

LA MESURE EN PSYCHOLOGIE DE BINET À THURSTONE, 1900-1930

Olivier MARTIN

RÉSUMÉ : Le psychologue français Alfred Binet est à l'origine du développement de tests mentaux destinés à diagnostiquer le « niveau intellectuel » des enfants. Initialement conçus comme des tests cliniques, leur importation aux États-Unis dans les années 1910 a considérablement modifié leur usage, leur portée pratique et leur interprétation. Devenus les instruments de politiques eugéniques ou héréditaristes, utilisés dans des processus de sélection de grande échelle, les tests ont transformé la conception que les psychologues se faisaient de l'intelligence : initialement conçue comme multiforme et qualitative, elle est devenue une grandeur unique quantifiable. C'est dans ce contexte que le psychologue américain Louis Léon Thurstone, critiquant le manque de rigueur des méthodes statistiques utilisées dans les tests, a développé une nouvelle méthode permettant de disposer d'une unité et d'un zéro absolu pour la mesure (échelonnement) de l'intelligence. Une fois objectivée, sa méthode est devenue une « loi » générale permettant de rendre compte de tous les processus mentaux, de disposer d'un outil d'analyse des opinions et des attitudes.

MOTS-CLÉS : histoire, intelligence, mathématiques, mesure, psychologie.

ABSTRACT : The French psychologist Alfred Binet is at the origin of the development of the mental test intended for diagnosing the « intellectual level » of children. Initially conceived as « clinical tests » their importation to the United States during the 1910's has considerably altered their use, significance and interpretation. Turned into political tools by eugenicists and hereditarists and used in large-scale selection operations, the tests have changed the psychologists' conception of intelligence : initially conceived as multifaceted and qualitative, intelligence had become a single quantifiable entity. Within this context, the american psychologist Louis Leon Thurstone criticized the lack of exactness of the statistical tools used in the mental tests, and developed a new method able to give an absolute unit and zero of intelligence scale. Once objectivized, his method became a general « law » intended to account for all mental processes.

KEYWORDS : history, intelligence, mathematics, measurement, psychology.

Revue de synthèse : 4^e S. N^o 4, oct.-déc. 1997, p. 457-493.

ZUSAMMENFASSUNG : Ursprünglich entwickelte der französische Psychologe Alfred Binet Intelligenztests, die dazu bestimmt waren, das « intellektuelle Vermögen » von Kindern zu ermitteln. Während sie zu Anfang als klinische Tests konzipiert waren, änderten sich in den Jahren um 1910 nach der Einführung in den Vereinigten Staaten ihr Gebrauch, ihre praktische Bedeutung und ihre Interpretation. Sie wurden als politisches Instrument von Vertretern rassistischer Vererbungstheorien im großen Stil in Auswahlprozessen eingesetzt. Mit den Tests hatte sich die Auffassung von der Intelligenz, die sich die Psychologen erarbeitet hatten, gewandelt : anfänglich als mehrgestaltig und qualitativ aufgefaßt, wurden sie nun zu einer einzigen, quantitativ meßbaren Größe. In diesem Zusammenhang entwickelte der amerikanische Psychologe Louis Leon Thurstone, der die Aussagekraft statistischer Methoden in den Tests kritisierte, eine neuartige Methode, die durch Festlegung einer Einheit und eines absoluten Nullpunktes eine Abstufung der Intelligenz erlaubte. In eine objektivierbare Form gebracht, ist seine Methode zu einem allgemeinen Gesetz geworden, welches die Erfassung aller geistigen Prozesse ermöglicht. Man verfügt damit über ein Instrument zur Analyse von Meinungen und Haltungen.

STICHWÖRTER : Geschichte, Intelligenz, Mathematik, Messung, Psychologie.

Olivier MARTIN, né en 1966, sociologue et statisticien, est maître de conférences en sociologie à la faculté des sciences humaines et sociales de la Sorbonne (Paris V). Il est l'auteur de *La Mesure de l'esprit. Origines et développements de la psychométrie, 1900-1950*, Paris, L'Harmattan (« Histoire des sciences humaines »), 1997.

Adresse : Département des sciences sociales, Faculté des sciences humaines et sociales de la Sorbonne, Université Paris V, 12 rue Cujas, 75230 Paris Cedex 05.

Courrier électronique : olmartin@citi2.fr

Le développement des sciences statistiques et probabilistes est intimement lié à l'essor des sciences humaines, au moins durant la fin du XIX^e et le début du XX^e siècle. La reconstruction de l'histoire de ce développement est pourtant loin d'être terminée : les liens entre psychologie, statistique et probabilité au XX^e siècle sont imparfaitement connus ; les voies par lesquelles les outils statistiques et probabilistes ont permis aux psychologues de disposer d'outils de mesure et de quantification de leurs grandeurs sont quasiment inconnues¹. Nous voudrions, dans cet article, contribuer à la connaissance de l'histoire de ces voies en retraçant l'une d'entre elles : la voie qui mène des premiers outils d'Alfred Binet aux théories et méthodes développées par Louis Léon Thurstone autour de 1925 pour améliorer et justifier la mesure en psychologie des opinions, des attitudes, des valeurs ou de l'intelligence.

Quand le XIX^e siècle bascule vers le XX^e siècle, Alfred Binet est un psychologue français qui serait aujourd'hui baptisé « clinicien ». Lorsqu'il développe les premiers tests d'intelligence et donc les premiers outils de « mesure » de l'intelligence, il n'imagine certainement pas que plus tard et plus loin ces outils connaîtront un destin remarquable. Les circonstances historiques ayant accompagné et forgé ce destin sont évidemment complexes mais certaines d'entre elles méritent, en raison de leur rôle prépondérant, une attention particulière : ce sont, d'une part, les circonstances sociales et intellectuelles de circulation et d'utilisation des outils et, d'autre part, les travaux du scientifique américain Louis Léon Thurstone.

Thurstone est une des figures emblématiques de la psychologie : cofondateur de la Psychometric Society et du journal *Psychometrika* (1935/1936), il est souvent considéré comme un des principaux acteurs — sinon le principal — du développement des méthodes d'analyse factorielle. Pourtant, c'est dans ses recherches sur la mesure que Thurstone estime voir sa principale contribution à la science psychologique². Dans les années vingt,

1. [Pour plus de précisions sur les références citées en notes, se reporter à la Bibliographie finale, p. 491-493.] Seule l'histoire des outils d'inférence en psychologie est à peu près reconstruite : sur ce point le lecteur pourra consulter les articles de Gerd GIGERENZER, Kurt DANZIGER et David MURRAY dans le second volume de *The Probabilistic Revolution*, Cambridge, MIT Press, 1987. Les ouvrages de BENZÉCRI, 1982, de REUHLIN, 1962, et de GOULD, 1983, constituent certainement les rares exemples d'étude des liens entre psychologie et statistique au XX^e siècle. Malheureusement aucun d'entre eux n'aborde avec précision l'analyse des premiers travaux de Thurstone (avant ses recherches sur les méthodes factorielles), ni les circonstances sociales de ce développement, ni les conséquences de ce développement sur les concepts psychologiques (intelligence en l'occurrence).

2. THURSTONE, 1952, p. 304, 307 ; ID., 1959, p. 15.

avant de s'intéresser à la théorie factorielle, reprenant dans leurs aspects formels et statistiques les outils de Binet tels que les conditions d'utilisation les ont transformés, Thurstone cherche à en améliorer la rigueur. Il élabore alors une méthode de mesure puis propose un cadre théorique unique pour la mesure de toutes les grandeurs psychologiques — aussi bien l'intelligence, les opinions ou les attitudes, que les processus de la psychophysique. Ce cadre unique est constitué par la « loi du jugement comparatif » (1927).

De très grandes différences séparent les travaux de Binet et ceux de Thurstone : différences dans les outils utilisés, dans le sens du terme « mesure » et dans les théories et concepts psychologiques constituant le cadre théorique des outils. Saisir ces différences suppose que soient reconstruits les sens du mot mesure et la signification des pratiques qui y sont associées à la fois chez Binet, chez les principaux utilisateurs américains de la mesure en psychologie et, bien sûr, chez Thurstone. C'est relativement à ces sens que les changements intervenus entre 1900 et 1930 pourront être compris. Ainsi, l'étude de ces sens et donc de l'outillage mental et technique est une étape obligée de la reconstruction de l'histoire des évolutions. À son tour, cet outillage ne se comprend que si une analyse détaillée de sa genèse et de ses évolutions est menée. Enfin, saisir les causes de cette genèse et des évolutions qui l'ont suivie suppose que soient reconstruites certaines circonstances sociales et institutionnelles. C'est à ces diverses tâches que s'attelle cet article.

Autant que faire se peut, cet article respecte la chronologie. La première section analyse la genèse des outils de Binet, leur insertion en Amérique et les transformations qu'ils subissent. La deuxième s'attache à reconstruire les changements introduits par Thurstone en tentant, dans un premier temps, d'identifier les espaces intellectuels et institutionnels au sein desquels Thurstone s'inscrit, puis de comprendre le sens des critiques qu'il formule à l'encontre des outils de mesure utilisés dans ces espaces, et enfin de saisir le sens et la portée de ses propres travaux : sa méthode d'échelonnement et sa loi psychologique.

1. — LES OUTILS DE BINET : DE LA FRANCE AUX ÉTATS-UNIS

1.1. Alfred Binet et la mesure des processus supérieurs

Les premiers tests psychologiques, au sens d'épreuves standardisées destinées à plusieurs individus, remontent au moins aux études des capacités psychosensorielles humaines menées au XIX^e siècle par les psychophysi-

ciens allemands. Par définition, ces épreuves ne permettaient d'étudier que les processus dits « inférieurs » tels que la vitesse de réaction à des stimuli sensoriels, la capacité à « discriminer » deux stimuli auditifs ou visuels... Wilhelm Wundt préconisait d'ailleurs de n'utiliser que ce type de test³.

Les premiers tests mettant en œuvre des processus « supérieurs » semblent remonter aux travaux de Hermann Ebbinghaus (psychologue allemand, 1850-1909) et surtout à ceux de James McKeen Cattell (psychologue américain, 1860-1944), ancien élève de Wundt à Leipzig, ancien assistant dans le laboratoire de Francis Galton à Londres et principal ouvrier de la construction d'une psychologie universitaire américaine. C'est dans un article de 1890 que Cattell présente des épreuves faisant appel à des processus dits « supérieurs » de l'esprit, comme la mémoire ou l'attention⁴. D'ailleurs, le fameux syntagme « *mental test* » a vu le jour dans cet article⁵.

Toutefois, malgré l'indéniable nouveauté que représentent les tests proposés par Cattell en 1890, leur champ d'application est restreint : l'auteur s'interroge sur leur portée réelle. Ils ne constituent pas encore des tests « d'intelligence ». Le mérite d'avoir franchi le pas revient à Binet. À travers plusieurs publications et plusieurs expériences qui s'échelonnent de la fin du XIX^e siècle à sa mort en 1911, Binet donne en effet naissance à des tests dits « tests d'intelligence » et à une notion passée dans le langage courant : le niveau intellectuel⁶.

1.1.1. Alfred Binet, la psychologie de l'enfant et l'instruction publique

Le développement des tests d'intelligence est le produit de la rencontre de deux préoccupations : d'une part, les préoccupations du ministère de l'Instruction publique qui s'interroge sur l'enseignement à procurer aux enfants en difficultés scolaires ; d'autre part, les préoccupations de Binet qui, depuis le milieu des années 1890, s'intéresse à la psychologie de l'enfant en étudiant les caractéristiques de ses nombreux mécanismes psychologiques (mémoire, imagination, compréhension, attention, coordination, etc.).

