des réponses des juges sont supposées normalement distribuées pour chacun des stimuli autour d'une valeur vraie, Guilford établit un nouveau tableau dans lequel il substitue les valeurs extraites d'une table normale réduite aux proportions cumulées du tableau précédent (cf. Annexe 1). Ainsi, la fréquence observée de 0,010 en faveur de la catégorie 1 pour l'item B correspond-elle à -2,326 sur l'axe des x d'une distribution normale réduite (note z). La colonne 7 reste, quant à elle, vide puisqu'aucune note z ne peut être attribuée à la valeur 1,000 (sinon $+\infty$). Aucune valeur ne peut être attribuée à la catégorie 1 pour l'item I puisque, de manière symétrique, il n'existe pas de note z correspondant à une probabilité nulle (sinon $-\infty$).

Les valeurs inscrites dans le tableau 4 situent, en quelque sorte, en terme de note z, le centre de chaque catégorie par rapport à la moyenne obtenue par un item donné.

¹ La démonstration est réalisée sur des résultats fictifs qui ne correspondent pas aux données qui ont permis à Thurstone d'établir son échelle.

Tableau 4 - Distances en notes z entre chaque centre de classe et la moyenne obtenue par chaque item

Item	Catégories successives						
	1	2	3	4	5	6	7
B	-2,326	-1,881	-1,405	-0,994	-0,228	+0,643	
C	-2,326	-2,326	-2,054	-1,405	-0,643	+0,332	
I		-1,881	-1,226	-0,358	+0,332	+1,126	
J	-1,555	-1,080	-0,553	+0,151	+0,806	+1,751	
K	-2,326	-1,728	-1,054	-0,025	+0,674	+1,538	
N	-2,326	-1,739	-1,468	-0,722	+0,128	+0,982	

A partir du tableau précédent, il est alors possible d'évaluer les écarts entre le centre de deux catégories voisines en calculant simplement la différence entre les notes z de chaque catégorie. Par exemple, l'écart entre la catégorie 1 et la catégorie 2, pour l'item B, est de $|-1,881 - (-2,326)|$, soit 0,445. Le tableau 5 reprend l'ensemble de ces différences. Nous avons reporté la différence entre la catégorie 1 et la catégorie 2 dans la colonne « 2-1 ». Aucune différence ne peut être calculée avec une catégorie pour laquelle une note z n'a pu être attribuée¹.

Tableau 5 - Ecart entre les centres de classes

ITEM	Catégories successives					
		2-1	3-2	4-3	5-4	6-5
B		0,445	0,476	0,411	0,766	0,871
C		0,000	0,272	0,649	0,762	0,975
I			0,655	0,868	0,690	0,794
J		0,475	0,527	0,704	0,655	0,945
K		0,598	0,674	1,029	0,699	0,864
N		0,587	0,271	0,746	0,850	0,854
$\sum d$		2,105	2,875	4,407	4,422	5,303
$M_d = w$		0,421	0,479	0,734	0,737	0,884
L_c	0,000	0,421	0,900	1,634	2,371	3,255
M_c		0,210	0,660	1,267	2,002	2,813
A_c		-1,057	-0,607	0,000	+0,735	+1,546

Ensuite, nous pouvons calculer la différence moyenne observée entre 2 catégories successives, quel que soit l'item considéré (M_d). Cette valeur indique la largeur moyenne - quel que soit l'item - de chaque catégorie.

¹Ce qui est toujours le cas de la catégorie 7, lorsque 7 catégories sont utilisées – ou 11, si on utilise 11 catégories – puisque nous avons affaire à des fréquences cumulées observées et que au moins la dernière catégorie correspond à 100% des individus.

