



Université Abderrahmane Mira de Bejaia
Faculté des Sciences Humaines et Sociales
Département des Sciences Sociales

Polycopié de cours en psychométrie

Destiné aux étudiants de deuxième année Psychologie

Préparé par

Dr : KHELOUFI Sihem épouse DRIS

Maitre de Conférences (B)

2020/2021

Sommaire

| | |
|---|----|
| Les objectifs de module : | 1 |
| Introduction | 2 |
| 2. Définition de la psychométrie: | 3 |
| 3. Les objectifs de la psychométrie: | 4 |
| 4. La différence entre la mesure physique et la psychométrie: | 5 |
| 5. Les Origines de la psychométrie: | 5 |
| 6. L'école française: | 6 |
| 6.1. Les fondateurs de la psychométrie en France: | 6 |
| 7. L'école allemande: | 9 |
| 7.1. Les fondateurs de l'école allemande: | 9 |
| 8. L'école anglaise: | 13 |
| 9. L'école américaine : | 18 |
| 9.1. Les fondateurs de l'école américaine: | 18 |
| 10. Les caractéristiques de la psychométrie : | 20 |
| 11. Interprétation psychologique des mesures psychométriques: | 21 |
| 12. Les erreurs de la psychométrie: | 22 |
| 13. Les niveaux de la mesure en psychométrie: | 24 |
| 13.1. Types d'échelles : | 25 |
| 13. Les outils de la psychométrie (les tests psychologiques) | 29 |
| 14.1. La définition d'un test : | 29 |
| 14.2. Les caractéristiques principales d'un test: | 29 |
| 14.3. Catégorisation des tests: | 30 |
| 14.4. Les tests « périphériques »: | 32 |
| 14.5. Principaux utilisateurs des tests : | 33 |
| 15. Les conditions d'un bon test : | 33 |
| 16. La standardisation : | 47 |
| 16.1. Qu'est-ce qu'une évaluation standardisée ? | 47 |
| 16.2. Les conditions de la bonne standardisation : | 47 |
| 17. Etalonnage d'un test : | 49 |
| 17.2. Phase d'échantillonnage : | 49 |
| 17.2. Construction d'un étalonnage : | 50 |
| Conclusion | 53 |
| Les références bibliographiques | 54 |

Les objectifs de module :

Le module de La psychométrie n'est pas moins important que d'autres modules dans l'unité méthodologique car il est nécessaire pour les chercheurs comme les praticiens de la psychologie d'utiliser des différents instruments de mesures (tests) pour évaluer et diagnostiquer les troubles psychologiques.

Deux grands objectifs que vise l'enseignement de la psychométrie à atteindre :

Le savoir : comprendre et connaître l'histoire de la psychométrie et comment que la psychométrie à donner un aspect scientifique pour la psychologie grâce aux différents instruments de mesure et connaître plusieurs instruments et leurs utilité dans des différents domaine de la psychologie ainsi reconnaître la pertinence de l'information qui va nous permettre de juger la qualité psychométrique d'un instrument de mesure.et définir correctement la notion de tests en psychométrie et énumérer les principales caractéristiques d'un tests en psychométrie.

Le savoir-faire : connaître les principes généraux guidant à l'élaboration d'un test et effectuer les étapes de la construction d'un test (instrument de mesure), ainsi émettre une affirmation (jugement) claire quant à la sensibilité, fidélité et validité et l'utilité d'un instrument de mesure (test).

Introduction

Le domaine de la mesure en psychologie et en sciences de l'éducation est un domaine extrêmement vaste et dont les techniques sont particulièrement variées. La psychométrie concerne l'ensemble des principes à la base de la mesure en psychologie. Dans l'enseignement universitaire en langue française (niveau licence), ce terme est associé plus particulièrement à la pratique et la construction des tests.

Comme le signale Larencelle (1998, p. 14), la distinction entre évaluation et mesure prend généralement peu d'importance en sciences et dans le monde des laboratoires scientifiques. Les mesures y sont pour la plupart physiques, purement matérielles, et ne donnent en elles-mêmes que peu de champ à l'interprétation subjective. Au contraire en psychologie, en éducation et dans les autres sciences humaines, l'évaluation peut et doit compléter la mesure, puisque souvent la valeur potentielle de celle-ci dépend entièrement du contexte dans lequel on la fait intervenir. Il conviendra donc bien de distinguer le travail d'instrumentation et les méthodes de mesure de l'interprétation qui pourra être faite du mesurage, même si la construction de l'instrument est déjà un parti pris. **(Demeuse, 2004, p3)**

Ce cours de Licence respecte cette tradition et restreint l'enseignement de la psychométrie à la pratique et la construction des tests. Il demande à l'étudiant d'apprendre (et comprendre) de très nombreux termes nouveaux, d'acquérir des connaissances connexes comme la notion de la mesure, ou de maîtriser des notions de statistiques descriptives ou différentielles (pré requis). Il est aussi fait référence aux règles éthiques qui accompagnent la pratique des tests.

Le cours de psychométrie est donc "rugueux" pour l'étudiant mais indispensable car s'il permet de mieux comprendre ce qu'est la mesure en psychologie, il est essentiel pour apprécier l'intérêt des tests mais aussi leurs limites. Tous les professionnels qui utilisent ou font référence à des tests devraient avoir une formation minimum en psychométrie.

Ce polycopié est un support pédagogique contient des cours de psychométrie destiné aux étudiants de L2 psychologie a pour but de connaitre des notions de base de la psychométrie et son développement comme discipline a part en tiers et sa collaboration dans la construction des tests psychologique qui sert à diagnostiquer les différents troubles dans différentes composantes de la psychologie.

Dans ce support pédagogique en va traiter les cours suivants : la définition de la mesure et la psychométrie, les écoles de la psychométrie, les erreurs et les niveaux de la psychométrie, les outils de la psychométrie ainsi qu'à les conditions d'un bon test.

1. Définition de la mesure:

✓ **Mesurer** : Déterminer, évaluer avec un instrument de mesure, le volume, la superficie, la quantité de quelque chose : Mesurer la longueur d'un mur. Mesurer une table avec un mètre. Mesurer des angles en degrés (Larousse .Fr)

✓ Selon Larousse la mesure est une évaluation d'une grandeur par comparaison avec une autre de la même espèce prise pour unité ; grandeur quantité ainsi déterminée.

(Larousse, 2008, p 266)

✓ **La mesure** est une notion (processus) complexe qui concerne des caractéristiques d'un objet ou d'une personne (la taille, la tension, l'extraversion, l'intelligence, la température corporelle, etc.). De façon générale mesurer c'est **attribuer des nombres aux objets, selon des règles déterminées**. Ces règles vont toujours avoir pour objet d'établir une correspondance entre certaines propriétés des nombres et certaines propriétés des objets.

(Roulin, 2018, p10)

2. Définition de la psychométrie:

Le terme « psychométrie » dérive du grec « psyché » (l'esprit) et « metron » (mesure), signifie littéralement : « la mesure des traits psychologique ».le but d'une évaluation psychométrique est donc de mieux comprendre les aptitudes fondamentales d'une personne comme la personnalité, les motivations, le raisonnement ou bien l'intelligence émotionnelle (ou encore d'autre capacités cognitives et comportementales). (Baromètre Ipsos/Edenred.2012.)

✓ selon **Larousse** la psychométrie est une « branche de la psychologie consacrée aux tests, à leur construction et à leur utilisation. (Larousse.fr)

✓ **la psychométrie** est un domaine d'étude qui concerne la théorie et la méthodologie de construction et d'utilisation des échelles de mesure des caractéristiques mentales. Celles-ci peuvent être des connaissances, des traits de personnalité ou encore des composantes de la cognition comme le langage, l'intelligence ou la mémoire.

(Grégoire.j sans date)

✓**La psychométrie** est la science qui étudie l'ensemble des techniques de mesures pratiquées en psychologie, ainsi que les techniques de validation et d'élaboration de ces

mesures. Ces techniques concernent tous les champs de la psychologie, ainsi que d'autres domaines de sciences connexes (comme la recherche en comportement du consommateur par exemple). À l'origine, elles ont surtout émergé par l'intérêt croissant pour la mesure des performances intellectuelles ou bien pour l'analyse des composantes de la personnalité (affectivité, émotions, relations avec les autres, etc.). La psychométrie est la science de la mesure de l'esprit et s'établit en tant que telle par l'usage de procédures rigoureuses s'appuyant notamment sur l'usage de techniques statistiques variées. **(wikipédia/psychométrie)**

- **la psychométrie** : La psychométrie concerne l'ensemble des principes à la base de la mesure en psychologie, ce terme est associé plus particulièrement à la pratique et la construction des tests. **(Roulin, 2018, p8)**
- **la psychométrie** est une branche de la psychologie qui élabore et étudié des théories, méthodes et techniques pour évaluer et mesurer les capacités psychiques d'un individu a partir de tests, inventaires et autres techniques valides statistiquement. **(Relevez vos talent. 2013)**
- **Définition opérationnelle:** la psychométrie concerne toutes les méthodes quantitatives utilisées en psychologie.

3. Les objectifs de la psychométrie:

- ✓ Mesurer les caractéristiques mentales : en élaborer les techniques de mesure. Et concevoir les lois mathématiques reliant ces caractéristiques au comportement. Aujourd'hui la psychométrie désigne principalement les méthodes employées dans la conception et construction des tests mentaux.
- ✓ Réaliser une évaluation des caractéristiques individuelles sur le plan intellectuel et de la personnalité. Le test permet d'évaluer la position d'un individu en comparaison d'un groupe de référence .on va donc situer la performance du sujet par rapport aux performances obtenues par la population d'étalonnage.
- ✓ Concevoir les lois mathématiques reliant ces caractéristiques au comportement. Aujourd'hui la psychométrie désigne principalement les méthodes employées dans la conception et construction des tests mentaux.

4. La différence entre la mesure physique et la psychométrie:

En physique, mesurer est le fait de déterminer la grandeur particulière d'un objet par rapport à un étalon ou une unité. la métrologie est l'étude de la mesure physique. En général une métrique est une échelle de mesure définie en termes d'étalon (ou standard), c'est -à-dire en termes d'unité clairement définie. la quantification d'un phénomène par le processus de mesure est basée sur l'existence implicite ou explicite d'une métrique, qui est le standard de référence des mesures.

(Vial. M.2012)

5. Les Origines de la psychométrie:

Il est toujours délicat de définir précisément les origines d'une méthode, d'une discipline, la psychométrie n'échappe pas à la règle pourtant le sujet n'est pas neutre : en introduisant la notion de mesure dans l'étude de l'esprit humain, la psychométrie permet à la psychologie de se différencier de la philosophie dont elle est issue, la question des origines permet donc de dater la bifurcation des deux disciplines.

Juan huart de San juan ,médecin et philosophe espagnol (1530-1588), marque un vrai départ il publie en 1573 son examen des esprits pour les sciences qui de diffuse progressivement à travers toute l'aurope, il s'agit du premier traité établissant une relation entre psychologie et physiologie ,mais c'est le philosophe et mathématicien allemand Christian Von Wolff qui est reconnu comme le premier à utiliser le terme (psychométrie).il écrit en 1732 dans son ouvrage « psychologie empirica » le mot psychométria.

En Angleterre, Francis Galton (1822-1911) cousin de Charles Darwin, s'intéresse à la transmission héréditaire des capacités intellectuelles. En 1869, dans (Hereditary Genius)(le génie héréditaire) il propose de mesurer le génie d'un individu par la fréquence des sujet qui dans un groupe d'individus donnés ,parvient à le dépasser, il conceptualise plus tard le calcul du coefficient de corrélation qui sera ensuite perfectionné par l'un de ses disciples Karl Pearson (1857-1936),et conduira à la formalisation par Charles Spearman (1863-1945) de la notion d'analyse factorielle.

Mais en Allemagne constitue la croissance de la psychométrie comme une discipline à part entière .à la suite de Gustav Théodore Fechner(1801-1887)qui dessine la route permettant à la psychologie de s'affirmer comme une véritable science.son compatriote Wilhelm Wundt(1832-1920)peut être reconnu comme le premier psychologue moderne et le père de la

psychologie expérimentale ,il crée formellement en 1879 le premier laboratoire de la psychologie a l'université de Leipzig ,en 1886 la première thèse dans le domaine de la psychologie est soutenue en Allemagne sous la supervision de Wundt ,par un Américain nommé James Mckeen Cattell(1860-1944),1890 de retour aux Etats-Unis James Mckeen Cattell introduit le concept de **test mental** dans un article publié dans la revue Mind et Institute « Mental tests and Measurement ».il fut l'un des psychologue américain les plus influents de son siècle à partir de la , la psychométrie connaîtra un essor croissant.

(Relevez vos talent.2013)

La psychométrie se développée grâce aux travaux de plusieurs personnes chaque une dans une école:

6. L'école française:

6.1. Les fondateurs de la psychométrie en France:

La biographie d'Alfred binet:

Né à Nice, fils d'un médecin et d'une artiste peintre .il commence d'abord des études de droit et exerce dans ce domaine jusqu'à ses premiers travaux avec J.M.Charcot en 1880.

En 1884 il épouse la fille de l'embryologiste Edouard-Gérard Balbiani, son nouveau beau-père l'encourage à commencer des études de sciences naturelles à la Sorbonne.

En 1891, il rejoint le laboratoire de la psychologie physiologique de la Sorbonne créé 2ans auparavant par Henri Beaunis dont il s'est toujours trouvé proche. En 1894 il en devient le directeur.

Au début des années 1900,binet se désintéresse peu à peu du laboratoire et se rend sur le terrain à l'école primaire de la rue de la grange-aux belles plus précisément ,ou il y observera des écoliers .reconnu pour ses qualités de pédagogues ,c'est à partir de ses observations des élèves de primaire que binet recueillera les données dont il se servira par la suite pour son test d'intelligence, mais aussi pour créer un laboratoire de pédagogie expérimentale en 1906, au sein même de cette école primaire.

En 1904, Binet est chargé par le ministre de l'instruction de réaliser un test permettant de diagnostiquer la débilité mentale d'écoliers en vue de leur perfectionnement.

(psyfact.2016)

La biographie de Théodore Simon:

Né le 10 juillet 1883 à Dijon, orphelin de très bonne heure, Théodore est recueilli chez son oncle à Sens, où il fait ses études secondaires. Il s'installe ensuite à Paris où il suit des études de médecine et passe le concours de l'internat des asiles de la Seine 1898, reçu 4ème dans la promotion de Roger Mignot de Gatian de Clérambault et Joseph Cap Gras, il prend un poste à l'asile de Vaucluse, dans le service du docteur Emery Blin, en charge de la colonie.

Il écrit donc à Alfred Binet, dont il connaissait les premiers travaux, pour lui proposer les observations qu'il avait recueillies sur les arriérés hospitalisés dans ce service. C'est le début d'une collaboration qui ne prendra fin qu'avec la mort de Binet en 1911.