Alfred Binet (1847-1911) est marginal dans l'espace de la psychologie française de la fin du XIX^e siècle : ni la Sorbonne ni le Collège de France ne l'accueillent et il doit se contenter du laboratoire de psychologie physiologique de l'École pratique des hautes études, « petit laboratoire, malgré son

3. BORING, 1950, p. 483-485.

4. Cet article de CATTELL, 1890, est accompagné d'une note de Galton.

5. BORING, 1950, p. 571.

6. En fait, cette notion est plutôt passée dans le langage courant sous l'expression « âge mental » sans que cette différence de signifiant cache une différence de signifié. L'autre notion centrale de l'étude quantitative de l'intelligence est le quotient intellectuel (QI) conçu par William Stern en 1916. Voir SCHNEIDER, 1992, p. 114, et HORNSTEIN, 1988.

grand nom⁷ ». Titulaire d'une licence de droit, d'un doctorat en sciences naturelles et ayant travaillé avec Jean-Martin Charcot à la Salpêtrière, il a mené des recherches sur des thèmes très différents : l'hallucination, l'hypnotisme, le magnétisme, les altérations de la personnalité, les grands joueurs d'échec, etc. Bien que Binet soit passé à la postérité pour ses recherches sur l'intelligence, les travaux précédents constituent une part importante de son œuvre. En 1892, il est nommé directeur adjoint du laboratoire de l'École pratique des hautes études, puis directeur en 1895. La même année il fonde, avec Henri Beaunis, la revue *L'Année psychologique*. C'est dans ce cadre institutionnel qu'il commence à s'intéresser à la psychologie de l'enfant.

Cet intérêt rencontre les préoccupations du ministère de l'Instruction publique relatives à la notion d'inadaptation scolaire : en cette fin du XIX^e siècle, cette notion est en pleine élaboration, tout autant du point de vue psychologique que du point de vue social. En 1899, la Société libre pour l'étude psychologique de l'enfant est fondée (SLEPE⁸). Cette société, composée avant tout de médecins et de pédagogues, est la concrétisation des « rapports qui se nouent, dès la mise en pratique de l'obligation scolaire, entre la pédagogie et une psychologie clinique et expérimentale⁹ ». Binet, nommé sous-directeur (1899) puis directeur (1902) de cette société, trouve en elle les moyens de poursuivre ses recherches et de les légitimer ; et ce besoin de légitimation est grand, tant la psychologie de l'éducation est culturellement dominée dans l'espace universitaire. Son objectif est d'étudier les différences entre les enfants « normaux » et « anormaux » et d'arriver à identifier les arriérés. Son propos n'est pas de classer nominativement les individus selon un ordre total mais de mieux connaître la nature, les qualités et le fonctionnement de l'activité psychologique. Ce n'est qu'en 1905, après avoir essayé plusieurs voies méthodologiques (céphalométrie, psychophysique), qu'il prend définitivement conscience que « c'est par l'étude des processus supérieurs qu'il faut établir la psychologie individuelle¹⁰ ».

Le regard que Binet porte sur ces processus est celui d'un clinicien : son univers est profondément emprunt de la culture médicale et clinique. Ces circonstances vont le conduire à concevoir une méthode proche, avant tout, d'une méthode de diagnostic médical¹¹.

7. PAICHELER, 1992, p. 14. Pour des éléments biographiques, se reporter à SIMON, 1912; J. LARGUIER DES BANCELIS, *L'Année psychologique*, vol. 18, 1912; ou à la biographie de WOLF, 1973.

8. En 1917, la SLEPE devient la société Alfred Binet et Théodore Simon.

9. PINELL, ZAFIROPOULOS, 1983, p. 21.

10. BINET, 1905, p. 71. Avant cette date, l'étude des « processus supérieurs » n'était qu'un moyen parmi d'autres; voir FRAISSE, SEGUI, dir., 1994, p. 22-23, 146.

11. La lecture de son article programmatique, BINET, HENRI, 1896, p. 411, et les analyses de WOLF, 1969, l'établissent avec précision.

1.1.2. *Le « Binet-Simon » : test clinique d'intelligence*

C'est à partir de 1904 que Binet, alors président de la SLEPE, s'attache à l'étude précise des enfants « anormaux » et esquisse ce qui va devenir son test d'intelligence : il prend en charge la Commission pour l'étude des enfants attardés dont la mission est « d'étudier les mesures à prendre pour assurer les bénéfices de l'instruction aux enfants anormaux¹² ». La tâche qui lui incombe est de développer une méthode capable d'attester l'inaptitude effective de l'enfant à profiter, « dans la mesure moyenne », de l'enseignement donné dans les écoles ordinaires. Ce n'est qu'assez tardivement que le terme « intelligence » apparaît dans les écrits de Binet. Le sens que Binet attribue à ce mot est universel : « Presque tous les phénomènes dont s'occupe la psychologie sont des phénomènes d'intelligence : une sensation, une perception, sont des manifestations intellectuelles, autant que le raisonnement¹³. » Il se refuse à voir dans l'intelligence « une fonction unique et invisible¹⁴ » : elle est constituée d'un grand nombre de processus ; elle est « multiforme¹⁵ ».

Dans une première étape, Binet et son assistant Théodore Simon s'attachent à définir la notion « d'état inférieur » de l'intelligence. L'absence de nomenclature unique crée la confusion et les pratiques cliniques sont souvent arbitraires et instinctives : « Ce qui fait défaut, disent-ils, c'est *une base précise de diagnostic différentiel*¹⁶. » La recherche de cette « base précise » est l'objet d'une série d'articles devenus célèbres depuis (1905b, 1908, 1911). Ils y exposent une méthode pour « le diagnostic du niveau intellectuel des anormaux » en accompagnant leur présentation d'innombrables précautions¹⁷ : ils excluent que leur « test » puisse servir à distinguer « idiotie acquise et idiotie congénitale » et à établir un pronostic sur « le caractère curable ou non de l'arriération » ; ils insistent sur le caractère clinique des épreuves et sur la nécessité méthodologique que celles-ci se déroulent face à face. Ils excluent, par ailleurs, que cette échelle permette de mesurer l'intelligence : elle établit un classement entre des intelligences mais pas une mesure car « les qualités intellectuelles ne se mesurent pas comme des longueurs, elles ne sont pas superposables ». Ils parlent donc de mesure par commodité et soulignent que sa qualité dépend de la qualité du diagnostic¹⁸. Ils insistent sur l'aspect conventionnel et arti-

12. BINET, SIMON, 1905a, p. 163.

13. BINET, SIMON, 1905b, p. 196.

14. FRAISSE, SEGUI, dir., 1994, p. 147-150.

15. SCHNEIDER, 1992, p. 111.

16. BINET, SIMON, 1905a, p. 168.

17. Voir BINET, SIMON, 1905b, p. 191-198, ou BINET, SIMON, 1908, p. 66 *sqq.*

18. BINET, SIMON, 1908. Cette conception de la mesure est exposée dès 1898, voir BINET, 1898.

ficiel de ce principe classificatoire, tout en reconnaissant que c'est l'aspect conventionnel et normalisé qui fait sa vertu.

Quel est donc ce fameux test de Binet-Simon? Il existe plusieurs versions du test. La première version date de 1905, les versions suivantes de 1908 et 1911, mais elles ne diffèrent pas dans l'esprit. Un test est une série d'épreuves (entre 10 et 30) mettant en œuvre différents traits psychiques. À chaque âge (compris entre 3 et 13 ans) correspond une série d'épreuves : les épreuves associées à un âge étant conçues de façon à être exécutées correctement par la majorité des enfants de cet âge. Pour déterminer le niveau intellectuel d'un enfant, il suffit d'examiner « l'âge le plus élevé dont il a accompli toutes les épreuves, avec une tolérance d'un insuccès dans une des épreuves de cet âge¹⁹ ». C'est ce principe que Binet et Simon appellent « l'échelle métrique de l'intelligence » et c'est l'âge ainsi calculé qui est nommé le « niveau intellectuel ». La dernière version publiée de leur test date de 1911, année de la mort prématurée de Binet. Après sa disparition, ses recherches seront progressivement marginalisées et leur champ d'application restera purement médical : aucune utilisation sociale de grande ampleur ne verra le jour en France²⁰.

Le sort réservé aux outils de Binet en Amérique est tout autre. La différence est double : d'une part, l'esprit de la méthode évolue ; d'autre part, l'ampleur des applications est considérable. L'introduction des outils de Binet aux États-Unis est due à un américain, Henri H. Goddard (1866-1957) qui, lors d'un voyage en Europe en 1908, découvre ses travaux. Goddard, qui est depuis 1906 directeur du département de la recherche d'une école pour « faibles d'esprit » dans le New Jersey, traduit et teste les outils de Binet dès 1909.

Pour quelles raisons les travaux de Binet connaissent-ils des changements lorsqu'ils traversent l'Atlantique? Quels changements subissent-ils? L'étude des espaces sociaux et universitaires américains qui accueillent les outils de Binet permet d'apporter des réponses à ces questions.

1.2. L'insertion des outils de Binet aux États-Unis

Importés aux États-Unis par Goddard durant la première décennie du xx^e siècle, les outils de Binet s'inscrivent dans divers espaces sociaux et scientifiques et, naturellement, évoluent au contact de ces espaces. L'identification de ces espaces et l'étude des changements que les outils de Binet vont subir supposent une connaissance minimale de la science psycho-

19. BINET, SIMON, 1908, p. 65.

20. SCHNEIDER, 1992. Ce n'est qu'après 1945 et l'importation des modèles américains de tests que leur usage social se répandra en France.

logique et, par là, des liens entre science et société aux États-Unis. Deux points doivent être soulignés : une tradition d'« ingénierie sociale » et la prédominance de la pensée héréditariste et eugéniste.

1.2.1. « L'ingénierie sociale » de la psychologie américaine

Dans l'espace de la psychologie américaine de la fin du XIX^e siècle et du début du XX^e siècle, la démarche scientifique est indissociable de la recherche d'outils d'action sociale, politique ou technique : les psychologues se conçoivent et sont perçus comme des « ingénieurs sociaux » capables d'améliorer le fonctionnement de la société, de modifier et de gérer au mieux les comportements humains. Les recherches entreprises par les représentants de la psychologie se sont en effet toutes orientées, dès les dernières années du XIX^e siècle, vers des objectifs utilitaires : la consommation, la publicité, l'organisation du travail, la répartition adéquate des tâches dans l'industrie, l'identification des déficients mentaux, la sélection des salariés ou des immigrants, le classement des élèves sont des exemples de processus sociaux dont la psychologie s'empare et tente d'y apporter des réponses pratiques²¹.

Pour des raisons semblables à celles ayant permis à Binet de légitimer ses outils, ou ayant fait évoluer l'économie et la sociologie françaises de l'après-Seconde Guerre mondiale²², la recherche de l'utilité dans la démarche scientifique est au moins autant voulue que subie par les principaux acteurs psychologues américains : en fournissant des outils décisionnels aux décideurs politiques et sociaux et, dans certains cas, en se trouvant en position de décideurs, les psychologues ont pu légitimer leur discipline. L'adaptation et le développement des outils de Binet aux États-Unis n'échappent pas à ce processus.

1.2.2. Eugénisme et héréditarisme dans l'Amérique du début du siècle

Le second trait majeur de la science psychologique américaine du début du siècle est la présence des thèses héréditaristes et eugénistes. Ces thèses prennent leurs racines dans l'Angleterre du XIX^e siècle et en particulier chez Galton, mais l'écho qu'elles rencontrent aux États-Unis est beaucoup plus fort qu'outre-Atlantique, spécialement à travers les lois sur le contrôle de l'immigration et des naissances. Les faits ne manquent pas pour montrer que l'immigration, l'eugénisme et l'héréditarisme sont, dans l'Amérique du XX^e siècle naissant, présents et intimement liés²³. Ces liaisons naturelles dans l'esprit des acteurs de l'époque prennent forme sociale à travers un

21. PAICHELER, 1992, en part., chap. IV ; sur les causes, voir BUSS, 1976.

22. POLLAK, 1976.

23. Le lecteur intéressé par une démonstration complète de ces faits pourra se reporter à PAICHELER, 1992, p. 154 ; GOULD, 1983 ; LEMAINÉ, MATALON, 1985, p. 31.

large tissu de comités, offices, associations ou laboratoires... toutes et tous dévoués aux causes héréditaristes et eugénistes.

Parmi tous les espaces sociaux et universitaires marqués du sceau de « l'ingénierie sociale » et de l'esprit héréditariste, les espaces au sein desquels les travaux de Binet vont précisément s'inscrire nous intéressent tout particulièrement : il s'agit de l'espace des *testing programs*²⁴.