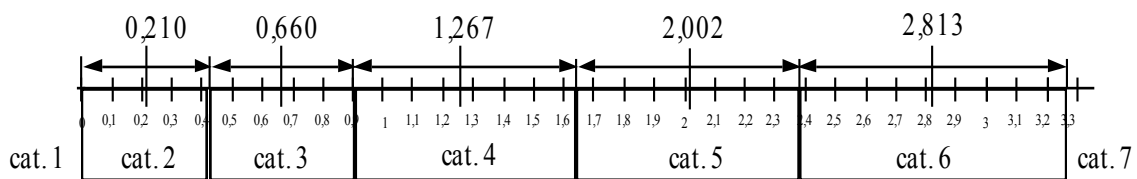
A partir de ces largeurs, il est alors possible d'établir les valeurs des centres de classes (M_c). La valeur L_c correspond à la limite supérieure de la classe considérée. Elle est obtenue en ajoutant la valeur de M_d à la valeur L_c de la classe précédente. Ainsi, pour la classe 4, la limite supérieure est égale à $0,900+0,734=1,634$. La valeur inférieure d'une classe est bien sûr égale à la limite supérieure de la classe précédente. Ainsi, pour la classe 4, la valeur inférieure est 0,900. Le centre de la quatrième catégorie se situe à équidistance des deux limites de classe, c'est-à-dire à $0,900 + \frac{0,734}{2} = 1,267$.

La limite supérieure de la première classe est fixée à 0. La limite inférieure et le centre de cette classe sont inconnus, comme la limite supérieure et le point moyen de la septième classe d'ailleurs.

Il est aussi possible de référer l'échelle à un point neutre, situé au centre de celle-ci (catégorie 4). Les centres de classes sont alors baptisés (A_c). L'échelle que nous venons de construire est reproduite graphiquement dans la figure 2.

Comme l'indique la figure 2, les différentes classes n'ont en fait pas la même largeur sur le continuum.

Figure 2 - Echelle d'appréciation du caractère favorable au divorce d'une série de propositions.

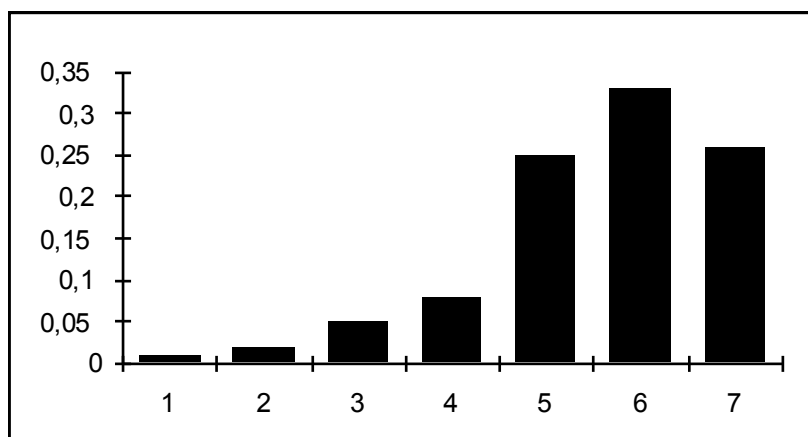


Il existe plusieurs façons d'établir pour chaque item sa valeur sur l'échelle, une fois les centres des catégories connus. On peut, par exemple, choisir pour chacun la valeur centrale de la classe modale ou la valeur de l'échelle correspondant au médian de la distribution des choix initiaux. Dans l'exemple proposé, l'item B (tableau 6 et figure 3), dont la classe modale correspond à la sixième catégorie, recevra la note de 2,813 (M_c) ou de +1,546 (A_c), si nous avons centré notre échelle sur la catégorie 4.

Tableau 6 - Distribution des réponses des juges pour l'item B

Catégorie	Fréquence relative
1	0,01
2	0,02
3	0,05
4	0,08
5	0,25
6	0,33
7	0,26

Figure 3 - Distribution des réponses des juges pour l'item B.