- ✓ En 1902, Théodore Simon prend un poste de médecin adjoint à l'asile de Dury.
- ✓ En juillet 1912, il est nommé médecin du cadre spécial des asiles d'aliéné de la Seine.
- ✓ En 1946 Théodore Simon crée la première école d'infirmière spécialisée, dont il sera le directeur technique de 1946 à 1957. L'école d'infirmière de Maison-Blanche (Neuilly-sur-Marne) initialement appelée « l'école des blues » (de la couleur de l'uniforme porté pour les élèves) ensuite institut de formation inter-hospitalier, porte aujourd'hui le nom de son fondateur.

(Caire. M. 2012)

Le point de départ de test Binet-Simon:

Ce test, créé en 1905 par Alfred Binet et Théodore Simon, est la première échelle de mesure de l'intelligence, composé d'une série d'épreuve de difficulté croissante faisant appel au jugement, à la compréhension et au raisonnement, il rend compte de l'âge mental des enfants testés.

« **L'échelle métrique de l'intelligence** » est née à la suite d'une commande du ministère de l'instruction publique qui souhaitait la mise en œuvre de méthode objective de dépistage de l'insuffisance intellectuelle et de la débilité mentale chez les enfants d'âge scolaire. En effet, l'école publique obligatoire de Jules Ferry avait montré que de nombreux enfants étaient

incapables de suivre cet enseignement ce qui rendait nécessaire la création de classes spécialisées. Le problème de l'orientation des enfants se posait.

Alfred Binet fut choisi pour créer l'instrument de cette orientation. élève de Wundt et directeur du laboratoire de psychologie expérimentale de la Sorbonne, il travaillait déjà sur l'évaluation de l'intelligence (étude expérimentale de l'intelligence 1903). la première mouture du test (1905) comporte des questions de difficulté progressive. sur trente questions, l'« idiot » ne réussit que les six premières, l'« imbécile » va jusqu'à la douzième et ainsi de suite.

En 1908, une nouvelle version de l'échelle métrique constitue un progrès décisif : en regroupant les questions par niveau d'âge on teste par rapport au résultat optimal d'un âge donné. C'est la notion d'âge mental. Ainsi un enfant de douze ans d'âge réel qui réussit à 100% les questions correspondant à la classe d'âge de huit ans et ne va pas au-delà aura huit ans d'âge mental.

Depuis sa création, l'échelle de Binet-Simon a été largement utilisée et surtout beaucoup remaniée et perfectionnée. la notion de QI apportée par Stern dès 1912 en a beaucoup facilité l'utilisation. aux USA, les psychologues Terman et Wechsler l'ont adaptée. En France, René Zazzo en a proposé une nouvelle révision en 1966.

(Larousse en ligne)

L'élaboration de l'échelle métrique:

Il y a maintenant un siècle que le premier test d'intelligence a été élaboré par Alfred Binet (1857-1911), le psychologue français le plus connu dans le monde. il fallait marquer la date du centenaire de cette création par la réédition des écrits fondateurs de Binet, encore trop peu connus, sur l'intelligence. L'élaboration de la première version de l'échelle métrique de l'intelligence sera réalisée au cours de l'année 1904-1905 dans le contexte de l'évolution des pratiques pédagogiques de l'époque. En octobre 1904, une commission ministérielle sera créée autour de la question de la scolarisation des enfants arriérés. Binet qui fait partie de cette commission, imposera l'idée d'un examen psychologique afin de déterminer l'admission des enfants anormaux dans une école spéciale. c'est dans ce contexte qu'il élabore avec l'aide de son collaborateur Théodore Simon (1873-1961) un test destiné à établir un diagnostic scientifique des états inférieurs de l'intelligence. c'est par le détour du diagnostic de la débilité mentale que la question de la mesure de l'intelligence va ainsi être formulée et résolue. C'est

pour dépister les enfants arriérés, pour distinguer avec le minimum d'erreurs entre les retards dus à un déficit d'intelligence et les retards dus aux conditions défavorables du milieu et de la scolarité que binet a construit son test dont la première version fut publiée en 1905 dans l'année psychologique. Le livre présente et analyse dans une perspective historique les écrits constitutifs de la première version du test d'intelligence. La suite de l'ouvrage est consacrée à la réédition des premiers articles de binet et Simon sur l'intelligence (1904-1905).ce livre s'adresse aux psychologues, aux pédagogues et aux historiens spécialistes ou simplement curieux de l'approche scientifique de l'intelligence et plus particulièrement des idées de binet sur la question. **(Serge. 2004. p 336)**

7. L'école allemande:

Au cours de la seconde moitié du 18 siècle, nombre de philosophes et de scientifiques réfléchissent aux moyens de rendre plus scientifique l'étude des processus psychologiques. Pour leurs travaux ils inventent et testent des méthodes expérimentales qui constitueront les bases de la recherche en psychologie. **(Marchand.G.2005)**

7.1. Les fondateurs de l'école allemande:

La biographie de Gustav Theodore Fechner:

Est un philosophe et psychologue allemand, né à Gross Sarchen en 1801 et mort à Leipzig en 1887.son père est pasteur mais ce dernier meurt quand le petit Gustav n'a que 5 ans.il est alors élevé par son oncle pasteur.

Néanmoins, il sera peu influencé par la religion protestante.il rentre à la faculté de médecine à l'âge de 16 ans, il délaisse ses études de médecine pour assister aux cours d'Ernst Heinrich weber qui enseigne la physiologie et les mathématiques à Leipzig.

Il est l'instigateur de la psychophysique, une science visant à mesurer des phénomènes d'ordre psychologique .il formula la loi de Weber –Fechner selon laquelle la sensation perçue varie proportionnellement au logarithme de l'intensité d'excitation.

L'unité métrique dans laquelle se mesurerait la sensation est pour lui le seuil différentiel relatif (rapport de Bouguer-weber), qui est la plus petite différence d'intensité remarquable. en 1839, il s'intéressa au problème de l'âme .il publie ses résultats en 1851, affirmant que la conscience est diffuse partout dans l'univers, l'âme ne meurt pas.

Le 22 octobre 1850 Gustav Fechner a une révélation ; il pense avoir résolu tout à coup une vieille énigme philosophique-celle de la relation entre l'âme et le corps, il pense même avoir trouvé comment la mesurer : grâce à une équation mathématique simple reliant l'intensité de la sensation à celle du stimulus. L'histoire des sciences a retenu ce 22 octobre comme date de fondation de la psychophysique. A partir de la Fechner va jusqu'à la fin de sa vie, poursuivre ses travaux mélangeant psychologie, physique et métaphysique.

(www.babilio.com/gustav-theodor-fechner)

A ses débuts, la psychologie expérimentale est essentiellement synonyme de psychophysique.

✓ **la loi de weber-Fechner:** c'est l'équation permettant de prédire le seuil différentiel.

Le seuil différentiel « est la petite différence entre deux stimuli qu'un individu peut détecter consciemment » supposons, par exemple que vous avez deux verres, A et B, contenant chacun 250ml d'eau et 100 grains de sel ;le seuil différentiel ,ce serait ici le plus petit nombre de grains de sel qu'il faut ajouter au verre B pour que vous détectiez une différence dans le gout salé de l'eau des deux verres.si cette quantité est égale à 40 grains, alors on dit que le seuil différentiel du stimulus 100 grains de sel dans 250 ml est de 40grains de sel.

Ce qui peut paraître un peu fascinant avec cette notion de seuil différentiel, c'est qu'il semble possible de le prédire. Voici l'équation permettant de le prédire:

$$SD=K.S$$

(Ou SD=seuil différentiel ; K=constante ; et S=le stimulus original)

Pour trouvé K, il s'agit de calculer un seuil différentiel et de faire le rapport:

$$K=SD /S$$

Revenons à l'exemple précédent .maintenant que vous avez calculé le seuil différentiel pour un stimulus de 100grains de sel dans 250ml d'eau. Supposons que vous voudriez connaître le seuil différentiel pour un stimulus de 400grains de sel dans 250ml d'eau:

=40 grains de sel / 100 grains de sel = 0,4 remplaçons maintenant K par 0,4 dans l'équation, et donnons à S la valeur 400 grains de sel : $SD = 0,4 \times 400 \text{ grains de sel} = 160 \text{ grains de sel}$.

(berthieume.f.sans date)

La biographie de Wilhelm Wundt:

Né près de Mannheim en 1832 meurt 1920, fils de pasteur, Wundt fait des études de médecine puis apprend la physiologie auprès du célèbre Johannes Muller à Berlin. Il enseigne successivement à Heidelberg à Zurich et à Leipzig, où il obtient en 1875 une chaire de philosophie et où il fonde en 1879 le premier laboratoire de la psychologie expérimentale qui ait une existence officielle.

Wundt a commencé par des travaux sur la perception en raison de sa formation de physiologiste. Ils s'intéressa aussi à la mesure de la complexité des processus psychologiques par la méthode des temps de réaction.

Du fait de l'influence qu'il a exercée, par l'intermédiaire de son laboratoire plus sans doute que par son œuvre, Wundt peut être légitimement considéré comme le véritable fondateur de la psychologie expérimentale.

(Richard. J-F .sans date)

La naissance de la psychologie expérimentale:

L'année de fondation reconnue de cette discipline est 1879, année où Wilhelm Wundt ouvre son laboratoire à l'université de Leipzig, (ville située à environ 135 km au sud de Berlin). La psychologie expérimentale est alors une branche de la physiologie et le lien entre ces deux disciplines est la psychophysique.

A ses débuts la psychologie expérimentale est essentiellement synonyme de psychophysique. La contribution du laboratoire de Wundt au domaine de la psychométrie est double. D'abord comme tout bon laboratoire de sciences, il se concentre sur la standardisation des conditions et sur la précision des mesures. Ensuite il s'intéresse aux processus élémentaires, notamment la sensation, les seuils, la perception et les réactions motrices simples. Le laboratoire de Wundt devient un lieu de formation de choix pour de nombreux pionniers de la psychologie. Ainsi, les intérêts et les méthodes de Wundt exercent-ils une grande influence sur ce domaine naissant.

(Hogan. 2017. p15)

La méthode de la psychologie expérimentale:

La psychologie expérimentale est une méthodologie systématique d'observation et d'acquisition de connaissances sur les attitudes et les comportements, basée sur l'expérimentation.

On peut l'opposer grossièrement à la méthode visant à demander aux personnes d'expliquer les causes, processus et conséquences de leurs comportements en psychologie expérimentale, on sait que les gens sont très peu doués pour l'introspection et qu'en général, ils se trompent pour expliquer leurs comportements.

La psychologie expérimentale se fixe pour objectif de contrôler de façon très pointue les différents paramètres d'une situation pour en évaluer de façon systématique, l'impact sur les attitudes et comportement. Le psychologue parle de variables.

La première étape du psychologue expérimental est de formuler une hypothèse, généralement élaborée sur les recherches des autres psychologues ou sur une idée nouvelle.

Le psychologue expérimental va ensuite mettre en place une situation expérimentale. il va en contrôler toutes les variables possibles ; s'il oublie de le faire, ses collègues ne manqueront pas de lui en faire le reproche en lui signalant les biais et sa recherche ne sera pas publiée.

En psychologie expérimentale, le travail des statistiques est très important. Le psychologue (l'expérimentateur) va donc travailler avec deux ou plusieurs groupes de personnes (20,25 personnes par groupe, ou parfois des centaines). il va mettre en place une manipulation expérimentale à laquelle seront soumis le premier groupe et pas le second. Et il va mesurer l'impact de la manipulation en observant la différence de moyenne entre les deux groupes.

Comme il doit pouvoir mesurer cette différence entre le groupe qui a reçu la manipulation et celle qui ne l'a pas reçue, il travaillera volontiers avec des réponses quantifiables ; les sujets sont ainsi invités à répondre sur des échelles de 0(=pas du tout) à 5(=tout à fait). une analyse statistique sur les moyennes lui permettra ensuite au scientifique de déterminer si son hypothèse de départ se vérifie ou pas. **(Vermeulen. sans date)**

8. L'école anglaise:

Les anglais ont contribué par leurs recherches et leurs études à développer la psychométrie, les fondateurs de cette école et leurs travaux comme suit:

La biographie de Francis Galton :

Sir Francis Galton est un anthropologue, explorateur, géographe, météorologue et statisticien britannique. Il est surtout connu pour ses travaux pionniers sur l'intelligence humaine et pour avoir créé la théorie de l'eugénisme. Il est né le 16 février 1822 à Sparbrook, près de Birmingham, d'une famille de gens aisés (son père est banquier) et instruit (il est cousin de Charles Darwin). Après avoir fréquenté plusieurs écoles de Birmingham, il rentre au King Edwards school, en 1836, mais l'enseignement classique et religieux ne lui convient pas. Ses parents décident alors qu'il doit devenir médecin, il étudie cette discipline un an à Birmingham et un an à Londres. Il profite d'un voyage en 1840 à l'université de Giessen, en Allemagne, dans le but de suivre des cours de chimie, pour découvrir le sud-est de l'Europe. Galton s'est intéressé à de nombreux autres sujets comme la météorologie. Il découvre aussi que les empreintes digitales d'un individu ne varient pas tout au long de sa vie, et c'est à la suite de ces travaux que les empreintes digitales furent utilisées par les polices du monde entier. Après avoir reçu de nombreux honneurs, dont la médaille Copley de la Royal Society, il est anobli en 1909 deux ans avant sa mort. (www.bibmath.net/Galton)

- les travaux en psychométrie qui lui revient:

✓ La transmission héréditaire (l'eugénisme):

Explorateur, géographe de talent, météorologiste, biométricien, sir Francis Galton est aussi le fondateur de l'eugénisme en 1883. L'eugénisme va se développer en se fondant sur une nouvelle théorie de l'hérédité, notamment exposée par Galton, mais aussi sur la théorie de l'évolution de Darwin, appliquée à la société humaine par Spencer, l'eugénisme de Galton est un programme de sélection artificielle pour produire une race humaine supérieure par un contrôle des mariages. Galton est favorable à un eugénisme positif espérant que les gens se fixeraient des objectifs eugénistes en choisissant leur partenaire en vue du mariage. En 1904 à Londres, il expose ses idées devant une foule de médecins et de scientifiques. Son discours largement diffusé, servira de points de départ aux mouvements eugénistes américains et européens qui se développeront dans la première moitié du XXe siècle.

(Aubert-Marsan.2009)

✓ **La définition de l'eugénisme selon Larousse:**

« Théorie cherchant à opérer une sélection sur les collectivités humaines à partir des lois de génétique ».