1.2.3. La naissance des testing programs

L'itinéraire de Goddard, traducteur et adaptateur américain des travaux de Binet, s'inscrit pleinement dans cet espace américain de l'héréditarisme, de l'eugénisme, de la politique restrictive de l'immigration et de la reproduction. Il est, depuis 1906, directeur du département de recherche d'une école pour déficients mentaux. Sa notoriété le conduit à être nommé en 1910 secrétaire du Comité d'étude de la faiblesse d'esprit au sein du Committee on Eugenics de l'American Breeders' Association et à être invité à Ellis Island²⁵ afin d'améliorer les procédures d'examen de l'intelligence des immigrants. C'est ainsi que Goddard est à l'origine de l'application systématique des tests de « Binet ».

Ces tests, apportant du « sang neuf à l'étude des différences individuelles²⁶ », vont dès lors voir leurs applications se multiplier grâce, en particulier, à Lewis Terman et à Robert Yerkes : tous deux acteurs centraux dans l'histoire de l'héréditarisme et de l'eugénisme américains, ils ouvrent aux tests d'intelligence la porte de l'utilisation systématique à grande échelle.

Terman, professeur à l'université de Stanford, se consacre dès 1910 à l'étude expérimentale des tests de Binet-Goddard. Il développe un test dit de « Stanford-Binet » (1916) destiné à la mesure de l'intelligence des adultes et utilise la notion de *QI*²⁷. L'ambition de sa démarche est, sans conteste, beaucoup plus vaste que celle de Goddard : les tests constituent pour lui l'outil de démonstration de l'existence d'un lien entre race, intelligence, classe sociale et vice ; lien dont l'existence prouve la nécessité d'accorder la position sociale d'un individu avec son intelligence. Stephen Jay Gould a montré la force de conviction de Terman²⁸.

Yerkes est, lui, président de l'American Psychological Association (APA) lorsque l'Amérique entre dans la Première Guerre mondiale en 1917. Conscients de l'existence de problèmes d'ordre psychologique dans

24. Nous préférons ce syntagme à ceux de *testing program* (singulier cachant la pluralité des applications) ou *mental testing* (expression cachant les aspects institutionnels).

25. Ellis Island est l'île située à l'entrée du port de New York, où les immigrants étaient accueillis avant leur entrée sur le territoire américain.

26. PAICHELER, 1992, p. 160.

27. Quotient intellectuel : rapport entre l'âge mental et l'âge chronologique, ramené à 100.

28. GOULD, 1983.

l'armée, l'APA et le National Research Council²⁹ décident de la constitution d'un Psychological Committee dont la direction est confiée à Yerkes et auquel Terman participe. Plusieurs tests (tests *alpha*, *beta* et individuel) destinés aux recrues sont conçus³⁰. Environ 1 750 000 jeunes recrues sont soumises à ces tests : cette expérience est la première à atteindre une telle envergure.

1.3. Succès et transformations des outils de Binet

Bien entendu, l'insertion et l'utilisation des outils de Binet aux États-Unis ne vont pas aller sans modification des outils. Les changements seront même nombreux : évolution de la conception de l'intelligence, évolution des propriétés attribuées à sa mesure, développement des applications des tests à d'autres grandeurs que l'intelligence. Le second fait marquant de l'introduction des tests de Binet aux États-Unis est le succès qu'ils rencontrent dans l'espace universitaire américain : la psychologie s'empare de ces outils et participe pleinement aux divers *testing programs*.

1.3.1. Intelligence et mesure dans les testing programs

L'histoire du développement des *testing programs* ne vaut, dans notre perspective, que par ce qu'elle révèle de la conception de la mesure de l'intelligence chez les concepteurs et utilisateurs des tests : les tests d'intelligence sont perçus comme capables, effectivement, de trier les individus selon leur « intelligence ». Le réalisme est double : réalisme de l'intelligence en tant que grandeur mesurable ; réalisme du test d'intelligence en tant qu'outil de mesure de cette grandeur.

Les propos tenus par Terman lorsqu'il définit l'intelligence et son principe de mesure dans sa contribution à un symposium consacré à *Intelligence and its measurement* (1921) sont marqués du sceau de ce double réalisme. Il affirme qu'il existe un type d'intelligence qui domine tous les autres types d'aptitudes humaines : l'intelligence abstraite. Plus une personne est douée d'une capacité à raisonner de façon abstraite plus elle est, selon lui, capable de s'adapter à toutes les situations. L'unidimensionnalité et l'additivité (au moins théorique) de l'intelligence abstraite ne semblent pas faire de doute : c'est une grandeur que les individus possèdent à différents degrés et l'opération qui consiste « à ajouter un peu de cette grandeur » à un individu permet de développer ses capacités globales³¹. Le

29. Organisation créée en 1916, placée sous la tutelle de la National Academy of Sciences, et en charge de la coordination de l'activité des recherches dans le pays.

30. Pour une critique de ces tests (conditions de passation, interprétations), voir GOULD, 1983, p. 238-290.

31. Sur tous ces points, voir le SYMPOSIUM, 1921, p. 128-132.

Alfred Binet	Lewis Terman (entre autres)
Intelligence : grandeur multiforme désignant l'ensemble des capacités	Intelligence abstraite : unilinéaire et dominant toutes les capacités (<i>réalisme de l'intelligence</i>)
Pas de théorie générale	Théorie générale de l'intelligence et de son caractère inné : instrument d'une politique eugénique
Diagnostic des problèmes scolaires	Instrument de classement des individus
Seul un souci de simplification fait parler de « mesure de l'intelligence »	Instrument de mesure de l'intelligence (<i>réalisme de la mesure</i>)

TABL. 1. — *Intelligence et mesure de l'intelligence chez Binet et Terman*

tableau 1 permet de synthétiser les oppositions majeures entre la conception de Binet et la conception de Terman (et, au-delà, des principaux utilisateurs américains) et donc de constater combien leurs conceptions de l'intelligence et de la mesure sont éloignées.

1.3.2. *Le succès universitaire des tests*

Il serait naïf de croire que cette conception de l'intelligence et de sa mesure est le seul fait de quelques idéologues marginaux, éloignés de la communauté des psychologues qui, elle, ne se ferait pas d'illusions sur la justesse des positions défendues... Tout montre, au contraire, que la psychologie scientifique américaine du début du xx^e siècle est, à part entière, partie prenante dans ces usages et théories ; les propos tenus par divers psychologues universitaires comme les positions institutionnelles des acteurs des *testing programs* au sein de la communauté des psychologues américains.

En 1921, par exemple, la revue *Journal of educational of psychology* publie un symposium consacré à la définition et à la mesure de l'intelligence. Les participants à ce symposium sont invités à répondre à deux questions. Premièrement, comment concevez-vous l'intelligence et les moyens pour la mesurer ? Deuxièmement, quels progrès souhaitez-vous ?

Ce symposium regroupe des professeurs, des chercheurs dont l'autorité scientifique est reconnue et pour certains connue, de futurs ou d'ex-présidents de l'APA et des universitaires d'origines très différentes. Mais cette hétérogénéité n'empêche pas ces scientifiques d'exposer tous le même point de vue de fond. Aucun d'eux — exception faite de Thurstone — ne remet fondamentalement en cause la conception de l'intelligence, de sa mesure et de la méthodologie de cette mesure.

Pour autant, affirmer que les participants à ce symposium ne souhaitent aucun changement serait erroné : les auteurs veulent voir se développer de nouveaux tests destinés à l'évaluation d'aptitudes particulières (tests de mécanique, musique, vision, logique, tests verbaux, tests pour enfants de moins de trois ans) ou des extensions du principe des tests d'intelligence à la mesure d'autres caractéristiques psychiques (personnalité, caractère, tempérament). De tels tests vont effectivement voir le jour et connaître un développement rapide. Les expressions « *educational measurement* » et « *educational testing* » vont peu à peu se répandre et, dans les années 1920, le *mental testing* regroupera en fait les études de l'intelligence et celles des aptitudes spécifiques. Toutefois, ces améliorations constituent des changements « à technologie constante » : la méthodologie n'est pas remise en cause.

Les positions universitaires et scientifiques des deux principaux acteurs cités jusqu'ici sont en elles-mêmes révélatrices du lien entre la psychologie universitaire américaine et les espaces du *mental testing* : Terman (1877-1956) est docteur de l'université de Clark, professeur de psychologie à Harvard de 1910 à 1942 et président de la toute-puissante APA en 1923 ; Yerkes (1876-1956) est docteur de l'université de Harvard, directeur du laboratoire de psychologie animale de Harvard de 1903 à 1917, en charge de responsabilité de l'administration de la recherche du National Research Council de 1917 à 1924, professeur de psychobiologie à l'université de Yale de 1924 à 1941 et président de l'APA en 1917. Et ces deux exemples ne constituent pas des cas isolés : Terman a lui-même dénombré onze utilisateurs de tests parmi les treize présidents qui se sont succédé à la tête de l'APA entre 1910 et 1921 et « sur les dix-neuf présidents de cette Association, qui se sont succédé depuis sa fondation en 1892 jusqu'en 1910, les noms de neuf d'entre eux se retrouveraient dans n'importe quelle bibliographie complète des tests mentaux³² ».

L'exemple le plus éclairant de la « symbiose » des tests avec la psychologie scientifique américaine est certainement le discours tenu par Terman lors de sa nomination à la présidence de l'APA en 1923 : « Je pense que l'on peut [dire] que la tentative de distinguer la méthode des tests de la

32. TERMAN, 1924, p. 110-112.

méthode expérimentale n'est fondée ni sur des principes logiques ni sur des principes historiques. Le test mental est partie intégrante de la psychologie expérimentale³³. »

1.3.3. *Les outils du mental testing*

Parallèlement aux transformations de la conception de l'intelligence et de sa mesure analysées précédemment, les outils vont eux aussi connaître des évolutions, à commencer par les tests eux-mêmes. Chez Binet, les résultats à chacune des épreuves du test d'intelligence ne sont pas quantifiés (le résultat final est une valeur qualitative ordinale). La seule information que retient Binet est l'échec ou la réussite d'un sujet à une épreuve. La valeur de l'indicateur « niveau intellectuel », exprimé en années, est fonction du succès ou de l'échec d'un sujet aux épreuves associées (*a priori*) à un âge chronologique donné. Lorsque les tests de Binet sont utilisés aux États-Unis, un score (exprimé en points) est associé à chaque épreuve d'un test et le résultat obtenu par un sujet à un test est le total des points des épreuves passées avec succès par le sujet. D'un modèle purement qualitatif, le test mental est passé à un modèle quantitatif.

Dans une enquête, une fois les indicateurs d'intelligence calculés (les scores ou un peu plus rarement les âges mentaux), l'étape suivante est d'analyser les séries d'indicateurs. Dans les années 1910 et 1920, plusieurs outils concepts sont utilisés : les pourcentages (taux de réussite ou d'échec); comparaison des niveaux intellectuels ou des âges mentaux (exprimés en années); les coefficients statistiques exprimant les positions relatives des points dans une distribution de résultats (quantiles, dispersion); moyenne, mode et médiane de la distribution des scores; indicateurs de variabilité... en peu de mots, les outils courants de la statistique descriptive. Bien entendu, la base de calcul de ces indicateurs peut varier : différents groupes de population (sexe, âge, nationalité, niveau scolaire, milieu social, etc.), différents tests...

La préoccupation finale des psychologues est de comparer les intelligences ou les aptitudes des différentes catégories d'individus. Lorsque l'indicateur d'âge mental n'est pas utilisé, les résultats sont très souvent présentés sous la forme de tables (ou de graphiques) de répartition des scores moyens par âge ou par catégories (sexe, nationalité, scolarité) : dans ce cas, la différence entre deux scores moyens sert à la comparaison des résultats de deux groupes différents. Mais, afin d'améliorer la qualité des comparaisons entre les résultats de deux catégories d'individus, les chercheurs utilisent fréquemment la comparaison des quantiles de dispersion.

33. *Ibid.*, p. 110-117.

Cette méthode, proposée par Edward Lee Thorndike en 1904³⁴, a le mérite « de tenir compte de la variabilité et de l'étendue des distributions, ce que ne permet pas la moyenne³⁵ ». Les résultats prennent alors cette forme : « 70 % des individus de tel groupe ont un score supérieur au score médian de tel autre groupe » ou encore « le score atteint ou dépassé par les 10 % meilleurs individus de ce groupe est dépassé par 30 % des sujets de cet autre groupe », etc.