4.2. Utilisation pratique de l'échelle

Une fois que chaque item a pu être caractérisé sur l'échelle, grâce la classification opérée par chacun des juges, il est possible d'utiliser cette échelle de manière à caractériser l'opinion d'une population en matière de divorce. Pour ce faire, on demandera à un nouvel échantillon de sujets représentatif de la population à étudier de dire pour chaque item s'ils sont ou non d'accord avec la proposition qui est faite. La note attribuée à chaque sujet sera la moyenne des notes de chacun des items avec lesquels il aura marqué son accord. De manière générale, Alexandre (1971, p. 26) indique que les questionnaires de type thurstonniens présentent *des items dichotomiques, ayant chacun une cote, auxquels les sujets sont priés de répondre en manifestant leur approbation ou leur désapprobation. On totalise la somme des cotes des items acceptés, puis on calcule la moyenne ou plus fréquemment le médian.*

Reprenons l'exemple de l'échelle d'attitude face au divorce et considérons cette fois les 22 items qui la composent. Cette échelle, qui a été mise au point, non pas sur la base d'un système de 7 catégories, comme dans notre démonstration ci-dessus, mais à partir de 11 classes, a permis d'attribuer à chaque item une valeur d'échelle¹. Le premier item, qui reçoit une valeur égale à 3,7 indiquera une attitude bien moins favorable au divorce que l'item 10 qui est crédité d'une valeur de 10,1, mais plus favorable que l'item 5, dont la valeur est de 0,5.

La dernière colonne, qui indique les valeurs de chaque item, n'est naturellement pas présentée aux individus à qui les items sont soumis. Il leur est simplement indiqué de noter leur accord ou leur désaccord avec chacune des propositions. Le tableau 7 présente les réponses d'un sujet à cette échelle.

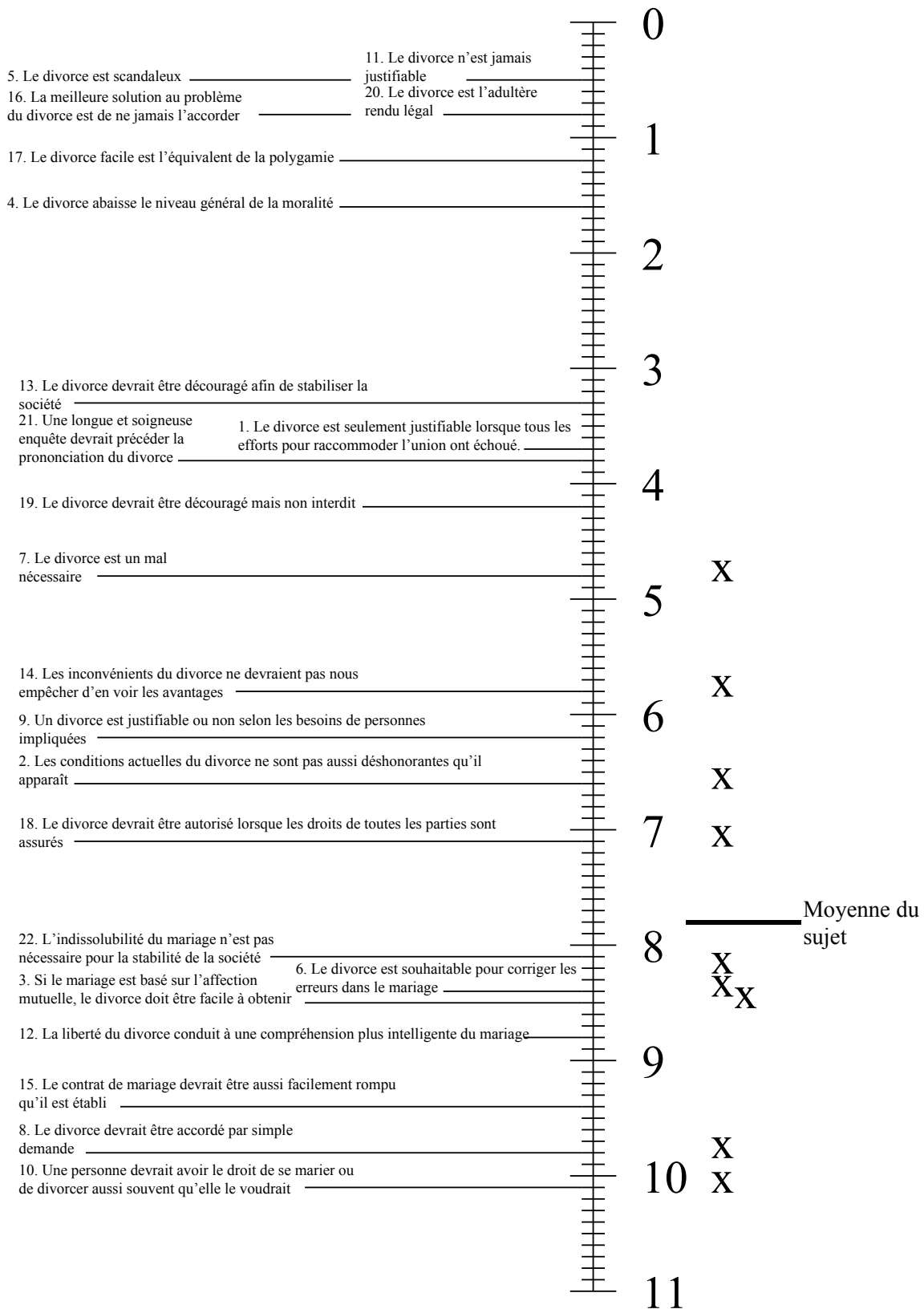
Pour obtenir le score du sujet dont les réponses sont présentées dans le tableau 6, on calcule donc la moyenne des valeurs sur l'échelle de chaque item pour lesquels l'individu a marqué son accord (c'est-à-dire les items 2, 3, 6, 7, 8, 10, 14, 18 et 22). Ce sujet obtient une valeur moyenne de $(6,6+8,5+8,4+4,8+9,8+10,1+5,8+7,1+8,1)/9 = 7,688$ et présente plutôt une attitude favorable – mais de manière assez modérée - face au divorce.