(larousse.fr)

- **La biographie de Karl Pearson:**

Né le 27 mars 1857 à Londres meurt 27 avril 1936 à Coldharbour Karl Pearson est un mathématicien anglais qui fut l'une des figures de proue de l'émergence de la statistique, au début du XX^e siècle. Issu d'une famille de la classe moyenne supérieure du Yorkshire son père est avocat, il entre à 9 ans à l'université Collège School de Londres, qu'il quitte à 16 ans pour des raisons de santé. Grâce à l'aide d'un tuteur privé, il réussit néanmoins l'examen d'entrée à Cambridge en 1875 et entre alors au King Collège. Il y étudie principalement les mathématiques. Après son succès aux examens terminaux de 1878, il bénéficie d'une bourse qui lui donne la liberté d'étudier et de voyager. Il en profite pour se rendre en Allemagne en 1880 où il étudie la philosophie et la littérature allemande, puis à son retour en Angleterre pour suivre des cours de droit. En 1884 il devient professeur de mathématiques appliquées à l'University College of London.

Karl Pearson a été un partisan convaincu de l'eugénisme : pour lui, il est souhaitable d'améliorer la race humaine en sélectionnant et en favorisant les plus doués de ses représentants, comme le fait la sélection naturelle pour les animaux.

([www.bibmath.net/ Pearson](http://www.bibmath.net/Pearson))

Les travaux de Karl Pearson:

✓ **Le coefficient de corrélation :**

Le coefficient de corrélation de Bravais-Pearson est un indice statistique qui exprime l'intensité et le sens (positif ou négatif) de la relation linéaire entre deux variables quantitatives, est une mesure de la liaison linéaire, c'est-à-dire de la capacité de prédire une variable (X) par une autre (Y) à l'aide d'un modèle linéaire.

Il permet de mesurer l'intensité de la liaison entre deux caractères quantitatifs. C'est donc un paramètre important dans l'analyse des régressions linéaires (simple ou multiples).

En revanche, ce coefficient est nul ($R=0$) lorsqu'il n'y a pas de relation linéaire entre les variables. Par ailleurs, le coefficient est de signe positif si la relation est positive (directe, croissante) et de signe négatif si la relation est négative (inverse, décroissante).

Ce coefficient varie entre -1 et +1 ; l'intensité de la relation linéaire sera donc d'autant plus forte que la valeur du coefficient est proche de +1 ou de -1, et d'autant plus faible qu'elle est proche de 0.

- Une valeur proche de +1 montre une forte liaison entre les deux caractères, la relation linéaire est ici croissante (c'est-à-dire que les variables varient dans le même sens)
- Une valeur proche de -1 montre également une forte liaison mais la relation linéaire entre les deux caractères est décroissante (les variables varient dans le sens contraire)
- Une valeur proche de 0 montre une absence de relation linéaire entre les deux caractères.
(Zarrouk. F.2012)

La biographie de Charles Spearman:

Charles Edward Spearman né en 1863 meurt en 1945, issu d'une famille anglaise aisée, après des études d'ingénieur et malgré une forte attirance pour la philosophie et les sciences s'engage dans l'armée britannique à l'âge de 22 ans. Il y restera 12 ans pendant lesquels son intérêt pour la psychologie, et plus particulièrement pour les travaux de Wundt et Galton, ira croissant. En 1897 il s'installe à Leipzig et entreprend un doctorat sous la direction de Wundt. En 1906 il soutiendra sa thèse.

De 1907 à sa retraite Spearman enseignera la psychologie à l'université de Londres. Il est surtout connu comme l'inventeur de la méthode d'analyse factorielle et par ses travaux sur l'intelligence. On lui doit également d'importants travaux dans le domaine de la psychométrie. Il est l'auteur de la théorie classique de la fidélité, où l'on considère que le score observé est la somme d'un score « vrai » et d'une erreur, dont il a tiré plusieurs implications : notamment le rapport entre la fidélité d'un test et sa longueur ainsi que l'atténuation des corrélations du fait du manque de fidélité des variables. Il est aussi l'auteur d'un coefficient de corrélation par rangs toujours utilisé.
(psyfact.2016)

Les travaux qui reviennent à Spearman en psychométrie

✓ Analyse factorielle:

L'analyse factorielle est une famille de techniques statistiques qui aident à déterminer les démentions communes sous-jacentes aux résultats de nombreuses mesures différentes. Ces techniques sont largement utilisées dans l'élaboration et la validation des tests. Elles jouent un rôle particulièrement important dans les inventaires de personnalité et les tests d'aptitudes intellectuelles. (Hogan, 2017, p 129)

✓ Les composantes de l'analyse factorielle:

A- La réduction des données:

La réduction des données est une des réponses nécessaire pour faire face à l'explosion du nombre d'indicateurs lorsque le nombre de mesure (variables) augmente. En statistique descriptive, la mesure sur un grand nombre de personnes (par exemple la taille, le poids ou une aptitude comme la mémoire, l'attention ou l'intelligence) peut être résumée par un indicateur de tendance centrale (ex. : la moyenne ou la médiane) et un indicateur de la dispersion de la distribution (ex. : l'écart-type ou l'écart inter-quartile). Pour une dimension mesurée, c'est-à-dire une variable, deux nombres minima sont donc nécessaires pour résumer les données. Si on étudie deux variables (la taille et le poids, les temps de réponse dans une tâche de mémorisation et la qualité de la réponse, les performances dans des tâches verbales et non verbales, etc.), 5 nombres seront nécessaires pour résumer les données : un indice de tendance centrale et un indice de dispersion pour chacune des deux variables, mais aussi un indice d'association (coefficient de corrélation de Bravais-Pearson par exemple) entre ces deux variables. Le nombre des descripteurs pour résumer les données explose avec le nombre des variables. Les techniques d'analyse factorielle auront pour intérêt (entre autres) de permettre de résumer et réduire les données.

B- Décomposition linéaire:

Le modèle utilisé en analyse factorielle est un modèle linéaire. Le principe général de l'Analyse Factorielle exploratoire (AFE) par exemple est de postuler que les variables observées sont des combinaisons linéaires de variables que l'on appelle facteurs (variables latentes sous-jacentes). Pour illustrer la démarche et comprendre les principes généraux, imaginons que deux facteurs (Dimensions psychologiques théoriques), V et W, expliquent les corrélations observées

entre 5 mesures (des tests par exemple, X1 à X5). La position d'un individu sur l'un des cinq tests s'explique alors par la position de l'individu dans chacun des facteurs (la note sur ces facteurs hypothétiques), et par sa position dans un facteur spécifique à ce test (nous nous placerons ici dans le cadre du modèle d'analyse factorielle exploratoire) c'est à dire par sa position sur un des facteurs non communs aux 5 tests.

C- Analyse en Composantes Principales (ACP)

L'objectif, lorsque l'on utilise une ACP, est de réduire les données, c'est-à-dire avoir une méthode pour obtenir un nombre réduit de composantes non corrélées. En terme clair c'est une technique d'analyse de données qui consiste à transformer des variables corrélées entres elles en nouvelles variables (composantes) non corrélées. Il faut savoir que :

- ✓ Le nombre de composantes extrait est initialement identique au nombre des variables initiales.
- ✓ Lors de l'extraction des facteurs ou des composantes (première étape de l'analyse), ceux-ci sont définis comme indépendants les uns des autres (« orthogonaux »), c'est-à-dire que la position d'un individu sur un facteur n'implique en rien sa position sur un autre facteur.
- ✓ Une hypothèse complémentaire est ajoutée pour permettre de résoudre le système : la première composante doit expliquer le plus de variance possible La seconde (orthogonale à la première) doit expliquer le Plus de la variance non expliquée, la troisième composante le plus de variance non Expliquée par les deux premières, etc.

Enfin, le plus souvent, l'analyse est faite sur des variables centrées-réduites En effet, si les variables n'étaient pas réduites (variances toutes égales à 1), et qu'une des variables composantes à une variance plus importante que les autres (quantitativement), la première composante aurait naturellement tendance à expliquer cette variable. Les réduire (ramener la variance à 1) fait que toutes les variables ont le même poids dans l'analyse.

(Raulin. 2018. p 145)

A côté de l'école française, allemande et anglaise on trouve la quatrième école qui a donné une valeur statistique et scientifique à la psychologie grâce a la discipline de la psychométrie :

9. L'école américaine :

9.1. Les fondateurs de l'école américaine:

La biographie de James McKeen Cattell:

Né en 1860 meurt en 1944, les historiens de la psychologie sont unanimes à reconnaître l'ampleur de la personnalité de James cattel et l'influence déterminante qu'il exerça sur le développement de la psychologie aux Etats-Unis. comme la plupart des psychologues américains de cette époque, cattel reçut sa formation scientifique en Allemagne il étudia avec W. Wundt à Leipzig .il rencontra ultérieurement sir Francis Galton ,qui l'accueillit un moment dans son laboratoire de Londres et dont les intérêts scientifiques coïncidaient largement avec les siens .Wundt l'introduisit à la psychologie expérimentale et c'est dans la revue de ce dernier que cattel publia ses premiers travaux .il fut en 1888,nommé professeur de psychologie à l'université de Pennsylvanie à Philadelphie, puis à l'université Columbia en 1891.

Outre celui des temps de réaction, topique majeur du laboratoire de Leipzig, cattel aborda très tôt le problème des différences individuelles.par ce biais, il fut amené à l'étude des tests, à partir de 1890.notons en passant que c'est lui qui forgea l'expression mental test .dans tous les domaines qu'il a abordés (problèmes de psychophysique, problèmes d'association, etc.)le point de vue de la performance différentielle domine nettement .il proposa, dans cette perspective, une méthode de détermination des seuils à partir de la mesure des temps de réaction.

(Thinés.G.sans date)

La biographie de David Wechsler:

David Wechsler est un psychologue américain ,né en Roumanie en 1896 et mort à new York en 1981.il est essentiellement connu comme créateur du test d'intelligence qui porte son nom et qui est aujourd'hui le principal outil de référence dans le domaine de l'examen clinique de l'intelligence.la version originale de son échelle d'intelligence ,la Wechsler-Bellevue en 1939.lors de sa première révision, en 1955,elle a pris le nom de Wechsler scale (WAIS),sous lequel elle est connue dans le monde entier.au coté de cette échelle ,Wechsler a développé un test d'intelligence spécifiquement destiné aux enfants de six à seize ans ,la Wechsler intelligence scale for children(WISC) , et un autre destiné aux jeunes enfants de trois à sept ans, la Wechsler preschool and primary scale of intelligence (WPPSI)ces trois tests ont été adaptés dans de très nombreuses langues et cultures .après la mort de Wechsler ,son éditeur a continué

à améliorer les différentes échelles ,en mettant à jour les normes et le contenu des épreuves, et en modifiant leur structure en fonction de l'évolution des modèles de l'intelligence.

Wechsler débute sa carrière en 1917 comme psychologue de l'armée américaine, ou il est chargé de l'examen individuel de l'intelligence de jeunes recrues qui, du fait de leur manque de maîtrise de l'anglais ou de leur faible niveau scolaire, ne peuvent passer les tests collectif d'intelligence. à cette occasion, il se rend compte que le QI est un indicateur imparfait des capacités intellectuelles d'un individu, car il est toujours influencé par des variables telles que la personnalité, la motivation et les opportunités d'apprentissage .comme facteurs ne peuvent être con. (Grégoire.J.sans date)

La différence entre Wechsler et cattel:

A. Les tests de Wechsler:

Les tests reposant sur l'échelle de Wechsler sont :

- ✓ Le WPPSI, pour les enfants de 2 ans à 7 ans et 7 mois.
- ✓ Le WISC (dans sa 5^e version actuellement) pour les enfants à partir de 6 ans à 16 ans et 9 mois.
- ✓ Le WAIS (dans sa 4^e version actuellement) pour les enfants de 16 ans à 69ans et 11 mois.

B. Le test de Cattell:

C'est un test de QI mais qui est spécifique à la culture nord-américaine. Le questionnaire est en américain et comportent des questions reliées spécifiquement à la culture américaine, pas seulement dans son histoire, mais aussi dans ses usages sociaux.il est donc très rare qu'une personne non américaine passe le test de cattel.

Les résultats du test de QI de cattel sont aussi valables que ceux des tests de Wechsler et il existe une table de conversion entre les deux échelles.

C. Les échelles:

Voyons des échelles de mesure du QI comme n'importe quel autre outil de mesure. Elles ont été conçues pour mesurer quelque chose de précis, ici l'effcience intellectuelle. La différence entre l'échelle de cattel et Wechsler vous pouvez la concevoir un peu comme la

différence entre une règle graduée en décimètres(Wechsler) et une autre graduée en millimètres (cattel).

Celle graduée en millimètre semble donner des résultats plus impressionnants, parce que les chiffres seront plus grands, mais en réalité, ils sont équivalents à ceux donnés par l'échelle en décimètre, mais c'est une histoire de virgule.

Un QI donnée sur l'échelle de cattel va être plus impressionnant qu'un QI donnée sur l'échelle de wechsler.et l'échelle de cattel va de 0 à 180.donc sur l'échelle de cattel, on peut obtenir un QI supérieur à 160 par contre l'échelle de wechsler elle est graduée de 0 à 160.

(Other 130-wordpress-com. la-difference-entre-wechsler-et- cattel)

10. Les caractéristiques de la psychométrie :

La psychométrie est la science de la mesure des caractéristiques psychologiques des individus. Elle consiste en l'évaluation quantifiée, objectives et standardisée des différences individuelles à l'aide d'instruments psychologiques .par différences individuelles, on entend toutes les caractéristiques psychologiques sur lesquelles les individus peuvent différer les uns des autres (par exemple : intelligence, extraversion, anxiété).

Trois caractéristiques permettent de qualifier un instrument de psychométrie soient la standardisation, l'objectivité de la mesure et les propriétés psychométriques.

- ✓ **La standardisation:** elle consiste en l'uniformité de la démarche d'administration et de correction des réponses. Un instrument psychométrique doit être administré de la même façon, selon les mêmes conditions, avec les mêmes moyens matériels et les mêmes délais, peu importe l'identité de la personne évaluée et de la personne évaluatrice. Sans standardisation la transformation des scores en comparaison à une norme est erronée (par exemple, si toutes les personnes n'ont pas exactement le même temps pour compléter les épreuves d'un test d'intelligence, ou ne reçoivent pas les mêmes consignes pour leur exécution, la comparaison de leurs résultats perd son sens).
- ✓ **L'objectivité de la mesure :** veut que les résultats obtenus résultent pas d'une évaluation subjective ; différents évaluateurs devraient arriver à des résultats similaires s'ils administraient un même test à une même personne.
- ✓ **Les propriétés psychométriques :** renseignent sur la qualité de la mesure fournie par l'instrument, notamment en termes de précision (fidélité, la sensibilité et la validité).