Cette méthode, initialement destinée à comparer des résultats entre groupes d'individus, va devenir au fil des années 1910 une méthode dite d'*educational scaling technique* qui permet les comparaisons entre tests ou entre épreuves de tests³⁶ : elle devient alors un outil d'échelonnement des tests (ou des épreuves), c'est-à-dire un outil pour la construction d'une hiérarchie quantifiée sur les tests (ou les épreuves). Cette méthode est censée apporter une solution au problème de l'étalonnage des tests, c'est-à-dire à la question de la comparaison des résultats obtenus par deux sujets identiques à deux tests différents. Sous l'hypothèse de normalité et d'égalité des variances des distributions des résultats à chacun des tests, la méthode consiste à déterminer la position relative des deux tests sur l'échelle non pas en fonction des scores moyens (rien n'assure que les scores aient la même signification dans les deux tests) mais en fonction des quantiles (ou des écarts à la moyenne mesurés en écarts-types). En effet, si les fréquences des résultats sont distribuées normalement, la simple connaissance des rangs des différents résultats (en oubliant leur cardinalité, c'est-à-dire leur valeur numérique) permet de déterminer un ordre sur les différences (grâce à la table de fonction de répartition de la loi normale). Il suffit de calculer, pour le test A, la proportion des scores des individus ayant obtenu un score situé dans une proportion arbitraire de meilleurs ou de plus faibles scores au test B. Si, par exemple, 30 % des meilleurs scores au test A sont obtenus par des individus ayant obtenu un score supérieur ou égal à 60 % des meilleurs scores au test B, alors le quantile d'ordre 40 % (100-60) de la distribution B sera située au quantile d'ordre 70 % (100-30) de la distribution des scores au test A. Le calcul des quantiles s'effectue à partir de la table de la loi normale de variance égale à celle de A. Il n'est pas tenu compte de la variance de B (point qui sera critiqué par Thurstone).

Au total, l'espace de la psychologie américaine du début du xx^e siècle, dominé par une approche utilitaire de la recherche scientifique et par les

34. THORNDIKE, 1913, 1^{re} éd. 1904. Selon THURSTONE, 1927c, p. 506, il est le premier à avoir préconisé l'usage systématique de cette méthode.

35. THORNDIKE, 1913, p. 128.

36. Selon THURSTONE, 1927c.

paradigmes eugéniste et héréditariste, est très étroitement liée aux *testing programs*. Cette liaison est suffisamment étroite pour qu'un jeune psychologue américain en formation pendant le premier quart de siècle s'insère naturellement dans cet espace et ait une connaissance précise des outils et de leurs usages. C'est, en tout état de cause, le cas du jeune psychologue Louis Léon Thurstone.

2. — THURSTONE ET LA MESURE EN PSYCHOLOGIE

Lorsque Louis Léon Thurstone (1887-1955) débute ses études supérieures à l'université Cornell (Ithaca, NY), le xx^e siècle est un jeune siècle. Sa formation initiale est indéniablement celle d'un ingénieur : il suit un cycle d'ingénieur civil puis d'ingénieur électricien et mécanicien, travaille dans le laboratoire de Thomas Edison, et est chargé en 1912 d'enseigner la géométrie et le dessin technique à l'université du Minnesota.

Cette première formation se double rapidement d'une formation en psychologie et psychophysique : portant de l'intérêt aux dimensions psychologiques du lien homme-machine (« l'ergonomie »), il assiste aux *lectures* de psychologie organisées par Madison Bentley et Edward B. Titchener et acquiert ainsi une première culture psychologique et psychophysique. Lorsqu'en 1917 il obtient son diplôme de docteur en psychologie de l'université de Chicago, son intérêt pour les aspects psychologiques semble en fin de compte l'avoir emporté sur son intérêt pour les sciences de l'ingénieur : durant toutes les années qui suivent, son activité sera consacrée à la psychologie. En 1920, il devient professeur et directeur de la division de psychologie appliquée du Carnegie Institute of Technology. De janvier 1923 à l'été 1924, il travaille à Washington pour le gouvernement américain à l'amélioration des examens organisés par le service public. Parallèlement, il mène des recherches sur l'apprentissage durant le sommeil pour la Marine. En 1924, il est nommé professeur associé à l'université de Chicago puis, en 1925, professeur titulaire. C'est au sein de cette université que sa carrière se déroulera jusqu'en 1952 : son nom reste indéniablement attaché à cette université dans laquelle il mènera tous les travaux et toutes les actions qui feront sa gloire (analyse factorielle multidimensionnelle, loi du jugement comparatif, fondation de *Psychometrika* et de la société de Psychométrie). De 1952 à sa mort, en 1955, il est professeur et directeur du laboratoire de psychométrie de l'université de Caroline du Nord.

Au-delà de ces éléments biographiques factuels, quels sont les espaces intellectuels, scientifiques et sociaux dans lesquels s'inscrit le parcours de

Thurstone? L'étude de ces espaces est nécessaire à une bonne connaissance de ce qu'étaient, pour un jeune psychologue américain du début du siècle, la psychologie et la mesure en psychologie; et cette bonne connaissance est elle-même nécessaire à une bonne compréhension du rôle joué par Thurstone dans l'espace de la psychologie.

2.1. *Espaces institutionnels et intellectuels de Thurstone*

Deux faits marquants, indispensables à notre reconstruction historique doivent être soulignés : l'insertion de Thurstone dans les *testing programs* et sa connaissance précise des outils statistiques et probabilistes.

En 1915, Thurstone est nommé assistant à la division de psychologie appliquée du Carnegie Institute of Technology. La filiation de cet institut avec les différents comités mis en place au sein de l'armée à partir de 1917 ne fait aucun doute : les principaux acteurs de l'institut (Bingham, Scott, Thurstone, Strong, Wipple, Yoakum...) se sont retrouvés, une fois les États-Unis engagés dans la guerre, à différents postes de responsabilité dans les différents comités mis en place pour l'armée. Au-delà de cette filiation entre le Carnegie Institute et le Psychological Committee, ce sont des liens entre les principaux acteurs des *testing programs* et de l'Institut dont nous pouvons être sûrs : Goddard, Terman, Yerkes, Bingham et Thurstone sont les membres d'un réseau tissant sa toile dans l'espace des *testing programs*, du Carnegie Institute, de l'université et de l'APA. En particulier, Thurstone est affecté par l'armée à l'étude des problèmes d'apprentissage, des méthodes d'instruction et de sélection.

À la différence de bon nombre de ces acteurs, Thurstone possède une connaissance des outils statistiques. Ingénieur et physicien de formation, ayant suivi les cours de statistique de Titchener³⁷, il a enseigné les statistiques au Carnegie Institute et, à la demande de Harvey Carr, professeur de psychologie, à l'université de Chicago (1924). D'ailleurs, l'arrivée de Thurstone à Chicago coïncide avec le développement d'un courant de pensée quantitatif au sein de cette université³⁸. Par le simple fait de devoir enseigner ces outils, Thurstone en a certainement une connaissance plus claire que celle des simples praticiens et le regard qu'il y porte est à coup sûr plus aiguisé. De plus, sa connaissance initiale des outils n'ayant pas pour unique source leurs applications aux tests mentaux, il est cer-

37. WALKER, 1929, p. 154-157.

38. Comme l'a montré BULMER, 1981, l'introduction de l'enseignement des statistiques dans le cursus de formation de psychologie n'est qu'une conséquence du développement de l'approche quantitative dans cette université. L'importance de cette approche au sein de l'université de Chicago à partir des années 1920 ne doit pas, selon Bulmer, être négligée : elle touche toutes les disciplines des sciences humaines.

tainement conscient du caractère pas nécessairement naturel des liens entre les outils et leurs applications, voire du caractère problématique de ce lien. En tout cas, il perçoit ses connaissances suffisamment sûres pour pouvoir, dans les années 1920, développer des critiques à l'encontre des outils en usage.

2.2. *Sa critique des outils de mesure*

Au début de ces mêmes années 1920, alors qu'il va devenir professeur associé à l'université de Chicago, Thurstone prend conscience des grandes imperfections des outils utilisés alors dans les divers *mental testing programs*, de la faiblesse, voire de l'absence de théorie de ces outils, ainsi que du peu d'intérêt porté à ces outils par les théoriciens :

« Mon souci était principalement tourné vers l'enseignement de la théorie des tests mentaux. Ce sujet était enseigné dans tous collèges américains principalement à partir des divers manuels, et en pratique tous les cours se limitaient à l'enseignement du test Stanford-Binet. Ni les enseignants ni les étudiants ne semblaient faire preuve d'intérêt pour la théorie des tests [...] Je décidais de chercher à contribuer à l'amélioration de cette situation et dès lors j'eus comme objectif de travailler sur les problèmes fondamentaux de la mesure en psychologie³⁹. »

Pour éclairant qu'ils soient, ces propos ne doivent pas éblouir : issus de l'autobiographie de Thurstone, ils datent de 1952, soit près de trente ans après les faits, et sont donc suspects de « rationalisation *a posteriori* ». Éviter le piège de l'anachronisme suppose de revenir aux écrits de Thurstone, à leur évolution et à leur signification historique. En fait, dès 1921, Thurstone formule quelques critiques fondamentales à l'encontre des tests, de leurs principes et de leurs fondements. Plusieurs années seront cependant nécessaires pour que ces critiques deviennent plus précises, plus ciblées et que Thurstone propose des outils et des méthodes alternatives. Chemin faisant, ses critiques vont d'ailleurs changer de nature : initialement de nature plutôt épistémologique, elles vont devenir, au milieu des années 1920, des critiques techniques de l'outillage statistique.

2.2.1. *Une voix discordante : celle de Louis Léon Thurstone*

Nos analyses précédentes ont permis de montrer que la conception réaliste de l'intelligence était suffisamment partagée pour qu'à une seule

39. THURSTONE, 1952, p. 304, c'est nous qui soulignons. En 1937, dans son discours de fin de présidence de la Psychometric Society, Thurstone tient des propos en tous points identiques.

exception près, tous les participants à un symposium publié en 1921 — date située au cœur de l'histoire des usages sociaux des tests mentaux — exposent une conception identique de l'intelligence et les mêmes outils peu théorisés. L'exception dont il est question est Thurstone.

Au début des années 1920, et en tout cas dans ce symposium, Thurstone semble bien être le seul à percevoir (ou en tout cas à ne pas se cacher) les difficultés rencontrées dans la mesure de l'intelligence, difficultés d'ordre théorique et fondamental et non simples obstacles pratiques dus aux aléas des expériences :

« Maintenant que nous avons accumulé assez de données pour nous donner le vertige, il est temps que nous commençons à formuler des principes pour les tests mentaux. La littérature abonde d'instructions spécifiques à chaque test et de données statistiques destinées à l'administration. Mais, jusqu'ici, *il n'y a pas de science, pas de principes, pas de psychologie dans la littérature sur les tests mentaux*. [...] Il m'a toujours semblé curieux que l'intelligence des enfants soit mesurée en terme d'âge. Je voudrais suggérer que les hypothèses sous-jacentes à la mesure paraissent moins troublantes si nous mesurons l'intelligence d'un enfant de huit ans en terme de rang exprimé en centiles par rapport à l'ensemble des enfants de huit ans, ou en terme d'écart type des scores au test pour les enfants de huit ans⁴⁰. »

Il affirme en substance que les outils de mesure de l'intelligence sont mal conçus, qu'ils manquent de « principes, de théorie et de sens psychologique ». Que Thurstone soit la seule « fausse note » dans un concert d'unanimité pour les tests mentaux peut paraître surprenant car les difficultés pratiques ne manquent pas : problèmes dans l'échantillonnage des tests, problèmes lors de la comparaison des résultats... Deux exemples symptomatiques — auxquels Thurstone cherchera plus tard à apporter une solution — méritent l'attention.