¹ Il s'agit d'une échelle de 22 items construite par Ellickson selon la méthode des intervalles apparaissant égaux de Thurstone dont la fidélité vaut 0,86. Un haut score indique une attitude favorable au divorce. Cependant, les caractéristiques de cette échelle étant peu connues, il serait souhaitable, en vue d'une utilisation éventuelle, de calculer à nouveau la cote des items et de s'assurer de la validité. Source : Thurstone L.L. – *The measurement of social attitudes* – University of Chicago Press, 1931. Reproduite dans M.E. Shaw et J.M. Wright in Alexandre V. – *Les échelles d'attitude*. – Editions Universitaires, encyclopédie universitaire, Paris, 1971, p. 129.

Tableau 7 – Réponses d'un sujet (x) à l'ensemble de l'échelle d'attitude face au divorce

Item	Accord	Désaccord	Valeur d'échelle
1. Le divorce est seulement justifiable lorsque tous les efforts pour raccomoder l'union ont échoué.		X	3,7
2. Les conditions actuelles du divorce ne sont pas aussi déshonorantes qu'il apparaît.	X		6,6
3. Si le mariage est basé sur l'affection mutuelle, le divorce doit être facile à obtenir.	X		8,5
4. Le divorce abaisse le niveau général de la moralité.		X	1,6
5. Le divorce est scandaleux.		X	0,5
6. Le divorce est souhaitable pour corriger les erreurs dans le mariage.	X		8,4
7. Le divorce est un mal nécessaire.	X		4,8
8. Le divorce devrait être accordé par simple demande.	X		9,8
9. Un divorce est justifiable ou non selon les besoins de personnes impliquées.		X	6,2
10. Une personne devrait avoir le droit de se marier ou de divorcer aussi souvent qu'elle le voudrait.	X		10,1
11. Le divorce n'est jamais justifiable.		X	0,5
12. La liberté du divorce conduit à une compréhension plus intelligente du mariage.		X	8,8
13. Le divorce devrait être découragé afin de stabiliser la société.		X	3,3
14. Les inconvénients du divorce ne devraient pas nous empêcher d'en voir les avantages.	X		5,8
15. Le contrat de mariage devrait être aussi facilement rompu qu'il est établi.		X	9,4
16. La meilleure solution au problème du divorce est de ne jamais l'accorder.		X	0,8
17. Le divorce facile est l'équivalent de la polygamie.		X	1,2
18. Le divorce devrait être autorisé lorsque les droits de toutes les parties sont assurés.	X		7,1
19. Le divorce devrait être découragé mais non interdit.		X	4,2
20. Le divorce est l'adultère rendu légal.		X	0,8
21. Une longue et soigneuse enquête devrait précéder la prononciation du divorce.		X	3,8
22. L'indissolubilité du mariage n'est pas nécessaire pour la stabilité de la société.	X		8,1

Figure 4 – Représentation graphique de l'échelle d'attitude face au divorce, d'après les données du tableau 8. (Chaque croix, à droite de l'échelle, représente une réponse positive du sujet. La moyenne du sujet est indiquée par un trait horizontal)



La méthode que nous venons de détailler peut aussi s'appliquer à des questionnaires relatifs à d'autres attitudes. Après avoir calculé le score de chaque item selon les classements des juges, cette opération étant établie une fois pour toutes, on peut ainsi attribuer à chaque sujet à qui l'échelle sera administrée, sa note. Il sera alors possible de situer chaque individu sur le continuum qui aura été établi. Cela permet d'utiliser un instrument « absolu » de mesure d'une attitude qui attribue en quelque sorte une seule valeur à chaque sujet plutôt que de considérer chaque réponse à une série d'items. Cette démarche est d'autant plus utile qu'on désire mettre en relation une dimension donnée avec d'autres mesures, notamment pour prévenir certains types de comportements ou de risques, comme le suicide ou l'infarctus, en fonction de profils de personnalité.

5. Critique de l'échelle

Selon Alexandre, les échelles de Thurstone sont souvent contestées à trois égards : l'absence d'unidimensionnalité, la complexité de la méthode et le recours à des juges qui appartiennent éventuellement à une autre population parent que les sujets dont on veut mesurer l'attitude.

L'unidimensionnalité, c'est-à-dire l'existence d'un véritable continuum, est une condition souvent difficile à rencontrer. C'est donc l'absence d'unidimensionnalité qui constitue la première critique que l'on formule généralement à l'encontre de ce type d'échelles. Celle-ci peut néanmoins s'éprouver à travers l'analyse statistique des réponses des sujets soit de manière classique, à travers la mesure de la fidélité par consistance interne, soit à travers l'analyse factorielle.

Le second type de critiques est liée de complexité de la détermination des intervalles supposés égaux. En effet, l'égalité des intervalles, condition nécessaire pour que nous ayons affaire à une véritable échelle de mesure et non à une échelle ordinale, est non pas fondée sur le continuum des stimuli mais sur le continuum psychologique. Alexandre (1971, p. 31) soutient néanmoins que si l'on n'atteint pas la perfection dans ce domaine *ces objections ne sont pas de nature à ôter à cette échelle son caractère métrique* (p. 31).