(Psychométrie. espace web/ instrument-psychometrique)

11. Interprétation psychologique des mesures psychométriques:

En psychologie, la notion de diagnostic est plus complexe et les examens psychométriques doivent permettre d'atteindre cinq niveaux d'objectifs:

- Décrire ce qu'il est pertinent de connaître du ou des sujets dans la situation traitée par le psychologue.
- Comprendre cette description dans son contexte.
- Prévoir les conséquences au niveau « écologique »
- Conseiller le ou les intéressés et participer à des décisions les concernant.

✓ **Décrire:** considérons par exemple la situation où une psychologue clinicienne est consultée pour l'instabilité et l'agressivité d'un enfant. Au vu des éléments dont on dispose, l'estimation du QI de l'enfant ne paraît pas particulièrement pertinente. On peut tout simplement considérer que la demande n'est pas assez travaillée pour qu'une démarche d'évaluation formelle soit envisageable. En revanche, ce serait le cas s'il s'agissait d'attester que les difficultés scolaires d'un enfant sont liées à des difficultés cognitives, ou à une mauvaise adaptation due à un QI particulièrement élevé. L'observation permise par les tests n'est pas une fin en soi, mais le moyen de faire avancer une problématique dans un contexte donné.

✓ **Comprendre :** c'est construire une représentation (J-F Richard, 1990). Pour ce qui nous concerne, il s'agit d'établir des relations entre les résultats obtenus et les conduites du sujet dans son contexte (activité et comportement). Dans le cas du conseiller d'orientation sollicité pour déterminer les compétences techniques d'un adolescent, une mesure de l'intérêt et du « don » pour la technique est opérationnalisée par:

- Un score d'intérêt « technique »
- Un score de compréhension mécanique
- Un score de raisonnement spatial.

Ces scores pourront être reliés aux résultats scolaires et à des expériences de vie significatives pour un enfant (montage d'un modèle réduit...).

- ✓ **Prévoir:** le fait de prévoir l'évolution des phénomènes à partir de différents indices constitue un objectif pour la plupart des sciences technologiques. la prédiction parfaite d'un phénomène suppose qu'il soit totalement connu du point de vue de son évolution causale, ou bien qu'il appartienne à une famille d'observations répétées. En psychologie, la prédiction se situe plutôt au plan légal que causal de l'explication. A l'exception de situations très spécifiques ou l'examen psychométrique apporte des éléments d'explication causale, deux types de sources permettent de fonder la capacité d'anticipation du psychologue : les études de validité prédictive qui font l'objet de critiques et d'efforts d'amélioration (M. Reuchlin, 1992) et l'accumulation d'observations qui construit l'« expérience » du praticien.
- ✓ **Conseiller, participer à une décision:** La capacité d'anticipation que l'on acquiert grâce à une vision et à des « paramètres » spécifiés fonde la compétence du conseiller. Deux situations radicalement différentes bien que non exclusives l'un de l'autre doivent être distinguées, selon que les résultats de la démarche sont destinés aux intéressés ou à d'autres. Lorsqu'ils sont destinés à l'intéressé, comme par exemple dans un bilan de compétences, le psychologue donne un avis sur les choix de carrière qui paraissent les plus favorables à l'épanouissement professionnel du sujet et sur les moyens de les mettre en œuvre. On travaille dans le même esprit en orientation scolaire. la situation où il s'agit de participer à des décisions concernant l'intéressé sont tout aussi délicates. Elles ne concernent pas uniquement les psychologues. les membres de commissions qui autorisent l'accès à certains cursus de formation, les responsables d'entreprise qui recrutent, renouvellent leur personnel, restructurent leurs unités de production, et autres professionnels de la justice et du social interviennent peu ou prou sur des événements qui modèlent la trajectoire de vie des personnes. Le fait d'exercer un métier qui implique le maniement d'un certain pouvoir sur les autres est une chose commune, et notre propos n'est pas d'en discuter les fondements éthiques. Le psychologue, s'il accepte de se situer dans cette logique, doit reconnaître cette responsabilité et la référer à la déontologie qui régit sa profession (J.Schlegel, 1994). (Guillevic et Vautier. 2005. pp 89-95)

12. Les erreurs de la psychométrie:

Le mot « erreur » se réfère à quelque chose de juste ou de vrai. On parle d'erreur sur une mesure physique lorsqu'on peut la comparer à une valeur de référence qu'on peut considérer comme « vraie » (par ex : mesure de la vitesse de la lumière, de la température de zéro absolu). En psychométrie les erreurs sont des types suivants :

- ✓ **Erreur systématique:** Dans une étude analytique elle peut se produire dans l'induction de sujets de l'étude, dans l'assignation aux groupes de traitement et dans la récolte données, dans l'analyse, dans l'interprétation, dans la publication et la révision des données.
- ✓ **Erreur de Berkson:** C'est un type d'erreur de sélection, qui peut se produire dans les études de cas et contrôles, dans lesquels le fait d'être malade et d'avoir été exposé au facteur de risque étudié augmente la probabilité d'être hospitalisé, ce qui donne lieu à un taux d'exposition systématiquement plus importante entre les cas hospitalisés, en comparaison avec les contrôles, aussi hospitalisés.
- ✓ **Erreur d'enquêteur:** Il s'agit d'une erreur systématique, variante de l'erreur d'information, due à la récolte sélective de données, de manière consciente ou inconsciente, par l'enquêteur. Afin de la limiter on utilise des questionnaires structurés.
- ✓ **Erreur d'information:** Erreur systématique résultante de la mesure de l'exposition (études de cas et contrôles) ou de l'évolution (études de cohortes et études cliniques non aveugle) avec une intensité différente entre les deux groupes comparés.
- ✓ **Erreur de mémoire:** Erreur systématique due à des différences dans la mémoire des faits ou expériences préalables. Dans des études de cas et contrôles il est possible que les cas se rappellent plus que les contrôles quelques unes des expériences préalables, ce qui en principe donne lieu à une surestimation de la raison d'avantages (ORA). L'erreur par sensibilisation de mémoire est une variante de l'erreur de mémoire, dans laquelle la mémoire des deux groupes de patients n'a pas été sensibilisée de la même manière ; l'entrevue avec questionnaire structuré aide à la limiter.
- ✓ **Erreur d'observation:** C'est la différence systématique entre la valeur réelle et celle enregistré. Dans une étude clinique, ceci est plus commun, elle est due à la connaissance, par l'observateur, du traitement reçu par chaque sujet.
- ✓ **Erreur de publication:** C'est une forme d'erreur de sélection, résultante de croire que les études publiés sont celles réellement effectuées.

Beaucoup d'essais cliniques ne sont pas publiés (parce qu'ils ne terminent pas, parce que le chercheur considère insignifiants ses résultats, parce que le promoteur ne le souhaite pas ou bien parce qu'ils ne sont pas acceptés pour publication). Cela se produit spécialement avec les essais cliniques qui n'enregistrent pas de différences entre les différents groupes de traitement.

D'autres essais sont publiés dans plus d'une revue médicale (publication dupliquée), elles sont parfois signées par différents auteurs, de sorte qu'il est difficile d'identifier quelques unes des publications dupliquées. Les deux phénomènes tendent à déterminer que les méta-analyses qui incluent seulement les résultats d'essais cliniques publiés surestiment en général l'effet du traitement expérimental.

- ✓ **Erreur de sélection:** Il s'agit d'une erreur systématique due au fait que les caractéristiques des sujets inclus dans une étude sont différentes des caractéristiques des patients qui ne sont pas inclus, de telle sorte que l'échantillon n'est pas représentatif de la population de référence.
- ✓ **Erreur de type I:** Il s'agit d'une erreur consistant en rejeter une hypothèse nulle, celle-ci étant vraie. L'erreur alpha est l'erreur qui est commise quand on rejette une hypothèse nulle quand celle-ci est vraie. La probabilité alpha est la probabilité de commettre une erreur de type I. Il est coutume de la fixer à $\alpha = 0.05$, c'est de dire une probabilité de 5%.
- ✓ **Erreur de type II:** Il s'agit d'une erreur consistant en accepter une hypothèse nulle, celle-ci étant fausse. L'erreur bêta est l'erreur qui est commise quand on ne rejette pas une hypothèse nulle qui est fausse. La probabilité bêta est la probabilité de commettre une erreur de type II. Il est coutume de la fixer à 10% (de sorte que $1 - \beta = 0.90$) ou bien à 20% (de sorte que $1 - \beta = 0.80$). Le terme $1 - \beta$ est le pouvoir statistique de l'étude.
- ✓ **Erreur de type III:** Il s'agit d'une erreur consistant à considérer supérieur l'efficacité d'un traitement, alors qu'en réalité celle-ci est inférieure.
- ✓ **Erreur alpha globale:** C'est l'erreur alpha qui est commise pour faire de multiples comparaisons. ([www.scientific-european-federation-osteopaths.org/Les erreurs](http://www.scientific-european-federation-osteopaths.org/Les_erreurs))

13. Les niveaux de la mesure en psychométrie:

Mesurer consiste à systématiser la réalité afin de classer et de différencier des objets en leur affectant des valeurs. En termes mathématique, c'est une opération qui consiste en une application d'un ensemble O d'objets réels, qui peuvent être des observations psychologiques, dans un ensemble N de catégories symboliques. Cette conception de la mesure va permettre de préciser les conditions d'attribution de propriétés numériques à des objets psychologiques.

(Guillevic et Vautier. 2005. p 51)

Stevens en 1946 propose de classer ces échelles de mesure en fonction des propriétés des nombres qui sont conservées. Cette classification est très critiquée par les statisticiens (Velleman et Wilkinson, 1993) mais il est traditionnel en psychologie (pour la formation) de distinguer ces quatre grands types de mesures.

13.1. Types d'échelles :

Les variables sont évaluées ou quantifiées à l'aide d'échelles de mesure. Stevens (1951) a classé les échelles en quatre types, auxquels la statistique et le testing font souvent référence. Il est important de bien distinguer ces types d'échelles, car certaines opérations arithmétiques et statistiques peuvent s'appliquer à certains d'entre eux, mais non à d'autres. Les échelles sont classées selon leur niveau de sophistication, au niveau le moins perfectionné ou le plus primitif est:

- **Echelle nominale:**

Une échelle nominale répartit les observations dans un certain nombre de classes disjointes, telles que chaque observation entre dans une seule classe. L'ensemble des classes utilisées constitue l'échelle nominale. En d'autres termes, on effectue une partition de l'ensemble des observations (application d'une relation d'équivalence) et tous les objets ou les sujets d'une même classe sont considérés comme équivalents. Attention : ce n'est pas un critère statistique qui définit la partition, c'est un critère empirique. Ce critère détermine la signification à attribuer à la mesure.

Exemple d'échelle nominale: lors d'une étude sur l'entretien clinique, toutes les interventions d'un psychologue sont classées en trois catégories : Interprétations (I), Clarifications (C) et Reformulations (R). Cette catégorisation constitue une échelle nominale si et seulement si on effectue une partition des interventions c'est-à-dire, si et seulement si, chaque intervention entre dans une seule catégorie I, C, ou R. Dans le cas où il serait impossible de faire entrer les interventions dans une seule catégorie, on devra ajouter d'autres classes à l'échelle ou redéfinir la règle de partition utilisée. **(Roulin. 2018. p 11)**

Un autre exemple sur l'échelle nominale: un psychologue qui veut travailler sur les relations entre l'attitude éducative des parents ou des enseignants (A) et l'agressivité des enfants (E).

Devra construire des échelles nominales significatives pour ces dimensions psychologiques telles que : A= (autoritaire ; non-directif ; laisser-faire) et E= (agressivité extériorisée ; intériorisée ; pas d'agressivité).s'il peut établir des relations significatives statistiquement entre les classes des deux échelles, ce psychologue disposera ainsi d'outils pour agir avec moins d'incertitude sur l'attitude des parents ou le comportement des enfants.

(Guillevic et Vautier. 2005. p 54)

- Echelle ordinale:

À un second niveau, celui des échelles ordinales, on prend en compte une nouvelle propriété des nombres, leur ordination. Si l'on peut regrouper les individus en classes ordonnées, il est possible, et même naturel, de désigner ces classes par des nombres. On peut par exemple, dans un domaine donné, définir cinq niveaux d'efficience intellectuelle et affecter les

Individus à chacun de ces niveaux. La variable « efficience intellectuelle » est alors une variable ordinale. L'ordre des classes d'une échelle ordinale est invariant, ce qui n'était pas le cas avec les échelles nominales (il peut seulement être inversé). Les distances entre les

Classes ne sont pas spécifiées. Si nous reprenons l'exemple de cinq niveaux d'efficience

Intellectuelle, rien ne permet de dire que la différence entre les deux premiers niveaux (1 et 2) est identique à la différence entre les deux niveaux suivants (2 et 3). Il n'y a donc pas de sens à calculer des différences arithmétiques entre les individus ou à parler de la forme de la distribution de l'efficience intellectuelle. Les nombres n'ont pas d'autres propriétés que les lettres rangées par ordre alphabétique. Au niveau ordinal, les indices permettant la

Description de la variabilité des individus deviennent plus puissants (médián pour la tendance centrale, écart interquartile pour la dispersion). La corrélation entre deux variables ordinales (coefficients de Spearman ou de Kendall) est une mesure de la similarité entre

Deux ordinations.

(Huteau.M. 2013. p51)

Sur une échelle ordinale les objets portent des numéros qui en indiquent l'ordre, comme un trait plus ou moins important, sans préciser les distances qui les séparent sur l'échelle. Le classement par ordre de grandeur constitue un exemple d'échelle de grandeur. par exemple, les classements des équipes de football (1, 2, 3,4,.....25) se font sur une échelle ordinale .sachant cela , il est impossible de supposer que la différence de performance entre l'équipe 2 et l'équipe

3 est égale à la différence entre l'équipe 1 et l'équipe 2. en fait ,il est possible que l'équipe 1 ne soit que légèrement meilleure que l'équipe 2 ,alors que l'équipe 3 peut être très inférieure à l'équipe 2. (Hogan. 2017. p 46)

Un autre exemple : si l'on reprend l'exemple du psychologue qui veut travailler sur les relations entre l'attitude éducative des parents ou des enseignants (A) et l'agressivité des enfants (E)

On perçoit la difficulté de construire de telles échelles : quels sont les critères pour décider que telle conduite est plus autoritaire que telle autre et que telle conduite est plus agressive que telle autre ? Atteindre ce niveau de la mesure permet d'établir une relation entre le développement de l'agressivité et la croissance de l'autorité, mais la vérification statistique d'une telle relation ne permet pas d'en déduire une relation causale : on peut tout aussi bien déduire que A détermine E ou l'inverse et même qu'une variable C (par exemple le niveau socio-économique des parents) détermine à la fois A et E. (Guillevic et Vautier, 2005. p 54)

- Echelle d'intervalle:

A la notion d'ordre va s'ajouter l'idée de distance entre les classes. Dans les échelles d'intervalles, on peut exprimer l'égalité des distances entre les classes. Non seulement on peut dire que la classe NC est plus grande que la classe NA, mais que la distance qui sépare NC de NA est deux fois plus grande que la distance qui la sépare de NB. Dans ce type d'échelle apparaissent les notions d'origine et d'unité arbitraire, dont l'exemple le plus classique est celui de la température qui permet de bien comprendre la signification de ce niveau de la mesure. Dans l'échelle Celsius, on a une origine arbitraire (le 0° de l'eau qui gèle) et une unité arbitraire (la division en 100 parties égales de la distance entre le 0° et le 100° de l'eau qui bout). On peut dire que la distance qui sépare 50° de 0° est deux fois plus grande que la distance qui sépare 50° de 25°, mais on ne peut pas dire qu'il fait deux fois plus chaud à 50° qu'à 25°. On peut simplement dire que l'écart entre 25° et 50° est le même qu'entre 50° et 75°.