Le premier exemple de difficulté est celui que rencontre Edwin G. Boring, alors assistant de Yerkes, dans l'exploitation statistique des résultats des tests dans l'armée⁴¹ : le problème de la définition d'une origine des échelles. Pour expliquer que certains individus obtiennent un score nul à des tests censés mesurer l'intelligence innée, Boring affirme qu'il était nécessaire de distinguer un « zéro mathématique » et le « zéro psychologique ». Le premier est le minimum absolu en dessous duquel il est impossible de descendre ; le second est un zéro arbitraire déterminé par un test donné. Boring étalonna donc les différents tests afin de faire correspondre, pour un individu, ses résultats aux divers tests.

40. THURSTONE, 1921, p. 206.

41. GOULD, 1983, p. 267.

Le second type de problème est celui qui se pose aux utilisateurs des tests lorsqu'ils souhaitent savoir si la difficulté complémentaire pour passer d'un résultat R au résultat $R + \Delta R$ est identique à la difficulté complémentaire nécessaire au passage de R' à $R' + \Delta R$ (avec $R \neq R'$). La méthode dite *educational scaling technique* apporte une réponse imparfaite à cette difficulté. Sa connaissance et son utilisation ne sont, de toute façon, pas très répandues : le plus souvent, les psychologues éludent la question.

Les premières critiques de Thurstone, celles évoquées ci-dessus, sont relativement peu précises. Elles vont progressivement le devenir au cours des années 1920 : ce sera, autour de 1925, la critique des concepts d'âge mental et d'unité de mesure. Mais, se précisant, ses objections vont quitter le terrain de l'épistémologie, de la logique et du sens psychologique des tests pour s'inscrire sur le terrain de l'analyse de la qualité statistique et probabiliste des outils d'analyse des résultats. À partir du milieu des années 1920, ses analyses n'auront jamais pour objet les tests et leurs principes générateurs mais seulement les aspects logiques et mathématiques des méthodes d'analyse des résultats aux tests.

Dans le même mouvement, il proposera des améliorations aux outils de mesure, améliorations qui lui permettront de résoudre le problème de la définition d'un repère absolu dans les échelles d'intelligence (zéro absolu). En fin de compte, ses réflexions vont le mener à proposer une nouvelle loi permettant de justifier les pratiques de mesure de toutes les grandeurs psychologiques.

Un exposé chronologique de ses observations est malaisé, voire impossible : les délais de publication aidant, la chronologie de ses travaux est très difficilement périodisable. La présentation de ces développements, produits d'un unique mouvement historique et d'un processus intellectuel unitaire, est donc nécessairement plus artificielle que la réalité historique.

2.2.2 La critique du concept d'âge mental

Dans un de ses premiers articles (1926a) où il développe une critique des pratiques de mesure, Thurstone attaque vivement un des concepts centraux des tests mentaux : l'âge mental. Sa condamnation repose sur une analyse des usages de cette notion ; usages qu'il estime inconsistants et illogiques car faisant appel à deux interprétations distinctes et logiquement incompatibles de l'âge mental.

Son argumentation repose sur la distinction entre les deux définitions possibles de l'âge mental : soit l'âge mental d'un test est défini comme l'âge chronologique moyen des personnes qui atteignent le niveau de performance de ce test ; soit l'âge mental d'un niveau de performance donné est défini par l'âge chronologique des personnes qui réussissent en moyenne à passer avec succès ce niveau de performance. Ces deux

définitions ne sont pas équivalentes et le choix de l'une ou l'autre n'est pas indifférent lors du calcul de l'âge mental d'un individu. Or, écrit en substance Thurstone, comment est utilisée en pratique cette notion numérique d'âge mental ? Une fois les résultats aux différents tests recueillis, la procédure d'étalonnage débute : on calcule la moyenne des résultats obtenus par chaque tranche d'âge ; ce calcul fournit donc un tableau à deux colonnes dont celle de gauche contient les différents âges (ou tranches d'âge) et celle de droite, en vis-à-vis, les résultats moyens obtenus par chaque âge. Thurstone nomme cette opération « la régression des tests sur les âges ». Une fois le test étalonné, il est possible de calculer l'âge mental d'une personne en déterminant quel âge correspond, dans le tableau évoqué ci-dessus, à la performance de la personne. L'âge obtenu est l'âge mental de la personne considérée.

La simple description de cette procédure laisse entrevoir « l'imperfection logique » : le tableau utilisé, celui de la « régression des tests sur les âges », doit être lu, par construction, de la gauche vers la droite. Le lire de la droite vers la gauche est, dans la pure logique, un non-sens. Un peu, écrit Thurstone, comme si nous estimions la taille d'un individu en fonction de son seul poids en utilisant un tableau donnant les poids moyens en fonction de la taille des individus et non pas un tableau fournissant la taille moyenne en fonction du poids. Certes, affirme-t-il en substance, il suffit de se tenir à l'une des deux définitions qui, tout en étant imparfaite, à le mérite d'harmoniser les usages et d'établir une convention. Mais, à ses yeux, le plus simple est certainement de remplacer l'âge mental, qui repose sur le calcul d'une valeur moyenne quelle que soit la définition, par des indicateurs de dispersions permettant de qualifier la position relative d'un enfant (en fait, de son score) par rapport à un groupe (dont la distribution des scores est connue).

Plus précisément, Thurstone préconise l'utilisation des quantiles pour caractériser l'intelligence d'un individu par rapport à sa classe d'âge ou à une classe d'âge quelconque. Par exemple, plutôt que d'affirmer qu'un enfant a un âge mental de 11 ans alors qu'il est âgé de 10 ans, il préfère comparer le score d'un enfant par rapport à l'ensemble des scores des enfants d'une classe d'âge donnée et affirmer que son score fait partie des 20 % des scores les plus élevés de la classe d'âge considérée. En cela il rejoint la méthode préconisée par Thorndike pour comparer les résultats de différents groupes. Thurstone ne fait, en quelque sorte, qu'appliquer une idée identique aux individus et plus seulement aux groupes.

Au total, il propose donc rien de moins que d'écarter définitivement la notion d'âge mental, qu'il qualifie « de mal conçue, de mal aisée (*awkward*) », au profit d'indicateurs de dispersion permettant de qualifier les positions relatives des individus.

2.2.3. La critique de « l'unité de mesure »

L'utilisation de ces indicateurs de dispersion doit cependant obéir à des règles précises. Constituant les « unités de mesure » de l'intelligence, il est nécessaire de garantir leur universalité, c'est-à-dire leur capacité à s'appliquer à tous les groupes ou à tous les tests. C'est en tout cas le souci que développe Thurstone dans son premier article sur les unités de mesure (1925). Thurstone ne remet pas en cause, sur le fond, la méthode généralement employée (méthode de Thorndike). Il critique simplement, dès 1925, l'hypothèse selon laquelle la dispersion des résultats aux tests est identique pour chaque groupe. À ses yeux la comparaison des différences entre deux quantiles différentes dans deux groupes distincts n'a pas de sens : comment comparer les écarts entre les quantiles d'ordre 80 % et 90 % de deux distributions si les dispersions sont différentes ?

La méthode proposée par Thurstone pour surmonter cette difficulté est exposée dans l'article au titre évocateur « A method of scaling psychological and educational tests » (1925). Sa méthode de mesure (échelonnement) permet de construire une unité de mesure « universelle », c'est-à-dire valable pour toutes les échelles mentales et autorisant donc leur comparaison. Elle repose simplement sur la prise en compte du coefficient de dispersion, c'est-à-dire de l'écart-type pour « normer » les distributions. Son nom est révélateur, à lui seul, de la capacité que Thurstone lui accorde à surmonter la difficulté : *absolute scaling method*.

2.3. Le renouvellement de la mesure en psychologie

2.3.1. Une méthode de mesure améliorée

L'*absolute scaling method* permet, à partir des résultats des sujets à différentes questions, de « localiser ces questions ou les résultats des groupes de sujet sur la même échelle⁴² ». Elle s'appuie sur quelques considérations probabilistes élémentaires et sur la figure 1. Pour un groupe d'individus i , Thurstone définit M_i comme la moyenne théorique des scores de ce groupe et, pour une question donnée, X_i comme l'écart théorique (exprimé en écart-type) entre M_i et cette question. Pour une même question (point 0 sur la figure 1) et deux groupes différents (1 et 2), la relation $M_1 + X_1\sigma_1 = M_2 + X_2\sigma_2$ est nécessairement vérifiée (puisque une même question a une et une seule position sur l'échelle). Cette relation affirme l'existence *théorique* d'une relation linéaire entre les positions *théoriques* X_1 et X_2 d'une question relativement à deux « référentiels » (deux groupes). En pratique, cette relation n'est pas exactement vérifiée mais, grâce à une régression, il

42. THURSTONE, 1925, p. 438.

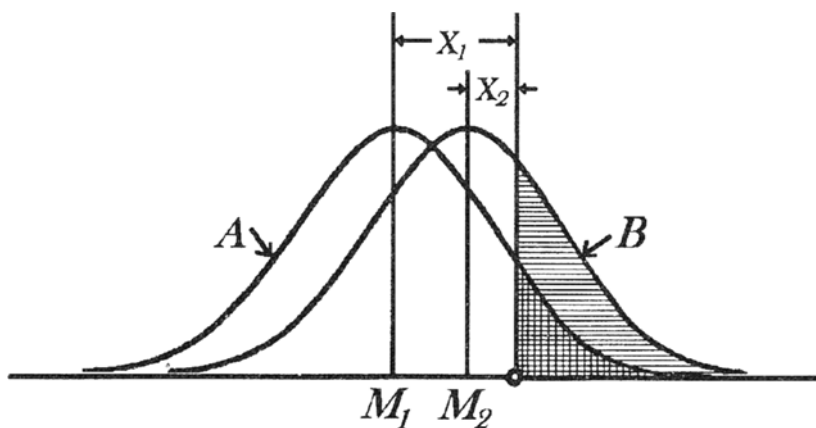


FIG. 1. — *Illustration de la méthode d'échelonnement absolu de Thurstone.*

est possible d'estimer empiriquement les coefficients permettant, une fois connus l'écart-type et la moyenne des scores d'un groupe, d'obtenir la moyenne des scores et de l'écart-type d'un autre groupe exprimés dans le même « référentiel » (la même échelle). En d'autres termes, une fois choisies une origine et une unité de mesure⁴³, il est possible de déterminer la position de tous les groupes (ou de toutes les questions) dans ce référentiel. Voilà donc réalisé le souhait de Thurstone : disposer d'une méthode permettant de mesurer (ou d'échelonner) de façon absolue groupes et/ou questions. En 1926, Thurstone publie le premier article qui utilise empiriquement cette méthode d'échelonnement. Beaucoup d'autres suivront.

Parmi les applications de la méthode, l'une d'entre elles a une importance symbolique particulière : en 1928, Thurstone apporte une solution à une difficulté parfois rencontrée par les utilisateurs de tests, celle de la définition d'un zéro absolu. Ce problème se pose de façon pratique dès que l'on veut représenter graphiquement des résultats et, théoriquement, dès que l'on veut disposer d'une échelle universelle pour pouvoir comparer des résultats. En 1928, dans l'article « The absolute zero in intelligence measurement », Thurstone affirme avoir découvert, en utilisant sa méthode de mesure absolue, une relation linéaire entre l'écart-type et la moyenne des résultats aux tests de plusieurs groupes d'âge. Il déduit de cette relation linéaire la position du zéro absolu de l'intelligence en considérant que la variation absolue (écart-type calculé sur l'échelle absolue) doit être nulle

43. C'est à partir de 1927 que THURSTONE, 1927c, désignera l'écart-type d'une distribution comme une « unité de mesure idéale ».

au point d'intelligence absolue nulle⁴⁴. Il en déduit ensuite la courbe du développement mental absolu des enfants, c'est-à-dire la courbe liant l'âge chronologique et le résultat absolu aux tests. Le calcul de l'âge « d'intelligence nulle » sur plusieurs tests permet de situer cet âge entre quelques mois avant la naissance et le moment de la naissance : preuve est donc faite que le développement intellectuel commence juste avant la naissance. Au-delà de l'intérêt de sa découverte, ce résultat fournit donc à Thurstone une belle confirmation de la pertinence de sa méthode d'échelonnement absolue.