La troisième critique concerne l'intervention de juges extérieurs pour décider de la position des items sur le continuum. Alexandre (1971, p. 31) précise ainsi « qu'il faut tenir compte dans l'utilisation d'une échelle de l'époque et du contexte culturel dans lequel elle a été construite ». En effet, il importe d'éviter d'établir une échelle au départ d'un ensemble de juges qui pourraient imaginer un classement des items complètement contradictoire par rapport à celui qui serait adopté par un nouveau groupe de juges issus de la même population que les sujets dont on souhaite mesurer l'attitude. Il y a malgré tout relativement peu de chance pour que le positionnement relatif des items soit modifié de manière fondamentale. On imagine plutôt un glissement de l'ensemble vers l'un des deux pôles de l'échelle. Dans l'exemple de l'échelle relative à l'attitude face au divorce, mais il en irait vraisemblablement de même pour d'autres domaines comme la contraception ou l'avortement, on peut s'attendre à un glissement relatif vers plus de permissivité et une acceptation globale plus large des items favorables. Ce glissement s'opérera sans doute en parallèle avec une atténuation des attitudes manifestement opposantes.

6. En résumé

Nous avons vu que l'on peut distinguer deux phases pour les échelle de type « Thurstone » : d'une part leur création, et, d'autre part, leur utilisation pratique. Ces échelles permettent de classer les stimuli sur un continuum et de fixer la taille des intervalles séparant les propositions sur ce même continuum.

Pour ce qui est de l'étape de la création, on demande à des sujets de classer des stimuli. La moyenne (ou le médian) des numéros de classe donnés par les différents juges permet de positionner les stimuli sur l'échelle. On estime la taille des intervalles grâce à une procédure statistique : d'abord, on calcule la probabilité qu'une attitude appartienne aux diverses classes. Ensuite, on convertit l'échelle ordinale en échelle d'intervalles via la transformation des fréquences cumulées de chaque item en note z. Enfin, on évalue les centres des différentes catégories et leurs écarts. On peut ainsi établir la valeur d'un item sur l'échelle en choisissant, par exemple, la valeur centrale de sa classe modale (ou la valeur de l'échelle correspondant au médian de la distribution des choix des juges). Cela permet de caractériser chaque item sur l'échelle, c'est-à-dire de lui attribuer une note sur le continuum.

Après avoir attribué à chaque item une valeur sur le continuum, on peut utiliser l'échelle pour mesurer l'attitude de sujets particuliers. Pour ce faire, on demande aux individus de marquer leur accord ou leur désaccord par rapport aux différents items de l'échelle. On calcule ensuite la moyenne des notes relatives aux items auxquels les individus ont marqué leur accord. Cette moyenne permet de situer chaque individu sur le continuum établi lors de la phase de création de l'échelle.

Bibliographie

- Alexandre, V. (1971). *Les échelles d'attitude*. France: Editions universitaires.
- Bonnet, C. (1986). *Manuel pratique de psychophysique*. Paris: Armand Collin.
- Green, B. (1954). Paired Comparison Scaling Procedures. In G. LINDZEY (Ed), *Handbook of Social Psychology, Vol. 1*, 344-347.
- Edwards, A.L. (1957a). Paired Comparison Attitude Scales. In *Techniques of Attitude Scale Construction*, Appleton-Century-Crofts, Educational Division, Meredith Corporation.
- Edwards, A.L. (1957b). The Method of Equal-Appearing Intervals. In *Techniques of Attitude Scale Construction*, Appleton-Century-Crofts, Educational Division, Meredith Corporation.
- Guilford, J.P. (1954), *Psychometric Methods*. Mc Graw Hill.
- Martin, O. (1997). *La mesure de l'esprit : origines et développements de la psychométrie, 1900-1950*. Paris : L'Harmattan. Histoire des Sciences Humaines.
- Thurstone, L.L. (1927a). Psychological Analysis. *American Journal of Psychology*, 38, 368-389.
- Thurstone, L.L. (1927b). A Law of Comparative Judgment. *Psychological Review*, 34, 273-286.
- Thurstone, L.L. (1928a). The absolute zero in intelligence measurement. *Psychological Review*, 35, 175-197.
- Thurstone, L.L. (1928b). An Experimental Study of Nationality Preferences. *Journal of General Psychology*.22. 415-430.