Sur le plan du traitement statistique , on peut alors calculer des indices plus puissants : la moyenne (somme des valeurs , divisée par les effectifs) et l'écart-type de la distribution (étalement des valeurs autour de la moyenne)qui permettent des comparaisons de performances en s'appuyant sur un test statistique comme le T de student ; la corrélation entre deux variables qui permettent d'établir des pronostics non plus fondés sur des rangs mais sur

les intervalles entre les catégories d'observations-ce que permettent les coefficients de corrélation comme le R de Bravais-Pearson.

La définition de critère permettant de définir des intervalles égaux étant particulièrement ardue, les psychologues utilisent très peu ce niveau de la mesure .il est en effet peu vraisemblable que l'on puisse estimer des distances entre attitudes éducatives qui permettraient d'affirmer que tel parent est deux ou trois fois plus autoritaire que tel autre. Rappelons que parfois les pédagogues et les psychologues manipulent leurs variables à un niveau d'échelles d'intervalles sans avoir pour autant satisfait à l'exigence de définir des critères d'intervalles. Par exemple, significatifs au plan pédagogique, qui leur permettent d'affirmer que les distances qui séparent 6 /20 et 18/20 de 12/20 sont égales, ou score que la distance de 4 à 10 est trois fois plus grande que celle séparent 8 de 10. On voit les absurdités auxquelles peuvent conduire de tels calculs et pourtant nul n'ignore les résistances à utiliser, en France, des échelles de notation en lettres qui permettent de se situer à un niveau plus faible de la mesure. **(Guillevic et Vautier. 2005. p 55)**

- **Echelle de rapport :**

L'échelle de rapport place les objets en ordre à intervalle égaux et comporte un zéro absolu. La plupart des mesures physiques qui font partie du quotidien, comme les longueurs et les poids, concernent des échelles de rapport.

Une échelle de rapport (ou de ratio) est une échelle d'intervalle dans laquelle le point zéro n'est pas arbitraire (comme le temps de réponse ou une mesure de vitesse). Ce type d'échelle est rarement utilisé ou plus exactement les propriétés de ce type d'échelle sont rarement utilisées en psychologie. Les échelles de rapport représentent des rapports car elles ont une origine absolue (correspondant à l'absence de l'attribut mesuré). Par exemple, la distance a pour origine 0 (absence de distance) et 40 mètres est deux fois plus loin que 20 mètres. Ce n'est pas le cas d'une échelle d'intervalle comme la température exprimée en Celsius ou le 0° est arbitraire. Une température de 40° n'est pas deux fois plus chaude que 20°. Pour connaître le rapport entre ces deux températures, il aurait fallu prendre une mesure absolue de la température en Kelvin (qui est une échelle de rapport) et on peut alors comparer les deux mesures en Kelvin et en faire le rapport. **(Raulin. 2018. p12)**

Remarquons cependant que des variables utilisées par les psychologues peuvent se situer au niveau d'échelles de rapports comme par exemple : l'âge, les revenus des parents, etc.

cependant aucun psychologue n'affirmera que tel enfant de quatre ans est deux fois plus développé qu'un enfant de deux ans ou que disposer d'un revenu de 12000 F donne un sentiment de richesse deux fois plus élevé qu'un revenu de 6000 F.

(Guillevic et Vautier. 2005. p 56)

13. Les outils de la psychométrie (les tests psychologiques)

Etymologiquement, le test est une épreuve dont les modalités d'administration des consignes de passation, de correction et de cotation ont été standardisées. Il permet de mesurer, de repérer ou simplement d'apprécier chez un individu une caractéristique sensorielle, sensori-motrice psychomotrice, mentale, tempéramentale, caractérielle ou affective. Il permet de situer le sujet dans une échelle dite étalonnage qui a été calculée à partir des résultats des autres membres de la classe sociale, économique, professionnelle ou culturelle à laquelle il appartient ce qui permettra par exemple de dire d'un manoeuvre de la région parisienne que pour l'intelligence il se situe entre 30ème et 40ème sujet sur 100 de la population de référence. (Robert et Denise. 2016. p 19)

14.1. La définition d'un test :

Dans sa définition la plus connue se définit par P.Pichot 1954 comme suit : « une situation expérimentale standardisée servant de stimulus à un comportement .ce comportement est évalué par une comparaison statistique avec celui d'autres individus placés dans la même situation, permettant ainsi de classer le sujet examiné, soit quantitativement, soit typologiquement ». (Noci .S.sans date)

14.2. Les caractéristiques principales d'un test:

- Echantillon d'un type de comportement spécifique de façon systématique : on veut insister sur la diversité de comportement et de type de tests.
- Standardisation, des conditions d'échantillonnage du comportement : différences de l'examineur peut influencer.
- Règles de cotation, pour attribuer un chiffre aux comportements : ce n'est pas forcément un chiffre mais peut être des qualificatifs, l'objectif est que quelque soit l'évaluation on obtienne le même résultat.

(Cours psycho.blog4ever.com/psychometrie)

14.3. Catégorisation des tests:

○ La sphère cognitive:

La plupart des tests dans le domaine cognitif relèvent d'une approche à la fois empirique et descriptive basée sur l'analyse factorielle (M. Reuchlin 1964). A partir d'épreuves variées, on cherche à dégager des facteurs statistiques qui rendent compte des performances obtenues et à les interpréter comme les manifestations de certaines formes ou modalités d'intelligence. A cette conception dite factorielle de l'intelligence s'opposent des tentatives d'opérationnalisation reposant sur des bases théoriques plus élaborées.

Sur le plan pratique, on peut se contenter de la valeur normative des mesures pour détecter des problèmes particuliers (retards de développement, détériorations, sous-fonctionnements ...) liées à des valeurs anormales. Souvent cependant, le psychologue a besoin de savoir ce que représente la mesure dans le contexte du problème qu'il a à résoudre. Jusqu'à présent, la prise en compte de ce que représentent concrètement les mesures de l'intelligence s'est faite principalement dans le cadre de validité prédictive.

(Guillevic et Vautier. 2005. p 36)

Cette catégorie se subdivise en tests individuels d'aptitudes intellectuelles, en tests collectifs d'aptitudes intellectuelles et en divers tests d'aptitudes autres que ceux portant sur l'intelligence. L'échelle d'intelligence de Wechsler pour adultes (WAIS) est un exemple de test individuel d'aptitudes intellectuelles. L'échelle d'intelligence de Sanford-Binet en est un autre. Ces deux tests sont administrés individuellement par des psychologues d'expérience afin d'établir un indice des aptitudes mentales générales de la personne.

Il existe de nombreux autres types de tests d'aptitudes intellectuelles, notamment les tests de mémoire, de raisonnement quantitatif, de pensée créatrice de vocabulaire et d'aptitudes spatiales. Ces fonctions mentales sont parfois incluses dans les tests d'aptitudes intellectuelles et sont d'autres fois évaluées séparément. (Hogan. 2017. p 05)

○ la sphère conative :

Dans ce domaine, on évalue des dimensions liées à la motivation, à la propension à prendre du plaisir et à s'intéresser à tel domaine d'activité scolaire ou professionnel.

Contrairement aux tests d'intelligence, les tâches consistent non pas à produire une performance, mais à répondre à des questionnaires qui font appel à la connaissance que l'on a de soi-même de la manière dont on fonctionne ou de ses goûts et préférences. On peut évaluer les intérêts pour différents domaines d'activités, les intérêts pour le travail scolaire, les valeurs socioprofessionnelles et certaines composantes motivationnelles comme le besoin de réussite.

Les tests sur les valeurs et les tests de motivation tentent de mettre en évidence des dimensions plus générales avec l'inconvénient d'avoir une valeur prédictive limitée. Les tests sur les valeurs professionnelles tentent de cerner l'importance accordée au salaire, à l'autonomie, au prestige à la sécurité... l'utilisation de ces tests est orientée vers une aide à la construction d'une image de soi. Les tests de motivation veulent mesurer le besoin de réussite, l'absence de peur de l'échec, le goût de l'effort et de la ténacité, la résistance d'un projet à l'échec. Les dimensions envisagées font que certains éditeurs rangent ces tests dans la rubrique de la sphère affective. **(Guillevic et Vautier. 2005. p 40)**

○ **La sphère affective :**

La troisième catégorie comprend un ensemble de tests visant à obtenir de l'information sur la personnalité d'un individu. La première subdivision inclut les tests objectifs de personnalité. Dans le domaine des tests « objectifs » signifient simplement que la notation des tests est objective, ou impartiale, et est basée sur des réponses à des items du type vrai ou faux ou autre type de choix de réponses. L'inventaire multiphasique de la personnalité du Minnesota (Minnesota Multiphasique personality Inventory, ou MMPI). L'inventaire de dépression de Beck (Beck Depression Inventory, ou BDI) et l'inventaire des troubles alimentaires (Eating Disorder Inventory, ou EDI) en sont des exemples. Le MMPI-2 fournit un profil comparant les réponses du candidat à celles de plusieurs groupes cliniques. Comme l'indiquent leurs noms le BDI-II et l'EDI-3 tentent respectivement de mesurer la dépression et les troubles alimentaires. Pour des raisons pratiques et pour plus de clarté, ces tests objectifs seront classés en deux groupes : les tests qui mesurent les traits normaux de la personnalité et ceux qui servent d'instruments cliniques pour mesurer les troubles pathologiques ou invalidants.

La deuxième subdivision des tests de personnalité est celle des techniques projectives. Toutes ces techniques présentent au candidat une tâche assez simple. Mais non structurée. On espère que les réponses du candidat révéleront les facettes de sa personnalité. La plus connue de ces techniques est le test de Rorschach, aussi appelé « test des tâches d'encre » ou simplement

« Rorschach ».les autres techniques consistent notamment à demander au candidat de dessiner des visages humains et de compléter des phrases ou à observer ses réactions à certaines images .les mesures de la personnalité comprennent une troisième subdivision ,simplement intitulée « autres approches »,qui regroupe les autres méthodes élaborées par les psychologues pour satisfaire notre fascination sans bornes pour la personnalité humaine. (Hogan. 2017. p 06)

14.4. Les tests « périphériques »:

Tous tests que nous avons mentionnés se situent dans le domaine de compétence exclusif du psychologue ; il peut aussi utiliser des tests qui ne mesurent pas des traits psychologiques mais qui sont indispensables pour compléter son diagnostic. On distinguera essentiellement trois grandes familles : les tests d'évaluation scolaire et de connaissance ; les tests moteurs ; les tests physiologiques.

A) les tests d'évaluation scolaire et de connaissances:

Ces tests devraient relever de la compétence des pédagogues mais le psychologue les utilise pour confirmer ses informations sur le niveau de réussite objectif des sujets qu'il examine .les épreuves sont particulièrement nombreuses en ce domaine et certaines offrent l'avantage de relier les connaissances au fonctionnement psychologique du sujet.

B) les tests moteurs:

Ces tests tentent de mettre en évidence la coordination psychomotrice des sujets.ils visent à prédire la réussite des dans des tâches impliquant une activité de manipulation sur outils et machines.de nombreux tests visent à mesurer la coordination et la dextérité manuelles, la coordination Visio-motrice, la précision des mouvements, la coordination motrice centrale (capacité à réagir rapidement et adéquatement avec ses mains et ses pieds)

C) les tests physiologiques:

Ces tests permettent de mesurer certains indices comportementaux sur les capacités physiologiques comme l'acuité visuelle ou auditive, la vision des couleurs, la capacité respiratoire. Le choix de ces tests est déterminé par le problème sur lequel doit intervenir le psychologue.

(Guillevic et Vautier. 2005. p 45)

14.5.Principaux utilisateurs des tests :

Les tests sont utilisés dans différents domaines et dans différents secteurs (école, santé mentale, orientation scolaire et professionnelle, médical, recherche, etc.). Parmi les utilisateurs on peut citer: les psychologues cliniciens, les neuropsychologues, les psychologues scolaires, les psychologues du travail (sélection du personnel et orientation). Ces exemples concernent essentiellement différentes facettes du métier de psychologue mais d'autres professions peuvent utiliser des tests comme les orthophonistes ou encore, avec des épreuves plus ou moins standardisées, les médecins, les psychomotriciens, les ergothérapeutes ou encore des enseignants. Dans le domaine de la recherche (en psychologie ou en sciences de l'éducation) le test peut jouer des rôles variés et contrairement aux exemples précédents, l'objectif est rarement de contribuer à une évaluation individuelle. Le plus souvent, en recherche, ces tests contribuent à sélection d'échantillon, à la description de population ou encore peuvent servir de mesures d'intérêt (soit comme variable dépendante ou comme variable contrôlée). **(Roulin, 2018, p 37)**

15.Les conditions d'un bon test :

Un test se définit par trois indicateurs : la sensibilité, fidélité et validité.

A)- Fidélité:

La fidélité est une propriété générale de tout test, échelle, épreuve, renvoyant au degré avec lequel une mesure est entachée d'erreur. a tout instrument de mesure induit des erreurs dans la mesure, ce qui ne signifie pas pour autant que cet instrument ne soit pas fiable. Mais il est illusoire de considérer un instrument comme parfait, aussi parle-t-on de fidélité pour refléter la proximité de la mesure avec la valeur réelle que l'on mesure.

1)- La définition de la fidélité:

- En effet, selon Beech (1994) et Bernaud (2007), un test fidèle donne la même mesure d'un « trait » quel que soit le moment où l'examineur propose le test, quelle que soit la partie donnée, et quel que soit l'administrateur de l'instrument.

(Thomas.M. 2019. p27)

- Un test fidèle est un test avec une erreur de mesure faible. La fidélité est donc un indicateur de la précision et de la constance des scores. Plus un instrument est fidèle, plus le score observé sera proche du score vrai. **(Roulin. 2018. p 96)**

- La fidélité d'un test évoque des dimensions technique et quantitative. En français les meilleurs synonymes de fidélité technique seraient « constante » et « reproductibilité ». un test fidèle, en psychométrie donne de façon constante le même résultat ou des résultats similaires pour un individu. Le résultat peut être reproduit à tout le moins à l'intérieur d'une certaine marge d'erreur. il importe de distinguer le changement réel du trait mesuré et les fluctuations de résultats attribuables à une modification temporaire des circonstances personnelles.