La découverte du zéro absolu de l'intelligence, quels que soient sa valeur anecdotique pour l'historien des sciences ou son intérêt réel, ne doit pas cacher d'autres évolutions dans le travail de Thurstone dont la portée pour l'histoire de la mesure en psychologie est moins éphémère. En effet, le développement de sa méthode d'échelonnement est accompagné de trois mouvements simultanés d'une importance considérable pour l'histoire du développement de la mesure en psychologie. Le premier de ces mouvements est la modification du statut de la méthode : de « méthode » destinée à échelonner, elle devient « loi » destinée à modéliser les processus de réponse. La seconde évolution est l'élargissement de la loi et de la méthode (qui se confondent formellement) à l'ensemble des processus psychologiques. Le troisième et dernier mouvement est la multiplication des utilisations expérimentales de la méthode et de la loi. Ces trois mouvements se déroulent selon la même unité de temps — la fin des années 1920 — et sont certainement le produit d'une volonté unitaire chez Thurstone — appliquer sa méthode à toutes les situations psychologiques. Toutefois, pour les commodités de la présentation, leurs exposés seront séparés.

2.3.2. *Évolution du statut : de la méthode à la loi*

Ce qui ne constitue qu'une simple méthode d'échelonnement dans l'article de 1925 devient, dans deux articles de 1927⁴⁵, la « loi du jugement comparatif ». L'examen de l'appareillage statistique mis en œuvre dans la loi permet d'établir sans ambiguïté leur proximité formelle, voire leur filiation : dans chacun des cas, la loi normale est utilisée pour représenter une distribution de grandeurs psychologiques caractérisées par leur dispersion et leur moyenne puis, à l'aide de ses propriétés, pour établir des résultats de nature probabiliste sur la distribution des grandeurs psychologiques ; dans chacun des cas, l'unité de mesure est définie comme l'écart-type de la distribution ; dans chacun des cas, les raisonnements reposent sur des calculs de quantiles (ou de distances exprimées en écart-type).

44. Cette idée est certainement le résultat de la culture statistique ou physique (cinétique des gaz) de Thurstone.

45. THURSTONE, 1927a et 1927b.

Au-delà de sa ressemblance formelle avec la méthode d'échelonnement, deux changements significatifs marquent l'apparition de la loi : changement de sens (statut épistémologique) et changement de portée pratique. D'un côté, la méthode d'échelonnement absolue est une technique, un simple outil permettant de traiter et d'améliorer l'analyse des résultats aux tests mentaux en « normalisant » les scores ; aucune hypothèse relative au comportement individuel des sujets soumis au test n'est faite. D'un autre côté, la loi est une théorie, un modèle psychologique : elle vise à « expliquer » ou à « modéliser » les réponses individuelles aux tests et plus exactement les processus psychologiques individuels de réponse aux stimuli (questions, épreuves).

Selon la loi, lorsqu'un individu est confronté à plusieurs stimuli de quelque nature que ce soit, la comparaison s'effectue selon une dimension psychologique que Thurstone désigne par les expressions *psychological continuum* ou encore *psychological scale*. L'individu met alors en œuvre un « processus discriminant » permettant de distinguer les stimuli les uns des autres. La loi affirme en substance que ce processus n'est pas « déterministe » : face à un *stimulus*, un individu réagit en mettant en œuvre un processus discriminant particulier mais inconnu *a priori*. Cependant, les processus discriminants sont supposés être distribués normalement le long du *psychological continuum* : leur dispersion est la *discriminal deviation* et est conçue comme une caractéristique du *stimulus* et non des individus⁴⁶. Ainsi, si deux stimuli A et B sont soumis à un individu, celui-ci met en œuvre deux processus discriminants *a* et *b* du même *continuum* et si, aux yeux de cet individu, A est préféré ou jugé supérieur à B alors (*a - b*) sera positif ; sinon (*a - b*) sera négatif.

Sous ces hypothèses, Thurstone établit la relation liant les deux processus discriminants en œuvre dès lors que deux stimuli peuvent être comparés par un individu. Cette relation constitue en fait l'équation de « la loi du jugement comparatif » :

$$S_1 - S_2 = x_{12} \cdot \sqrt{\sigma_1^2 + \sigma_2^2 - 2r\sigma_1\sigma_2},$$

où S_1 et S_2 sont les positions des deux processus discriminants mis en œuvre dans le jugement comparatif des stimuli 1 et 2 ; x_{12} est le quantile associé à la probabilité p_{12} que 1 soit préféré à 2 ; σ_1 et σ_2 sont les écarts-types ; et r est la corrélation entre les deux processus.

Bien entendu, cette construction et les définitions qui l'accompagnent sont purement théoriques : tout en qualifiant sa construction de « loi », Thurstone est parfaitement conscient de l'artifice modélisateur qu'elle ren-

46. Attribuer la variabilité au *stimulus* et non pas au sujet est certainement une attitude héritée de la psychophysique : les sujets soumis aux tests ne valent pas par leurs caractéristiques différentielles mais seulement par leur capacité à représenter l'ensemble des individus.

ferme et de la part de définition qu'elle contient⁴⁷. En particulier, la normalité de la distribution des processus discriminants constitue une définition : la dimension psychologique et la métrique sont définies et choisies de façon à rendre normale la distribution des processus. Le caractère artificiel de sa loi et la conscience que Thurstone en a sont renforcés par son refus de vouloir fournir une explication quant à la nature intrinsèque des phénomènes dont sa loi rend compte : dès son premier article, il se refuse explicitement⁴⁸ à s'avancer sur le terrain de l'interprétation de sa loi en termes psychologiques, physiologiques ou biologiques et il ne changera pas d'avis plus tard.

De prime abord, à la différence des raisons qui ont guidé la naissance de la méthode d'échelonnement absolu, il ne semble exister aucune raison pratique poussant Thurstone à imaginer sa loi du jugement comparatif. Quelles peuvent donc bien être ses motivations ?

La portée pratique de sa loi, telle que Thurstone la conçoit initialement, est considérable car elle fournit un cadre général aux méthodes d'échelonnement et un cadre théorique aux lois psychophysiques de Gustav Fechner et E. H. Weber : « La loi du jugement comparatif s'applique à la loi de Weber comme à celle de Fechner. [Cette loi] s'applique non seulement à la comparaison des stimuli physiques mais également aux jugements comparatifs tels que ceux utilisés dans les échelles éducatives⁴⁹. » Autrement dit, Thurstone voit dans sa loi un cadre « essentiel » pour l'étude de tous les processus psychologiques et pas seulement, comme ses préoccupations précédentes le laissaient présager, les processus mobilisés dans les tests mentaux. Sa loi a pour objet de lier, dans un même cadre théorique et pratique, la voie suivie par la psychophysique et la voie prise par les tests mentaux.

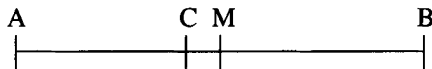
Pour quelles raisons Thurstone est-il conduit à s'intéresser à la psychophysique et à lui fournir un cadre commun avec le *mental testing* ? La réponse se situe tout simplement dans son itinéraire biographique et dans les propriétés communes aux deux « cultures » psychologiques. Avant de s'intéresser aux outils du *mental testing*, Thurstone a acquis une bonne connaissance des outils et de la philosophie de la psychophysique allemande. Les esprits de la psychophysique et de la psychologie des tests mentaux sont assurément incomparables mais il est vraisemblable que ses recherches sur les outils du *mental testing* et ses connaissances des outils de la psychophysique ont conduit Thurstone à découvrir ou percevoir des similitudes purement formelles entre les outils des deux « cultures » et donc à chercher à les unifier dans un même cadre théorique.

47. THURSTONE, 1927b, p. 274.

48. THURSTONE, 1927a, p. 369.

49. THURSTONE, 1927b, p. 273, 276.

Plus de précision s'impose et un retour aux années durant lesquelles Thurstone découvre la psychologie et commence à y porter un intérêt est nécessaire. Lorsqu'il débute ses études d'ingénieur à l'université Cornell, Thurstone s'intéresse aux « aspects psychologiques des machines techniques », c'est-à-dire aux relations des dimensions technico-physiques des machines avec les dimensions psychologiques des utilisateurs. Les questions qui le préoccupent relèvent de ce qui constitue à l'époque la psychophysique et il est naturellement mené à suivre les *lectures* que deux professeurs de psychologie de renom consacrent à cette discipline : Edward Titchener et Madison Bentley. L'un comme l'autre appartiennent à la tradition de la psychophysique allemande⁵⁰. Deux traits saillants, au moins, caractérisent la psychophysique telle que Thurstone la découvre. D'une part, les tests sont en effet au cœur de la méthodologie des expériences de psychophysique et sont formellement proches des tests mentaux⁵¹. D'autre part, les outils probabilistes de la « théorie des erreurs » du XIX^e siècle sont utilisés pour rendre compte des aspects non déterministes des comportements psychologiques. C'est à Fechner, semble-t-il, que revient le mérite d'avoir introduit la théorie des probabilités dans le but de disposer « d'une mesure de la sensibilité des individus qui ne dépende pas de façon critique [déterministe] des stimuli⁵² ». Stephen M. Stigler a conduit une analyse fouillée de la démarche probabiliste entreprise et présentée par Fechner dans son célèbre ouvrage *Elemente der Psychophysik* publié en 1860 à Leipzig⁵³. Il montre que Fechner utilise la loi normale pour « modéliser » les erreurs de jugement des sujets, c'est-à-dire les écarts entre la vraie réponse et la réponse fournie par le sujet. Fechner a utilisé, en 1860, la loi normale pour modéliser les erreurs dans les estimations des poids et des différences entre poids. Mais l'illustration la plus éclairante de l'usage de la théorie des erreurs dans le cadre de la psychophysique est certainement l'expérience imaginée par un de ses collègues, le célèbre August Ferdinand Möbius. L'expérience consiste à présenter à un sujet une ligne AB sur laquelle un point C est placé, puis à demander au sujet de juger quel est, des points A et B, le point le plus proche de C. En théorie, ce jugement revient à estimer la distance algébrique entre C et le point milieu M du segment de droite AB. Möbius, et avec lui Fechner, supposent que les erreurs



50. Titchener est un des plus fameux élèves de Wundt et a traduit les travaux allemands en anglais. Bentley est un ancien élève de Wolf, lui-même ancien élève de Wundt.

51. D'autant plus proches que, depuis Binet, les épreuves élémentaires composant les tests mentaux ont eu tendance à se simplifier pour devenir des questions à choix dichotomique.

52. STIGLER, 1986, p. 246.

53. Voir STIGLER, 1986, chap. VII.

commises par le sujet sur l'estimation de la longueur CM sont distribuées selon la loi $N(0, 1/\sqrt{2})$. Si h est la sensibilité du sujet, si F_0 est la fonction de répartition de la loi normale centrée réduite, et si CM est en réalité positive, alors le sujet commettra une erreur de jugement s'il estime la distance CM inférieure à $-CM$. La probabilité d'un tel événement s'écrit :

$$\frac{h}{\sqrt{\pi}} \int_{-\infty}^{-CM} e^{-h^2 t^2} dt = 1 - F_0(h.C.M. \sqrt{2}).$$

L'estimation de la sensibilité d'un individu est donc donnée par la relation :

$$h.C.M. = \frac{1}{\sqrt{2}} \times F_0^{-1} (\text{taux de bonnes réponses}).$$

Cette dernière formulation s'apparente à la loi du jugement comparatif et, au-delà, à la démarche systématiquement suivie par Thurstone dans ses recherches sur la mesure : estimer des grandeurs à partir des quantiles associés aux fréquences des réponses correctes.

Au total, au moins sur les deux points évoqués précédemment (utilisation de tests et des outils probabilistes), la psychophysique et la psychologie des tests mentaux sont formellement très proches. Issu des deux « cultures », possédant une connaissance universitaire des outils probabilistes, il n'y a rien d'étonnant à ce que Thurstone ait voulu concilier les préoccupations de la psychophysique et celles du *mental testing* ; rien d'étonnant à ce que la proximité des outils formels utilisés dans les méthodologies des deux psychologies l'ait incité à unifier les deux séries d'outils — que cette proximité soit réelle ou partiellement imaginée. De toute façon, quoi qu'il en soit des causes de sa naissance, la loi du jugement comparatif apparaît comme le produit des deux psychologies à la fois sur le fond et sur la forme. Sur le fond, sa loi rejoint les préoccupations des psychophysiciens : lier ou modéliser les liaisons entre les stimuli et les réponses des sujets. Sur la forme, c'est-à-dire techniquement, les outils utilisés dans la loi du jugement comparatif sont ceux déjà utilisés par Fechner ou par Thurstone dans sa méthode d'échelonnement, selon une mise en ordre proche de cette méthode.