(Hogan. 2017. p 74)

La fidélité d'un test renvoie ainsi au degré avec lequel une mesure est obtenue avec peu d'erreur. il y a dans cette idée, la **notion de stabilité dans le temps** (une mesure indépendante du temps- réalisée a un moment donné doit fournir la même valeur à un autre moment pour lequel les mêmes conditions sont réunis), **la notion de précision** (une mesure doit être suffisamment précise pour déterminer la valeur sans trop d'erreur), **la notion de discrimination**(ou plus généralement de sensibilité).

- Toute procédure de mesure s'accompagne d'erreurs de mesure. Celles-ci sont de deux types : des erreurs aléatoires, variables d'une observation à l'autre et imprévisibles au niveau de contrôle des observations choisi, et des erreurs systématiques qui se manifestent de la même manière d'une observation à l'autre. La théorie de la fidélité traite uniquement des erreurs aléatoires. Cette théorie s'applique à des mesures ayant les propriétés des échelles d'intervalles. **(Huteau et Lautrey. 2003. p 97)**

A. les erreurs aléatoires, spontanées réalisées au moment de l'observation:

Afin d'évaluer la fidélité relative au moment de l'observation (stabilité ou constance), les observations sont répétées à deux moments différents (méthode test- retest). La stabilité est alors mesurée par le coefficient de corrélation entre les deux observations. Une corrélation élevée signifie que les individus se classent de la même manière aux deux moments et témoigne d'une bonne stabilité, c'est à- dire d'un faible poids des facteurs aléatoires associés au moment de la passation. (Portant sur les classements cette stabilité est tout à fait compatible avec des changements lorsque ceux- ci ne bousculent pas trop l'ordre des sujets.) Une corrélation faible signifie que les individus se classent de manière différente d'une

Observation à l'autre et indique une mauvaise stabilité des observations. La conduite des sujets est alors affectée par des facteurs qui se manifestent seule ment à un moment et

Conduisent ainsi à la différence des classements (un sujet, par exemple, peut être très motivé pour des raisons diverses et aléatoires lors de la première observation, et l'être beau coup moins lors de la seconde, ce qui, le conduira alors à être moins bien classé).

Mais pour expliquer la corrélation non parfaite entre le test et le retest on peut évoquer d'autres facteurs qu'il est difficile de considérer comme des erreurs aléatoires. Lorsque l'observation sollicite l'activité du sujet, ce qui est le cas dans les tests, celui-ci, nous l'avons déjà noté, se modifie. En passant un test, le sujet apprend au moins à passer ce test et les scores à la seconde passation sont toujours plus élevés. Cet apprentissage étant plus ou moins marqué selon les individus, il contribuera à la réduction du coefficient de stabilité. Lorsque les sujets passent le test pour la seconde fois, on leur propose une tâche déjà connue et certains d'entre eux peuvent ne plus être très motivés ce qui contribuera aussi à la réduction du coefficient de stabilité. Si l'intervalle entre le test et le retest est long, si les sujets ont été soumis à des apprentissages différents et s'ils se sont développés à des rythmes divers et dans des directions différentes, les coefficients de stabilité ne peuvent plus être interprétés comme des coefficients de fidélité. Le coefficient de stabilité est donc un indice ambigu qui ne nous renseigne sur le poids des erreurs aléatoires que lorsque l'intervalle entre le test et le retest est court

(Huteau.M. 2013. p40)

B. les erreurs systématiques relatives au choix des situations :

L'erreur systématique. Cette erreur est une "déviations «constante, négative ou positive introduit par l'instrument. De façon plus générale on parle d'erreur systématique quant, par rapport à une valeur de référence x , l'instrument donnera toujours comme valeur observée $x+b$ (déviations positive ou négative). Par exemple, pour un instrument comme votre balance, si elle affiche systématiquement "+ 2 kilogrammes" par rapport au poids réel, l'erreur systématique est de +2 kg. Dans les tests mentaux une des causes possibles de cette erreur systématique est le biais d'échantillonnage lors de l'étalonnage de l'épreuve.

(Roulin. 2018. p 98)

La définition de la situation d'observation est toujours fonction des idées que

L'on a sur la nature des différences individuelles que l'on se propose d'observer. Ces idées étant relativement générales, elles peuvent donner naissance à diverses Opérationnalisations.

On doit alors se demander si la variabilité que l'on observe pour un type d'opérationnalisation est équivalente à celle que l'on observe pour d'autres opérationnalisations. Si oui, on a

Distingué les individus selon une propriété relativement générale, sinon ils ont été différenciés sur un caractère spécifique. Afin d'apprécier le degré de généralité des Individuelles (elles constituent des formes parallèles). Par exemple, deux questionnaires de personnalité parallèles seront constitués de questions différentes, mais se rapportant aux mêmes traits. On examinera ensuite dans quelle mesure les individus se classent de la même manière dans les deux questionnaires. Le coefficient de corrélation entre les deux situations est un indice de leur équivalence (on l'appelle coefficient d'équivalence). S'il est élevé, on évalue une caractéristique psychologique relativement générale. On dira alors que le poids des erreurs attribuables au choix de la situation particulière est faible ou que, relativement à ces sources d'erreur, l'observation est fidèle. **(Huteau.M. 2013. p42)**

Donc le concept de fidélité est étroitement lié à l'idée de mesure. Or depuis Heisenberg et son principe d'indétermination (incertitude), toutes les sciences ont conscience du fait qu'aucun instrument de mesure ne sera jamais parfait, aucune mesure ne reflètera parfaitement la réalité à cause de ces grand type d'erreurs. (Les erreurs aléatoires et les erreurs systématiques).

Les composantes de la fidélité générale d'un test:

A. la fidélité relative au choix des items:

Il s'agit de vérifier la cohérence des items entre eux. On vérifie donc que le choix des items n'a pas entraîné de biais dans les mesures, notamment, à l'aide de ces deux principales techniques :

○ méthode des formes parallèles:

Dans la pratique deux formes parallèles peuvent se rencontrer dans deux situations que nous allons maintenant décrire :

La stabilité:

L'on administre le même test à deux reprises, la corrélation entre les scores observés au test-retest nous donne une indication de la stabilité des résultats dans le temps. Le test administré au temps A est considéré comme parallèle au même test administré au temps B. si

les résultats au test-retest ne sont pas stables, alors la corrélation entre les deux sera faible et l'effet du passage du temps s'ajoutera à l'erreur de mesure. Il faut noter qu'une telle procédure suppose que le retest est sans effet particulier sur les sujets, c'est-à-dire qu'il n'y a pas eu d'effet d'apprentissage ou de contamination des résultats. Si par exemple, les sujets plus forts lors de la première administration sont aussi ceux qui, au moment du retest, se rappellent mieux des questions posées la première fois, il risque d'y avoir corrélation entre le score vrai de l'élève au premier test et erreur aléatoire de mesure au second.

-L'équivalence :

Si l'on administre deux versions d'un même test, la corrélation entre les scores de chaque test nous renseigne sur le degré d'équivalence entre les tests. Ceci suppose que les deux formes ont été administrées en même temps ou à l'intérieur d'une période de temps très courte, sinon la stabilité et l'équivalence des deux tests seraient mesurées simultanément. Ce type de fidélité requiert que deux tests soient créés. Ce n'est pas toujours nécessaire cependant. On peut décider de considérer comme équivalentes les deux moitiés d'un test (méthode bissection). Le calcul de la corrélation nous fournit alors une estimation de l'équivalence des résultats pour chaque moitié du test. A la limite, on peut étendre ce concept jusqu'aux items et déterminer à quel point tous les items entrant dans le calcul d'un score total sont homogènes, c'est-à-dire équivalents ou encore parallèles. L'ennui avec le calcul de la fidélité par la méthode de bissection c'est que l'estimation fournie se fonde sur une partie du test, alors que c'est la fidélité du test entier que nous recherchons.

(Laveault et Grégoire. 2014. p 109)

o Méthode de bissection ou Split-Half:

On éprouve la de la mesure au niveau des différents éléments qui permettent ; ainsi par exemple, on calculera la corrélation entre les résultats aux items pairs et aux items impairs d'un test. On vérifie par là que le test se comporte de manière homogène. Si toutes les épreuves ne se prêtent pas facilement à cette partition, en particulier quand le nombre d'items est réduit, cette procédure présente l'avantage d'être peu coûteuse (une seule passation) et elle permet de supprimer les causes de variations externes au test.

Dans la perspective qui est de mettre l'accent sur la signification psychologique des objets cette notion d'homogénéité nous paraît particulièrement importante. Elle permet de vérifier que les

items qui composent l'épreuve renvoient au même type de difficulté ainsi qu'au trait psychologique.

Des procédures spécifiques ont été mises au point pour vérifier l'homogénéité d'un test .En voici deux exemple :

- **la méthode item-score:** s'appuie sur l'idée qu'un test est d'autant plus homogène que la liaison des résultats de chaque item aux résultats de l'ensemble du test est forte.

L'évaluation de cette liaison peut se faire à l'aide de différents tests statistique choisis en fonction du niveau de la mesure (T de Student, P de Spearman, R de Bravais-Pearson ...).

(Guillevic et Vautier. 2005. p 59)

B. la fidélité relative au moment de l'expérience:

*** Interprétation du coefficient de corrélation de Pearson :**

Pour être interprété, le coefficient de corrélation doit être significatif (la valeur de p doit être plus petite que 0,05). Si le coefficient est non significatif, on considère qu'il est semblable à $r = 0$. Par contre, lorsqu'il est significatif, le coefficient de corrélation donne deux informations importantes :

- **Le sens de la relation linéaire entre les deux variables :** Le coefficient de corrélation, qui présente finalement la covariance standardisée, varie entre - 1 et 1. Un coefficient de 1 indique une corrélation positive parfaite entre les deux variables. À l'inverse, un coefficient de - 1 indique une corrélation négative parfaite: lorsque la variable x augmente, la variable y diminue dans la même proportion. Dans les deux cas, les points tombent parfaitement sur la droite. Un coefficient de 0 indique qu'il n'y a aucune relation entre les deux variables. Ainsi, la variation de l'une n'est aucunement associée à la variation de l'autre.
- **La force de la relation linéaire entre les deux variables :** Plus la valeur du coefficient est proche de + 1 ou de - 1, plus les deux variables sont associées fortement. Au contraire, plus le coefficient est près de 0, moins les variables partagent de covariance et donc, moins l'association est forte. On peut qualifier la force de cette relation avec les balises de Cohen concernant la taille d'effet.
- **La taille d'effet :**
Pour la corrélation, nous n'avons pas à effectuer de calcul particulier pour connaître la taille d'effet. Nous regardons seulement la valeur du coefficient et nous l'interprétons selon les balises de Cohen (1988) :
 - Autour de 0,10 : effet de petite taille, corrélation faible.
 - Autour de 0,30 : effet de taille moyenne, corrélation moyenne.

- Plus de 0,50 : effet de grande taille, corrélation forte.

(Psychométrie dans l'usage/corrélation)

C. la fidélité inter –juges :

Cette technique, différente des précédentes, est utilisée dans les cas où il peut y avoir ambiguïté dans l'évaluation (cotation) des résultats au test, évaluation qui peut être entachée de subjectivité (par exemple : certaines mesures utilisant des techniques projectives). La principale source de l'erreur de mesure étant le cotateur, la mesure de la fidélité consiste à évaluer s'il existe un degré d'accord suffisamment élevé entre les jugements de plusieurs observateurs. Plusieurs indicateurs statistiques permettent d'évaluer l'accord inter-juge, mais contrairement aux autres indices de fidélité, il n'y a pas de consensus véritable sur l'interprétation de ces coefficients qui font parfois l'objet de critiques importantes (pour le kappa de Cohen, cf. par exemple le billet de Stéphane Vautier "Le kappa de Cohen : une solution à un faux problème"). Pour ceux que cela intéresse, on peut aussi utiliser des coefficients de corrélations intra classe (Shrout & Fleiss, 1979). Cette méthode se base sur les résultats d'une analyse de variance prenant en compte comme sources de variation, le facteur sujet (facteur aléatoire), le facteur juge et l'interaction entre ces deux facteurs. Ce coefficient s'interprète comme les autres coefficients de fidélité. **(Roulin. 2018. p 104)**

- **Méthodes d'amélioration de la fidélité :**

Améliorer la fidélité d'une épreuve peut prendre beaucoup de temps. Pour élever la fidélité on peut :

- Améliorer la standardisation de l'épreuve.
- Améliorer les critères de cotation.
- Faire diminuer les erreurs aléatoires en contrôlant un maximum de variables.
- Ecarter les items atypiques.
- Améliorer l'échantillonnage des items.
- Augmenter le nombre des items.

B- La validité:

En psychométrie, la validité d'un test a toujours été considérée comme le concept le plus fondamental et le plus implorants .pour les concepteurs comme pour les praticiens, l'essentiel est en effet d'être assuré de mesurer ce qu'ils veulent mesurer et uniquement cela. La précision

de la mesure est certes importante, mais elle est inutile si le test n'évalue pas, ou évalue mal le phénomène visé. **(Laveault et Grégoire. 2016.p 163)**

La définition la plus connue de la validité est « le degré auquel un test mesure ce qu'il prétend mesurer ». **(Hogan. 2017. p 102)**

Donc la validité d'un test est sa qualité à mesurer effectivement ce qu'il est censé mesurer. Un test psychologique valide mesure de façon correcte la dimension psychologique pour la mesure de la quelle il a été conçu.

1- Terminologie plus ancienne:

Dans la plupart des manuels de psychologie, le concept de validité est présenté en reprenant la distinction proposée initialement par Cronbach(1984) puis messick (1989).cette terminologie ne doit pas conduire à penser qu'il existe plusieurs formes de validité. Ce sont des aspects ou des méthodes qui participent au processus de validation ou preuves de la validité d'un test. **(Roulin. 2018. p 108)**

Pour être valide, un test ou une épreuve doit respecter la validité de plusieurs des ses composantes : validité de contenu, validité empirique et la validité théorique ou hypothético-déductive.

A. La validité de contenu:

La validité de contenu concerne la relation existant entre le contenu d'un test et certains domaines bien définis de la connaissance et du comportement, pour que le contenu d'un test soit valide, il doit bien correspondre au contenu du domaine étudié .l'application de la validité de contenu implique souvent la notion d'échantillonnage, c'est-à-dire que le contenu du test doit couvrir un échantillon représentatif de tous les éléments possible du domaine, ce qui n'est pas toujours le cas. Le test peut couvrir toute la matière du domaine, mais celui-ci est habituellement trop étendu pour qu'une couverture complète soit possible, d'où l'utilisation d'échantillons. **(Hogan. 2017. p 106)**

Rassembler des preuves de validité sur la base du contenu d'un test consiste à apprécier dans quelle mesure les différentes composantes de ce test permettent une évaluation correcte du concept visé. Le terme composant est utilisé à dessein trop souvent, les preuves basées sur le contenu se focalisent sur les seuls items. Cette composante est essentielle, mais ne constitue

pas la totalité des preuves de la validité de contenu des résultats à un test .il est également nécessaire d'évaluer les instructions données aux sujets les modalités de présentation des stimuli (p.ex. présentation papier-crayon ou sur écran) les contraintes de temps, les modalités de réponse et les critères de cotation. Toute ces composantes du test contribuent à une mesure valide du concept visé.la récolte des preuves de validité inclura bien entendu une évaluation de l'adéquation du contenu des items.

(Laveault et Grégoire. 2014. p 171)

Donc la validité de contenu correspond à la possibilité de décrire complètement le contenu de l'épreuve et ce que l'on veut mesurer, le test doit donc présenter des items mesurant effectivement la dimension psychologique .ces items doivent être exhaustifs et représentatif de l'ensemble des questions que l'on veut se poser.

B. La validité empirique

Une observation psychologique est dite valide lorsqu'elle permet d'atteindre de manière satisfaisante les objectifs visés par celui qui l'a élaborée. Il y a donc autant de types de validité que de catégories d'objectifs. On distingue classiquement des objectifs pratiques et des objectifs théoriques.

On peut attendre de l'observation psychologique qu'elle fournisse une aide à la décision et à l'action dans des situations très diverses : par exemple, diagnostic d'un état des connaissances ou des compétences en vue de l'organisation ou de la régulation d'une formation, d'un conseil d'orientation, d'un recrutement. L'observation aura une bonne validité empirique si elle apporte une aide significative dans ce genre de situation. Prenons l'exemple de la validité prédictive d'un test en vue d'une opération de sélection professionnelle. Avant d'utiliser le test à des fins de sélection, on doit s'assurer qu'il a une bonne validité prédictive (cette forme de validité est par fois dite critérielle). Pour cela, on commence à définir un critère de validité. Ce peut être par exemple un critère d'efficacité

professionnelle. Dans un premier temps, on caractérisera les sujets uniquement au moyen du test. Dans un second temps, plus ou moins éloigné du premier, on les caractérisera au moyen du critère. Il est alors souhaitable, ce qui n'est pas toujours possible, que tous les sujets ayant passé le test puissent être caractérisés selon le critère. Une forte corrélation entre le test et le

critère indique qu'il aurait été possible de choisir à l'avance, sur la base du test, les sujets qui ont révélé par la suite une bonne efficacité professionnelle. Ce résultat pourra alors être

Utilisé dans les cas semblables pour procéder à une sélection professionnelle.

(Huteau.M. 2013. p45)

C. La validité théorique ou hypothético-déductive:

Elle réside dans l'élaboration théorique de l'outil de mesure, permettant ainsi d'émettre des hypothèses sur le fonctionnement et l'élaboration de ses réponses par le sujet (M. Reuchlin, 1992).F.Longeot(1969) a élaboré une échelle d'intelligence basée sur les théories de l'intelligence de Piaget ; le Thematic Aperception Test, qui relève d'une approche projective, est fondé sur une théorisation psychanalytique du processus d'élaboration des réponses.

Le diagnostic en psychologie suppose l'évaluation des caractéristiques humaines. Cette démarche fait nécessairement appel à la quantification qui si elle est imparfaite, permet néanmoins de définir un cadre descriptif autorisant un meilleur contrôle et une plus grande fiabilité de l'évaluation. Ce cadre descriptif situe l'évaluation à différents niveaux. On peut différencier des catégories (niveau normal), ordonner des niveaux d'intensité sur des grandeurs (niveau ordinal), ou encore mesurer des écarts à une norme (échelle d'intervalle). Le choix du niveau de la mesure renvoie à une question de plausibilité psychologique, c'est –à-dire aux connaissances théoriques qui en supportent l'opérationnalisation.

Ajuster l'utilisation des nombres aux dimensions qu'ils permettent d'appréhender n'est toutefois pas suffisant pour assurer qu'elles sont mesurables. La mesure d'une dimension psychologique nécessite que soient garanties la sensibilité, la fidélité et la validité de l'instrument utilisé .c'est en fonction de la qualité de ces critères que le praticien va accorder plus ou moins de confiances aux outils proposés sur le marché.

(Guillevic et Vautier. 2005. p61)

Donc ce type de validité est utilisé lorsque l'on cherche à savoir si une mesure donnée est associée à d'autres mesures selon des hypothèses théoriques concernant les concepts qui sont mesurés. Cette démarche n'est pas spécifique à la méthode des tests, mais a des méthodes générales de construction et de vérification d'une hypothèse en science expérimentale. Il s'agit d'étudier et de vérifier les liaisons constatées entre les variables et les hypothèses qui ont guidé

les modalités de détermination de la dimension psychologique que l'on veut évaluer.
(Roulin. 2018. p 114)

- **Validité vs fidélité :**

La fidélité concerne la précision avec laquelle un test mesure certaines caractéristiques, elle est donc en relation avec l'erreur de mesure et elle est formellement définie comme le rapport de la variance vraie à la variance totale du test. On définit par contre la validité comme la qualité de ce qui est mesuré ; c'est à dire la ressemblance existant entre ce que l'on veut mesurer et ce que mesure le test. De façon formelle, la validité est donc la portion de variance vraie qui est pertinente aux buts de l'utilisation du test .

- **Remarque:**

On peut déduire que si un test est fidèle, il n'est pas obligatoirement valide et que pour qu'un test soit valide il est nécessaire que celui-ci soit fidèle, c'est à dire que la variance totale ne soit pas que de l'erreur de mesure. S'assurer de la validité d'un test ne donne cependant pas lieu, comme pour la fidélité, à un ou plusieurs indices sur lesquels il existe un consensus. La validation d'un test est une démarche progressive qui commence dès la construction du test (validation de contenu).

* **distinction Validité – Fidélité** : L'absence de fidélité traduit une erreur non constante ou aléatoire autour d'un point moyen (qui peut être la cible ou non) l'absence de validité traduit une erreur constante qui éloigne le résultat de la cible visée. Fidélité et validité sont deux notions distinctes liées par une relation d'implication

* **Relation d'implication [validité \supset fidélité]**: la fidélité est une condition nécessaire mais non suffisante pour la validité d'un test.

(1) Un test non fidèle est nécessairement non valide.

(2) Un test valide est nécessairement, a minima, un peu fidèle.

(3) Un test fidèle n'est pas nécessairement valide. **(Roulin. 2018. p 115)**

C-la sensibilité:

Selon Bernaud (2007), il s'agit des possibilités de discrimination du test, de son pouvoir à différencier les sujets entre eux. **(Bernaud.J.L. 2007. p67)**

La sensibilité d'un test est, commune à toute épreuve construite pour tester un composant de la réalité. La sensibilité d'un test est également la première propriété à vérifier lors de la construction de l'épreuve.

La sensibilité réside dans la capacité d'un instrument de mesure à différencier deux objets avec la finesse requise par l'objectif poursuivi par le psychologue. Cette puissance discriminative de l'outil de mesure peut être obtenue en multipliant le nombre d'items d'un questionnaire ou d'un test et en jouant sur des variables de passation, par exemple le temps accordé pour répondre à un test. Il faut d'autant plus s'interroger sur la signification psychologique de cette manipulation des variables que celle-ci ne conduit pas forcément à une plus grande sensibilité de la mesure : un psychologue peut avoir distingué huit classes dans son échelle mais si l'une d'entre elle regroupe 80% des sujets on comprendra que la sensibilité de l'échelle est faible. **(Guillevic et Vautier. 2005. p 57)**

La sensibilité d'un test représente le pouvoir classant du test, sa capacité à discriminer les éléments testés lorsqu'ils diffèrent et sa capacité à regrouper ces éléments lorsqu'ils convergent ou sont similaires. Un test n'est pas assez sensible lorsqu'il regroupe trop d'éléments testés qui diffèrent visiblement.

Lors de la construction des items d'un test, on doit vérifier sa sensibilité à deux niveaux :

- ✓ la sensibilité de chaque item.
- ✓ la sensibilité du score total.

Un test qui n'est pas sensible ne peut être ni fidèle, ni valide, c'est pourquoi cette propriété est cruciale dans la construction du test, et pourquoi elle doit être étudiée en premier lieu.

Une méthode simple pour éprouver la sensibilité d'un test consiste à le faire passer à un nombre suffisamment important de sujets. A partir des résultats obtenus, on peut construire un histogramme par lequel on va pouvoir vérifier à première vue, la distribution des fréquences. Si le test est sensible, les scores devraient être distribués normalement (courbe en cloche).

++ Trois types de distribution peuvent signaler un problème de sensibilité :

- **Un effet plancher** : Effet observé dans l'utilisation d'une échelle (expérimentations, tests psychométriques...) dont les réponses sont trop « difficiles » pour la population testée. On observera par exemple un effet **plancher**, par un échec massif des sujets à une épreuve, toutes les notes individuelles se regroupant sur le bas de l'échelle (ou sur le haut si l'échelle est inversée) qui perd son pouvoir discriminatif.
- **Un effet plafond** : effet observé dans l'utilisation d'une échelle (expérimentations, tests psychométriques...) dont les réponses sont trop « faciles » pour la population testée. On observera par exemple un **effet plafond**, par une réussite massive des sujets à une épreuve, toutes les notes individuelles se regroupant sur le haut de l'échelle (ou sur le bas si l'échelle est inversée) qui perd son pouvoir discriminatif.

(Admin.2012)

- **Un effet barres** : Un effet barre : Les différents éléments sont regroupés dans un nombre limité de valeurs. On peut donc différencier des groupes d'éléments mais la distinction des sujets au sein du groupe n'est pas possible. (Thomas .M. 2019. p22)

✓ Sensibilité et mesure d'une dimension :

Lorsque l'on cherche à évaluer une dimension (exemples : une aptitude, un trait de personnalité) le test doit permettre de différencier le plus possible les personnes. **La sensibilité est alors le pouvoir séparateur, différenciateur d'un test.** La sensibilité est donc la capacité d'un test à détecter une variation du score vrai sur le trait mesuré. La méthode de sélection des items permet normalement de s'assurer de la sensibilité des tests Pour étudier la sensibilité d'un test, une première méthode consiste à établir la distribution des résultats et d'examiner sa forme via le calcul d'indices de dispersion (écart-type ou autre), d'asymétrie ou d'aplatissement Si l'épreuve est trop facile ou trop difficile, on observe une distribution asymétrique (effet plancher = trop difficile ou effet plafond = trop facile). On préfère en général une distribution plutôt normale, symétrique, au mieux légèrement aplatie qui présente une dispersion et un pouvoir différenciateur plus important. Si la distribution n'est pas une distribution normale, la sélection des questions étaient probablement incorrecte et le choix des questions doit être revue et/ou les questions remaniées. Lorsque l'on

Sélectionne les items on cherche à rendre la courbe « plus normale » d'une part et, d'autre part, à maximiser la dispersion de l'épreuve. Ce remaniement de l'épreuve s'effectue souvent en Augmentant le nombre d'items de difficulté moyenne. (Roulin. 2018. p 90)

--Les défauts de sensibilité :

Plusieurs facteurs peuvent expliquer un défaut de sensibilité :

1-Motifs techniques

Il peut y avoir un défaut dans la construction du test comme par exemple un nombre trop limité d'items, des items mal sélectionnés ou alors un test naturellement peu différenciateur.

Ces différents facteurs peuvent donc influencer sur la sensibilité d'un test.

2-Motifs conjoncturels

Un emploi de l'instrument de mesure dans un contexte particulier peut par exemple, influencer sur la sensibilité du test. Il en est de même si le sujet examiné est inadapté à la tâche.

***-- Que faire pour élever la sensibilité d'un test ?**

D'une manière générale, pour augmenter la sensibilité d'un test, on peut :

- augmenter le nombre d'items ; Il est important de noter que si on augmente le nombre d'items d'un instrument de mesure, il est nécessaire d'augmenter proportionnellement le nombre de sujets examinés.
- affiner les critères de cotation (ce qui peut se faire pour chaque item) ;
- travailler la sensibilité des items. On peut donc, en fonction des résultats obtenus :
- éliminer les items réussis ou échoués par un trop grand nombre de candidats ;
- garder les items réussis par environ la moitié des candidats.
- ajouté éventuellement de nouveaux items. Il faudra cependant, les tester à nouveau pour voir s'ils sont plus discriminants que les précédents. **(Thomas .M. 2019. p22)**

16. La standardisation :

16.1. Qu'est-ce qu'une évaluation standardisée ?

Une situation standardisée C'est une évaluation qui va respecter certaines règles ou conditions, tout au long de la procédure d'évaluation: de la passation de l'épreuve à la cotation des résultats. (chartier.ph. 2018. p 12)

La standardisation des tests est essentielle à respecter. C'est elle qui donne la valeur aux tests et permet de réaliser la comparaison : « **toutes choses étant égales par ailleurs** ». Elle permet donc de s'assurer que les différences entre les scores observés habituellement et les scores observés lors de la passation du test ne sont pas la conséquence de variations de la situation (les événements particuliers à la situation d'évaluation devront donc être pris en compte lors de l'interprétation d'un test). La standardisation doit permettre d'assurer :

- que les conditions de passation sont les mêmes que celles du groupe de référence utilisé

Pour l'étalonnage :

Le manuel des tests est un outil important pour les psychologues. Un bon manuel (donc un manuel qui respecte les règles de la déontologie professionnelle) doit préciser à minima

- Les objectifs généraux du test
- que le matériel et la procédure sont les mêmes ; que la cotation et le calcul des scores sont identiques pour tous afin de garantir le résultat quel que soit l'examineur.

(Roulin. 2018. p11)

16.2. Les conditions de la bonne standardisation :

la passation, comme la cotation, doivent être effectuées selon les indications du manuel du test (indiquons à ce propos qu'un test doit toujours être accompagné d'un manuel qui doit fournir un certain nombre d'informations sur l'épreuve) afin que chaque personne soit placée dans les mêmes conditions et garantir alors la comparabilité des résultats. Pour la passation par exemple, il est fréquemment indiqué qu'elle doit se dérouler dans une pièce calme (sans bruits nuisances...) afin que le-la participant·e soit concentré·e et impliqué·e au maximum. Matériel (la personne peut-elle utiliser un brouillon, un objet...), conditions d'utilisation (temps libre ou limité...) et comportement de l'évaluateur·rice sont aussi soigneusement décrits et ces éléments

doivent être respectés. L'objectif principal de la standardisation de la situation d'évaluation est bien de permettre la comparabilité des résultats, de s'assurer que les différences observées relèvent bien de caractéristiques de la personne et non pas de différences dans la passation ou de la subjectivité de l'évaluateur·rice.