Bien entendu, en offrant un cadre identique aux deux psychologies, Thurstone transforme le sens et la portée des outils utilisés dans chacune d'elles. Par exemple, à la différence de Fechner, Thurstone ne parle pas d'erreur : la dispersion des processus psychologiques mis en œuvre par un même individu pour un *stimulus* identique semble constituer chez Thurstone une caractéristique naturelle de chaque *stimulus*. Chez Fechner, un individu est caractérisé par sa sensibilité (h) et la dispersion de la distribution des réactions d'un individu à un *stimulus* — c'est-à-dire ses erreurs —

est facteur de sa sensibilité h . Chez Thurstone, la variabilité est attribuée au *stimulus*. Techniquement, rien n'est changé mais l'interprétation est bien différente : la dispersion est passée du statut d'erreur à celui de caractéristique intrinsèque attachée au *stimulus*.

Cependant, la transformation la plus importante introduite par Thurstone lors de l'élaboration de sa loi se situe dans le rapprochement des processus psychophysiques tels que les psychophysiciens les perçoivent et les processus de la psychologie de tests mentaux tels que les psychologues du *mental testing* les perçoivent. D'ailleurs, cette volonté d'unification ne va pas s'arrêter là : dès 1927, et surtout à partir de 1928, Thurstone publie en effet une série d'articles élargissant sa loi à l'étude des opinions, des attitudes ou des valeurs. En d'autres termes, il généralise sa loi à tous les processus psychologiques couramment analysés à l'époque ; ce qui fait dire à Maurice Reuchlin que « le modèle proposé s'applique à tous les stimuli susceptibles d'être comparés⁵⁴ ».

2.3.3. Élargissement de la loi à toute la psychologie humaine

C'est à partir de 1927 et à travers plusieurs publications⁵⁵ que Thurstone ouvre le champ d'application de sa loi à toute la psychologie humaine ; plus précisément à toutes les recherches sur les processus psychiques pouvant s'étudier selon un schéma « questions-réponses » ou « stimuli-réactions » comme les opinions, les attitudes, les valeurs et, bien sûr, l'intelligence et les capacités sensorielles. L'utilisation répétitive, à travers tous ses écrits postérieurs à 1927, du terme « psychophysique » pour qualifier sa loi est d'ailleurs révélatrice de la capacité que Thurstone attribue à sa loi de rendre compte de tous les processus psychologiques : sa loi concerne tous les processus de base (c'est-à-dire de nature psychologico-physique ou biologique) de l'activité mentale ; elle est universelle.

Il apparaît, par ailleurs, que le dispositif formel emprunté à la méthode d'échelonnement absolue (1925) pour concevoir le cadre de la loi du jugement comparatif reprend une grande partie de sa signification et de sa portée initiale : ce qui ne constituait qu'une loi très générale et abstraite dans les articles les plus théoriques de Thurstone devient (ou redevient) à travers son utilisation empirique une méthode d'échelonnement pratique. Dans les écrits de Thurstone, « loi » et « méthode » se confondent désormais pour ne

54. REUHLIN, 1962, p. 118.

55. THURSTONE, 1928, « Attitude can be measured », *American Journal of sociology*, vol. 33, p. 529-554 ; 1928, « The measurement of opinion », *Journal of abnormal and social psychology*, vol. 22, p. 415-430 ; 1929, « Theory of attitude measurement », *Psychological Review*, vol. 36, p. 222-241 ; 1931, « The measurement of social attitudes », *The Journal of abnormal and social psychology*, vol. 26, p. 249-269 ; 1954, « The measurement of values », *Psychological Review*, vol. 61, p. 47-58. Certains de ces articles ont été réédités in THURSTONE, 1959.

constituer qu'une « loi-méthode » et si Thurstone n'emploie pas un tel syntagme, il utilise néanmoins de façon alternative les qualificatifs « loi » et « méthode ».

2.3.4. *L'utilisation expérimentale de la « loi-méthode »*

Exception faite d'un article de transition (1927e), le premier véritable article d'application pratique, non pédagogique et non illustrative, de la « loi-méthode » de Thurstone paraît en 1928 sous le titre « An experimental study of nationality preferences » (1928b) et le premier exemple d'utilisation devenu célèbre est la recherche menée par Thurstone et Chave sur les attitudes des américains à l'égard de l'église (1929).

Certes, Thurstone n'est pas le premier à mener des études sur le comportement social des individus, sur leur personnalité, sur leurs attitudes, sur leurs opinions : plusieurs recherches ont, depuis la fin de la Première Guerre mondiale, devancé les siennes. Mais Thurstone est certainement le premier à proposer une méthode explicite et objectivée dans un cadre théorique et pratique cohérent. Ce n'est qu'à partir de 1931 et les travaux de Rensis Likert, Louis Guttman, Clyde Coombs et Paul Lazarsfeld que d'autres méthodes seront explicitées⁵⁶. La « loi-méthode » de Thurstone est longtemps restée un des outils traditionnels d'échelonnement pour les psychologues : au moins jusqu'aux années 1960, les principaux ouvrages méthodologiques destinés aux psychologues présentaient la méthode de Thurstone comme une des trois ou quatre méthodes d'échelonnement disponibles. En 1962, Maurice Reuchlin écrivait d'ailleurs que la loi de Thurstone avait suscité des publications « ininterrompues jusqu'[alors]⁵⁷ ».

3. — L'IDÉE DE MESURE DE BINET À THURSTONE

Au terme de ce parcours dans l'histoire de la mesure en psychologie, divers enseignements méritent d'être tirés : les enseignements relatifs à l'histoire de la psychologie et de ses concepts ; ceux relatifs à l'histoire des probabilités et des statistiques dans le cadre de la psychologie ; enfin, et surtout, quelques éléments de réponse à la question des rapports entre outils statistiques et grandeurs psychologiques.

56. L'ouvrage *Experimental Social Psychology* publié en 1931 et réédité en 1937 confirme la primauté de la méthode de Thurstone. Cet ouvrage, central pour les *social scientists* américains du milieu du siècle et donc pour les historiens, ne fait état dans son édition de 1937 que de deux méthodes d'échelonnement explicites : Thurstone et Likert, voir MURPHY, NEWCOMB, 1937.

57. REUHLIN, 1962, p. 118.

3.1. Parcours des concepts de mesure et d'intelligence dans l'histoire de la psychologie

Le processus historique retracé dans cet article marque, à de nombreux égards, le passage de la psychologie du XIX^e à celle du XX^e siècle. En une trentaine d'années et en passant de l'univers médical français à l'univers psychologique américain, les outils de mesure et la conception de l'intelligence se métamorphosent. D'une part, nous voyons apparaître des outils statistiques et probabilistes permettant aux psychologues de disposer de méthodes explicites de mesure et de quantification des grandeurs étudiées : quasiment absents chez Binet et toujours marginaux par rapport à son propos principal, les instruments formels de mesure se développent aux États-Unis afin de pouvoir faire face à la demande grandissante de méthodes systématiques. Thurstone les a rendus très explicites et en a fait une fin en soi dans ses recherches. D'autre part, conjointement à l'automatisation et à la mathématisation progressive des outils, les grandeurs étudiées ont connu une simplification : l'intelligence est par exemple passée du statut de grandeur plurielle et multiforme au statut de grandeur mesurable unilinéairement dont le réalisme ne pose pas question.

Un des autres enseignements que cette reconstruction de l'histoire de la mesure en psychologie dans les trente premières années de ce siècle nous fait voir est qu'une unification théorique de deux champs disciplinaires ne signifie pas nécessairement l'unification institutionnelle de ces deux champs : l'élargissement de la portée de la loi du jugement comparatif à toute la psychologie n'a pas entraîné, dans sa suite, une unification de toutes les psychologies concernées par cette loi (psychophysique, socio-psychologie et psychologie des tests mentaux).

3.2. Diffusion et statuts des outils statistiques en psychologie

Une des grandes leçons de cette reconstruction historique est la confirmation de l'importance grandissante des rôles des outils statistiques et probabilistes dans certains espaces de la recherche et de la pratique psychologique durant les premières décennies de ce siècle. Mais au-delà de cette confirmation, l'histoire que nous venons de parcourir nous fait voir les changements de statuts et de rôles que les outils statistiques de la psychologie connaissent.

Le constat qui s'impose le plus naturellement est relatif à l'évolution des cadres d'utilisation des méthodes et de la loi développées par Thurstone : parties du seul cadre des tests mentaux en tant que méthode d'échelonnement, les « constructions formelles » de Thurstone deviennent, grâce à l'éloignement progressif des raisons qui leur ont donné naissance, une

méthode et une loi dont le champ d'application couvre aussi bien l'étude des opinions, des attitudes et des valeurs que celle de l'intelligence et des aptitudes spécifiques. Cet éloignement et cette autonomisation constituent certainement le processus historique dont l'importance pour l'histoire des outils statistiques en psychologie est la plus grande : elle est une des conditions de la circulation et de l'adaptation des outils statistiques. Ce processus ne fait pas figure d'exception : comme cela avait été le cas au tournant des deux siècles avec les outils de la corrélation, « une rhétorique s'efface [...] mais une grammaire formelle se détache et trouve d'autres usages⁵⁸ ».

Le constat suivant est relatif au rôle joué par la loi normale dans les outils et théories psychologiques : initialement outil empirique, la loi devient, au fil des années et des travaux de Thurstone, un concept théorique lié à la réalité supposée (c'est-à-dire modélisée) des grandeurs psychologiques. Dans la méthode d'échelonnement de Thurstone (et donc celle de Thorndike), la loi normale n'a pas de statut théorique précis : la distribution des résultats est souvent normale par construction mais il n'y a aucune nécessité à cela. Dans la loi du jugement comparatif, la courbe normale est une *propriété théorique* que les stimuli sont supposés posséder. Du coup, la loi normale acquiert un statut ontologique, lié à la réalité intrinsèque supposée des grandeurs étudiées. Ce passage du statut d'outil ou de constatation empirique au statut de propriété ontologique n'est d'ailleurs pas nouveau : Alain Desrosières a montré que Galton a, lui aussi, transformé « la distribution normale (ou « d'allure normale ») observée en une distribution normale *supposée* d'une grandeur hypothétique⁵⁹ ». Une transformation identique, bien que plus profonde, aura lieu après la Seconde Guerre mondiale lorsque les outils d'inférence statistique vont devenir les métaphores des processus cognitifs humains. Gerd Gigerenzer a en effet montré, à travers une analyse de l'histoire du rôle des méthodes d'inférence statistique dans les années 1940⁶⁰, qu'après avoir été utilisées comme simples outils scientifiques, ces méthodes sont devenues des métaphores des théories du fonctionnement cognitif : l'homme est pensé comme un individu raisonnant selon les schémas de l'inférence statistique. En passant du statut de simple outil au statut de théorie ou de concept, les outils statistiques et probabilistes élargissent leur rôle, augmentent leur portée et font désormais partie des accessoires des psychologues, au point de constituer parfois le cœur de leur appareillage théorique et conceptuel.

58. DESROSIÈRES, 1993, p. 179, évoque ici la « rhétorique de l'eugénisme et de la psychométrie » et la « grammaire » des outils statistiques des corrélations.

59. *Ibid.*, p. 143, c'est l'auteur qui souligne.

60. GIGERENZER, 1991.

3.3. *Statistiques, probabilités et psychologie : tensions et libertés*

Toutes ces réflexions nous amènent à revoir notre conception des rapports entre, d'une part, les outils statistiques et probabilistes formels et, d'autre part, les concepts et grandeurs utilisés en psychologie. Entre un réalisme naïf des outils utilisés pour la mesure et un relativisme fort, il existe une autre voie que nous avons cherché à explorer. Quelques remarques préalables sont nécessaires.