(Chartier.PH. 2018. p 13)

❖ Quand ne pas utiliser un test ?

Voici selon Urbina (2014), les 10 raisons qui devraient conduire un psychologue à ne pas utiliser un test :

- 1- Les objectifs du test sont inconnus ou peu clairs pour le psychologue.
- 2- Le psychologue n'est pas familier avec le test et pas assez entraîné à l'utilisation de ce test.
- 3- Le psychologue ne sait pas à qui est destiné le test ou comment seront utilisés les résultats au test.
- 4- Les informations que pourraient fournir le test sont déjà disponibles ou peuvent être Obtenus par d'autres moyens ou des sources plus sûres.
- 5- La personne devant être testée n'est pas d'accord ou pas prête pour coopérer à une situation De test.
- 6- Le test ou la situation de test peut engendrer un préjudice à la personne testée.
- 7- L'environnement et/ou les conditions de passation ne sont pas adaptés à la situation de test.
- 8- Le format du test n'est pas adapté en raison de l'âge, d'aspect linguistique, culturel, ou de tous les autres facteurs qui rendent invalides les données obtenues.
- 9- Les normes sont trop anciennes ou inadaptées et inapplicables à la personne testée.
- 10- Le manuel du test (documentation) concernant le test ne donne pas d'informations Suffisantes concernant la fidélité et la validité des scores observables. **(Roulin. 2018. p 38)**

17. Etalonnage d'un test :

Une fois le test élaboré et mis à l'essai, le concepteur s'affaire à générer des normes. Ce procédé s'inscrit dans une programmation de la recherche et s'appelle plan de normalisation, ou d'étalonnage. Cette partie importante de l'élaboration d'un test commence une fois que les items ont été sélectionnés, au terme de l'étape d'analyse. Lorsqu'il est prêt pour l'étalonnage, le test ne devrait plus changer jusqu'à sa publication. Les consignes, le nombre d'items, la durée d'administration et les autres caractéristiques devraient être fixés, sans quoi les normes découlant de l'étalonnage pourraient être compromises.

(Hogan, 2017, p.162)

~~Les méthodes d'étalonnage les plus courantes~~

17.1 - Les procédures d'étalonnage :

L'étalonnage est la procédure qui va être utilisée pour permettre qu'un score soit interprétable. L'étalonnage désigne aussi le résultat de cette technique. C'est le psychologue différentialiste qui impose cette procédure, lorsqu'il construit le test : les utilisateurs se serviront par la suite des résultats. L'étalonnage permet d'obtenir des notes standards (les notes obtenues par beaucoup de sujets) que l'on ordonne et classe. Grâce à ce processus on obtient un tableau de correspondance, entre la note au test et la dimension psychologique observée. Ainsi le test passé ensuite par un sujet donne une note que l'on pourra comparer au groupe de référence. L'individu sera donc classé (on saura combien de sujets ont un score inférieur au sien et inversement).

17.2. Phase d'échantillonnage :

L'expérimentateur fait passer le test à un échantillon représentatif de la population. Pour établir un échantillon représentatif, plusieurs méthodes peuvent être prises en voici les exemples les plus courants :

***-Méthode des quotas :**

La méthode des quotas est utilisée quand il n'existe pas de base de sondage. L'objectif est d'assurer la représentativité de l'échantillon en conformant la structure de l'échantillon aux caractéristiques de la population. Cela suppose des statistiques fiables concernant la population parente. Cette méthode est, d'une certaine façon, proche de la méthode stratifiée mais il n'y a plus de hasard. On choisit les personnes « au gré des rencontres » (et on est libre de choisir telle ou telle personne) mais on s'impose de respecter les proportions de diverses catégories de la population parente. On cherche les sujets représentatifs de la population totale. Par exemple, si dans la population de référence il y a 10% d'agriculteurs et 51% de femmes et 49% d'hommes parmi eux, pour un échantillon de 100 personnes, on devra sélectionner 5 agriculteurs et 5 agricultrices

En dehors de la méthode des quotas, il existe d'autres méthodes non probabilistes mais ces Méthodes peuvent induire des biais plus ou moins importants dans la représentativité de L'échantillon.

-Échantillonnage sur la base du volontariat (pas de garantie de représentativité) mais utilisé non pas dans la construction des tests mais dans les études lorsque l'on recherche des groupes Témoins.

- La méthode des itinéraires : on fixe à celui qui interview un itinéraire (du point de vue de l'espace et du temps) à parcourir pour interroger des personnes de façon à avoir la population la plus représentative possible (méthode utilisée en sociologie dans certaines enquêtes).

- Technique boule de neige ("snowball") : on utilise le parrainage ou les amis et collègues des répondants pour construire l'échantillon (permet en fait de sonder ou d'interroger une population spécifique).

17.2. Construction d'un étalonnage :

Il existe différents étalonnage et les règles de construction sont différentes selon le type

D'étalonnage. De façon générale, pour construire un étalonnage il faut choisir un type d'étalonnage (quantilage, échelle normalisée, échelle réduite, etc.) et éventuellement le nombre de catégorie de référence que l'on va utiliser. Le choix du nombre de catégorie d'étalonnage

(pour des étalonnages de type quantilage, échelle réduite, échelle normalisée) est fonction du but poursuivi et dépend du besoin que l'on a de discriminer finement ou non les sujets. Il dépend aussi de la nature des données, de la taille de l'échantillon normatif, du nombre de notes brutes observées dans cet échantillon et de la sensibilité de l'épreuve. En règle générale on considère que :

Le nombre d'observations indépendantes (personnes) doit être au moins 10 fois plus grand que le nombre des catégories de l'étalonnage. Pour un étalonnage en 10 catégories, il faut au moins 100 sujets.

Le nombre des notes brutes observées de l'épreuve doit être 3 à 4 fois supérieur à celui des Catégories de l'étalonnage (pour un étalonnage en 10 classes, il faut que 30 à 40 notes Brutes).

Plus le nombre des catégories est élevé, plus la fidélité du test doit être grande : il ne servirait à rien de discriminer finement des sujets dont les performances ne sont pas stables Par exemple. **(Roulin. 2018. p 137)**

- Les différentes formes de quantilage :

Quantilage, décilage, centilage, constituent des techniques d'étalonnage. Elles fournissent le pourcentage d'individus ayant obtenu un résultat supérieur à celui de l'individu examiné. Au lieu de classes contenant toutes le même pourcentage d'éléments comme celles des quantilages, on peut utiliser des classes contenant des pourcentages d'éléments correspondant à ceux que déterminent, sur une distribution normale, des intervalles égaux entre les mesures. On peut également indiquer à l'utilisateur de l'épreuve à quel écart réduit correspond, dans la population de référence, chacune des notes pouvant être obtenues. On parle alors d'échelle en écarts réduits. **(Jean.N.2010)**

- Exemple de quantilage :

- **Décilage:** division de l'effectif de la population d'étalonnage en 10 classes égales comportant chacune 10% de l'effectif (si la population d'étalonnage comporte 100 sujets, chaque décile comprendra 10 sujets).
- **Centilage:** division de l'effectif de la population d'étalonnage en 100 classes égales comportant chacune 1% de l'effectif.

Chaque décile est défini par une **note minimale** et une **note maximale** (les limites de la classe). En raison de la forme de la distribution, **l'intervalle de classe** (écart entre note minimale et maximale) **n'est pas constant**: le décilage ne discrimine pas de façon égale les sujets ayant des notes faibles ou fortes et ceux qui obtiennent des notes moyennes.

(carnets2psycho.net/ La définition de Étalonnage)

Conclusion

La psychométrie est considérée comme étant une matière essentielle quelles que soit la spécialité du psychologue et son orientation professionnelle, chercheur ou praticien. Cette discipline a contribué au développement de la psychologie grâce aux différents et plusieurs domaines tel que le coefficient de corrélation, l'eugénisme, la transmission héréditaire, analyse factorielle et les tests d'intelligence...

A partir de cela, les chercheurs ont pu utiliser la psychométrie dans les différents domaines de la vie et compris la vie psychologique grâce aux tests dans le but est résoudre les problèmes psychique, éducatifs et sociaux.

Les références bibliographiques

Ouvrages

- 1-Bernaud. J.L. (2007). **Introduction à la psychométrie**. paris.dunod.
- 2-Chartier. Ph. (2018). **L'évaluation en psychologie, test et questions**. Paris. édition in press.
- 3-Denise .R et Guyot.S, (2016), **la psychométrie, évaluation et pronostic par les tests de psychométrie**, l'harmattan.
- 4-Hogan .TH, (2017), **introduction à la psychométrie**, chenellière éducation, 2^{ème} édition.
- 5-Huteau.M. (2013). **psychologie différentielle**, paris. Dunod, 4^{ème} édition.
- 6-Laveault.D et Grégoire.J.(2014), **introduction aux théories des tests en psychologie et en sciences de l'éducation**, Noto, 3^{ème} édition.
- 7-Rolland.J.P. (2006). **comment évaluer un test**. Paris (PDF)
- 8-Roulin.J.L. (2018). **savoir, comprendre, apprendre. Leçons de psychométrie**, 3^{ème} édition.
- 9-Vial. M. (2012).**Se repérer dans les modèles de l'évaluation**. Édition de Boeck supérieur.

Dictionnaires

- 10-Larousse(2008). dictionnaire de français, France.
- 11-Larousse.fr/dictionnaires /français/mesurer/50793.
- 12-www.larousse.fr/encyclopedie/divers/binet-simon/183418

Articles :

- 13-Admin.2012. **Psychologie effet plafond**.retrouvé dans le site : psychologie.savoir.fr-
Consulté le 23-11-2020.
- 14-Aubert-Marson.(2009).**sir francis galton :le fondateur de l'eugénisme**. Retrouvé dans le site :<http://www.medecinesciences.org>.
Consulté le 17aout 2020.
- 15-Baromètre Ipsos/Edenred.2012.**le livre blanc de l'évaluation psychométrique**. Retrouvé dans le site <http://www.centraltest.fr>.
Consulté le 17 aout.

16- Berthieume.François. **fechner gustav theodor**. Retrouvé dans le site : <http://www.agora.qc.ca>.

Consulté le 23 novembre 2020.

17- Demeuse.M.(2004),**introduction aux théories et aux méthodes de la mesure en sciences psychologiques et en sciences de l'éducation**. Université de Mons. Retrouvé dans le site : www.resarchgate.net.consulté le 18-9-2020 à 16h25.

18- Grégoire.J , « **psychométrie** », Encyclopædia Universalis [en ligne] :

<https://www.universalis.fr/encyclopedie/psychometrie>

Consulté le 23 novembre 2020

19- Grégoire.J, « **Wechsler david-(1896-1981)** »,encyclopaedia universalis (en ligne),retrouvé dans le site :<http://www.universalis.fr>.

Consulté le 18 aout 2020.

20- **Histoire du monde de la psychométrie**.2013. Retrouvé dans relevez vos talent.fr. Consulté le 11 aout 2020.

21-[Http://www.babilio.com/auteur/-gustav-theodor-fichner/335564](http://www.babilio.com/auteur/-gustav-theodor-fichner/335564).

Consulté le 15 aout 2020.

22- [Http://www.fr.wikipedia.org/psychométrie](http://www.fr.wikipedia.org/psychométrie). Consulté le 8 aout 2020.

23- [Http://psychometrie.espaceweb.usherbrooke.ca/instrument-psychometrique](http://psychometrie.espaceweb.usherbrooke.ca/instrument-psychometrique).

Consulté le 8 aout 2020

24- [Http://www.bibmath.net/bios/index.php?action=affiche&quoi=galton](http://www.bibmath.net/bios/index.php?action=affiche&quoi=galton).

Consulté le 17aout 2020.

25- [Http://www.bibmath.net/bios/index.php?action=affiche&quoi=pearson](http://www.bibmath.net/bios/index.php?action=affiche&quoi=pearson).

Consulté le 17aout 2020.

26-[Https://carnets2psycho.net/dico/sens-de-etalonage.html](https://carnets2psycho.net/dico/sens-de-etalonage.html)- La définition de Étalonnage

27- [Http://courspsycho.blog4ever.com/psychometrie-1](http://courspsycho.blog4ever.com/psychometrie-1).

Consulté le 8 aout 2020.

28- [Http://overthe130.wordpress.com/la-difference-entre-wechsler-et-cattell](http://overthe130.wordpress.com/la-difference-entre-wechsler-et-cattell).

Consulté 18 aout 2020.

29- [Http://Scientific european federation of osteopath](http://Scientific%20european%20federation%20of%20osteopath). **Les erreurs** .retrouvé dans le site: www.scientific-european-federation-osteopath.org.

Consulté le 20 aout 2020.

30- Jean.N.2010-**psychométrie- introduction-a-la-mesure**.retrouvé dans le site <http://reflexions-psycho.over-blog.com>.consulté le 23 novembre 2020.

31- Marchand. G. (2005).**naissance de la psychologie expérimentale**-article issu du numéro enquêtes sur la lecture, au-delà des idées reçues(mensuel N°161-juin2005).

32- Michel.Caire.(2012).**théodor simon**. Retrouvé dans le site :<http://psychiatrie.histoire.free.fr>.

Consulté le 18 aout 2020.

33- Noci .Stéphanie.**les tests psychologiques**. Retrouvé dans :interview "5(pdf).consulté le 8 aout 2020.

Consulté le 8 aout 2020.

34- **Psychométrie et examens psychologiques**. Retrouvé dans : les cours psycho.blog 4 ever.com.

Consulté le 8 aout 2020.

35- **Psychométrie : introduction à la mesure**(2010).retrouvé dans <http://www.reflexions-psycho-over.blog.com>. 24 novembre 2020.

35- Psyfact.(2016).**alfred binet(1857-1911) ; histoire et grands noms de la psychologie**. Retrouvé dans le site :<http://psyfact.e-monsite.com>.

37- Psyfact.2016.**charles edward spearman(1863-1945) dans histoire et grands nom de la psychologie** .retrouvé dans le site :e-monsite.com.

Consulté le 18 aout 2020.

38- Richard. J-F , « **Wundt wilhem- (1832-1920)** », Encyclopædia Universalis [en ligne], URL : <https://www.universalis.fr/encyclopedie/wilhelm-wundt/> consulté le 23 novembre 2020.

39- Thinés.G.**cattell James Mckeen-(1860-1944)**.retrouvé dans le site:www.universalis.fr. Consulté le 18 aout 2020 à 13h26.

40- Vermeulen.J.**la psychologie expérimentale**. Retrouvé dans le site :<http://www.lepsychologue.be>.

Consulté le 16 aout 2020.

41- Zarrouk.F.2012.**le coefficient de corrélation**. issep ksar-said.PDF

Consulté le 17aout 2020.