La première remarque conduit à constater que des outils formels identiques peuvent donner lieu à des interprétations très différentes. Par exemple, dans les travaux de Fechner, la loi normale est censée représenter la distribution des erreurs d'un sujet lors de l'évaluation d'une grandeur alors que dans la loi de Thurstone cette même loi normale représente la dispersion des processus discriminants d'un *stimulus* et est une caractéristique de ce *stimulus*. Le second exemple nous est donné par le passage de la méthode à la loi : les mêmes outils formels, agencés de façon différente, peuvent donner lieu à deux constructions dont le sens empirique est radicalement différent. Le sens donné aux objets statistiques ou probabilistes formels n'est donc pas une propriété intrinsèque de ces objets : il leur est largement extérieur. À cet égard, les contributions de Thurstone apparaissent davantage comme des réinterprétations que comme de véritables innovations techniques. Notons, d'ailleurs, que l'autonomisation des outils statistiques et probabilistes ne peut se faire que sous la condition de séparation effective du sens et de la forme.

La remarque suivante concerne le rapport que Thurstone entretient avec le « réalisme » des grandeurs mesurées. À aucun moment, malgré toutes ses critiques, il ne semble s'interroger sur la réalité des grandeurs mesurées : il critique les aspects mathématiques des outils, les principes logiques qui les fondent, mais il ne remet jamais en cause l'hypothèse selon laquelle les grandeurs psychologiques sont susceptibles d'être mesurées et les grandeurs mesurées sont bien celles souhaitées. Il modifie, dans une large mesure, les outils sans pour autant toucher à leur « contenu métaphysique ». Cette remarque, établie pour Thurstone, vaut *a fortiori* pour les autres psychologues.

La troisième et dernière remarque naît du constat qu'au fil des années, malgré les évolutions des outils de mesure et des concepts psychologiques, ces outils formels ont toujours été adaptés à la mesure de ces concepts. Les outils ont évolué, les conceptions des grandeurs ont changé, mais dans l'esprit des utilisateurs les outils remplissent bien leur rôle. Ce ne sont pas les quelques critiques de Thurstone, exclusivement de nature logique et mathématique, relatives aux seuls aspects formels des outils, qui changent cet état de fait : malgré les évolutions, outils formels et grandeurs psychologiques coïncident et se correspondent parfaitement.

Ces trois remarques conduisent au tableau « paradoxal » suivant : bien qu'en apparence les significations accordées aux outils et à leur mise en œuvre ne soient pas des propriétés intrinsèques de ces outils, et bien qu'à aucun moment les psychologues ne s'interrogent sur le réalisme de leurs pratiques et n'élaborent une théorie mettant en correspondance psychologie et mathématiques, il apparaît que la signification accordée par les utilisateurs à leurs outils coïncide exactement avec leurs besoins et les grandeurs psychologiques étudiées. Quatre éléments composent donc cette coïncidence « paradoxale » : les psychologues, les instruments formels (outillage statistique et probabiliste), les concepts psychologiques et, enfin, l'épreuve de la « réalité ». Le renouvellement de notre conception du rapport entre outils statistiques et grandeurs psychologiques viendra de notre capacité à résoudre ce « paradoxe » et donc à rendre intelligible le rôle de chacun des acteurs et leur articulation. Comment expliquer à la fois la stricte séparation des formes et des sens qui leur sont attribués, et les liens entre les évolutions que connaissent ces formes et ces sens ?

Refusant de voir, dans le seul cynisme politique des psychologues, les causes de cette situation, c'est dans les « dispositifs mentaux » (catégories de pensée ou d'entendement) servant à saisir les expériences des sens, que se situe certainement la clef de la situation. Les grandeurs psychologiques étudiées, dont la conception est le produit de ces « dispositifs mentaux », fournissent leur sens aux outils formels utilisés. Inversement, lorsque les outils formels évoluent selon une logique qui leur appartient (logique formelle) ou lorsque de nouveaux outils sont introduits (loi normale, dispersion), les conceptions des objets mesurés sont refaçonnées et redéfinies en fonction de ces outils et de leur logique propre (interprétation de la dispersion, modélisation par la loi normale). Entre les instruments formels et les concepts psychologiques, la médiation passe par les dispositifs mentaux (des psychologues). Leur adéquation est la simple conséquence de cette médiation qui apparaît dès lors comme le lien de l'épreuve de la « réalité ».

Olivier MARTIN
(septembre 1994).

BIBLIOGRAPHIE

- BENZÉCRI (Jean-Paul), 1982, *Histoire et préhistoire de l'analyse des données*, Paris, Dunod.
- BINET (Alfred), HENRI (Victor), 1896, « La psychologie individuelle. Description d'un objet », *L'Année psychologique*, vol. 2, p. 411-465.
- BINET (A.), 1898, « La mesure en psychologie individuelle », *Revue philosophique de la France et de l'Étranger*, t. 46, 2, p. 113-123.
- BINET (A.), 1905, « À propos de la mesure de l'intelligence », *L'Année psychologique*, vol. 11, p. 69-82.
- BINET (Alfred), SIMON (Théodore), 1905a, « Sur la nécessité d'établir un diagnostic scientifique des états inférieurs de l'intelligence », *L'Année psychologique*, vol. 11, p. 163-190.
- BINET (A.), SIMON (T.), 1905b, « Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux », *L'Année psychologique*, vol. 11, p. 191-244.
- BINET (A.), SIMON (T.), 1908, « Le développement de l'intelligence chez les enfants », *L'Année psychologique*, vol. 14, p. 1-94.
- BINET (A.), 1911, « Nouvelles recherches sur la mesure du niveau intellectuel chez les enfants d'école », *L'Année psychologique*, vol. 17, p. 145-201.
- BORING (Edwin G.), 1950, *A history of experimental psychology*, 2^e éd. revue, New York, Apleton-Century-Croft.
- BULMER (Martin), 1981, « Quantification and Chicago social science in the 1920's. A neglected tradition », *Journal of the history of the behavioral sciences*, vol. 17, p. 312-331.
- BUSS (Allan R.), 1976, « Galton and the birth of differential psychology and eugenics. Social, political and economic forces », *Journal of the history of the behavioral sciences*, vol. 12, p. 47-58.
- CATTELL (James McKeen), 1890, « Mental test and measurements », *Mind*, 1^{re} sér., vol. 15, p. 373-381. (Article accompagné d'un commentaire de GALTON.)
- DANZIGER (Kurt), 1990, *Constructing the subject. Historical origins of psychological research*, Cambridge, Cambridge University Press.
- DESROSIÈRES (Alain), 1993, *La Politique des grands nombres. Histoire de la raison statistique*, Paris, La Découverte.
- FRAISSE (Paul), SEGUI (Juan), dir., 1994, *Les Origines de la psychologie scientifique. Centième anniversaire de L'Année psychologique (1894-1994)*, Paris, Presses universitaires de France.
- GELB (Steven A.), 1986, « Henry H. Goddard and the immigrants, 1910-1917. The studies and their social context », *Journal of the history of the behavioral sciences*, vol. 22, p. 324-332.
- GIGERENZER (Gerd), 1991, « From tools to theories. A heuristic of discovery in cognitive psychology », *Psychological Review*, vol. 98, 2, p. 254-267.
- GOULD (Stephen Jay), 1983, 1^{re} éd. 1981, *La Mal-mesure de l'homme. L'intelligence sous la toise des savants*, Paris, Ramsay.

- HORNSTEIN (Gail A.), 1988, « Quantifying psychological phenomena. Debates, dilemmas and implications », in *The Rise of experimentation in American psychology*, dir. Jill G. MORAWSKI, Yale, Yale University Press, p. 1-34.
- LEMAINE (Gérard), MATALON (Benjamin), 1985, *Hommes supérieurs, hommes inférieurs ? La controverse sur l'hérédité de l'intelligence*, Paris, Armand Colin.
- MURPHY (Gardner), MURPHY (Lois B.), NEWCOMB (Theodore M.), 1937, 1^{re} éd. 1931, *Experimental Social Psychology. An interpretation of research upon the socialization of the individual*, New York, Harper & Brothers Publ.
- PAICHELER (Geneviève), 1992, *L'Invention de la psychologie moderne*, Paris, L'Harmattan.
- PINELL (Patrice), ZAFIROPOULOS (Markos), 1983, *Un siècle d'échecs scolaires (1882-1982)*, Paris, Les Éditions ouvrières.
- POLLAK (Michael), 1976, « La planification des sciences sociales », *Actes de la recherche en sciences sociales*, vol. 2, 2-3, p. 105-121.
- REUCHLIN (Maurice), 1962, *Les Méthodes quantitatives en psychologie*, Paris, Presses universitaires de France.
- SCHNEIDER (William H.), 1992, « After Binet. French intelligence testing, 1900-1950 », *Journal of the history of the behavioral sciences*, vol. 28, p. 111-132.
- SIMON (Théodore), 1912, « Alfred Binet », *L'Année psychologique*, vol. 18, p. 1-14.
- STIGLER (Stephen M.), 1986, *The History of statistics. The measurement of uncertainty before 1900*, Cambridge, MA, The Belknap Press of Harvard University Press.
- SYMPOSIUM, 1921, « Intelligence and its measurement. A symposium », *The Journal of educational psychology*, vol. 12, 3 et 4, p. 123-147 et 195-216.
- TERMAN (Lewis M.), 1924, « The mental test as a psychological method », *Psychological Review*, vol. 31, 2, p. 93-117.
- THORNDIKE (Edward Lee), 1913, 1^{re} éd. 1904, *An introduction to the theory of mental and social measurements*, New York, Teachers College Columbia University.
- THURSTONE (Louis Léon), 1921, « Intelligence and its measurement », *The Journal of educational psychology*, vol. 12, 4, p. 201-207.
- THURSTONE (L. L.), 1925, « A method of scaling psychological and educational tests », *The Journal of educational psychology*, vol. 16, 7, p. 433-451.
- THURSTONE (L. L.), 1926a, « The mental age concept », *Psychological Review*, vol. 33, 4, p. 268-278.
- THURSTONE (L. L.), 1926b, « The scoring of individual performance », *The Journal of educational psychology*, vol. 17, 7, p. 446-457.
- THURSTONE (L. L.), 1927a, « Psychophysical analysis », *American Journal of psychology*, vol. 38, p. 368-389.
- THURSTONE (L. L.) 1927b, « A law of comparative judgment », *Psychological Review*, vol. 34, 4, p. 273-286.
- THURSTONE (L. L.), 1927c, « The unit of measurement in educational scales », *The Journal of educational psychology*, vol. 18, 8, p. 505-524.
- THURSTONE (L. L.), 1927d, « A mental unit of measurement », *Psychological Review*, vol. 34, 6, p. 415-423.
- THURSTONE (L. L.), 1927e, « The method of paired comparisons for social values », *The Journal of abnormal and social psychology*, vol. 21, p. 384-400.
- THURSTONE (L. L.), 1927f, « Three psychophysical laws », *Psychological Review*, vol. 34, 5, p. 424-432.
- THURSTONE (L. L.), 1928a, « The absolute zero in intelligence measurement », *Psychological Review*, vol. 35, 3, p. 175-197.
- THURSTONE (L. L.), 1928b, « An experimental study of nationality preferences », *Journal of general psychology*, vol. 1, p. 405-425.

- THURSTONE (L. L.), CHAVE (E. J.), 1929, *The Measurement of attitude*, Chicago, University of Chicago Press.
- THURSTONE (L. L.), 1952, « Leon Louis Thurstone », in *A history of psychology in autobiography*, éd. Edwin G. BORING, Herbert Sidney LANGFELD, Heinz WERNER, Robert M. YERKES, Worcester, MA, Clark University Press, p. 295-321.
- THURSTONE (L. L.), 1959, *The Measurement of values*, Chicago, Chicago University Press.
- WALKER (Helen M.), 1929, *Studies in the history of statistical method*, Baltimore, The Williams & Wilkins Company.
- WOLF (Theta H.), 1969, « The emergence of Binet's conception and measurement of intelligence. A case history of the creative process », *Journal of the history of the behavioral sciences*, vol. 5, p. 113-134, 207-236.
- WOLF (T. H.), 1973, *Alfred Binet*, Chicago, University of Chicago Press.
- YERKES (Robert M.), 1919, « Report of the Psychology Committee of the National Research Council », *Psychological Review*, vol. 26, 2, p. 83-149.
- YERKES (R. M.), dir., 1921, « Psychological examining in the army in the United States army », *Memoirs of the National Academy of Sciences*, vol. 15.