

FICHE PEDAGOGIQUE

L'étalonnage en psychométrie

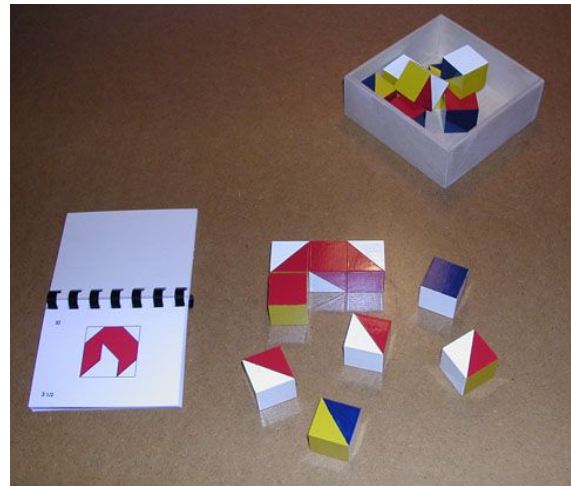
V1.3.00

29/12/2023

Si vous avez téléchargé cette fiche ailleurs que sur le site, assurez-vous d'avoir la dernière version ici :

<http://deporientation.free.fr/Ressources/Ressources.html>

Fabien BELTRAME, Ph.D.



1 Pourquoi un étalonnage ?

Imaginons un test d'aptitude cognitive qui comporte 48 items. On pourrait être tenté d'adopter un réflexe « scolaire » et dire que 24 bonnes réponses correspond à la moyenne. Donc toutes les notes au-dessus de 24 sont considérées comme « fortes » et toutes celles situées en dessous sont considérées comme « faibles ».

Mais imaginons également que nous fassions passer le test, les 48 items, à une population de 500 personnes. Le résultat est le suivant : 499 personnes obtiennent 0 bonne réponse, et une personne obtient 5 bonnes réponses. Dans cet exemple un peu caricatural, on comprend que le constructeur du test a conçu 48 items d'une extrême complexité. Réussir à trouver la bonne réponse à cinq items apparaît donc comme une performance extrêmement supérieure. Et 5 bonnes réponses, c'est très loin de la « pseudo moyenne » de 24 !

S'il n'y avait pas d'étalonnage on pourrait penser que cinq bonnes réponses sur 48 est un score extrêmement faible. Mais on se rend compte que dans cette population de 500 personnes une seule a réussi à répondre juste à cinq items. Cette personne a donc réussi une performance largement supérieure aux 499 autres.

L'objet de l'étalonnage c'est donc d'offrir des repères en termes de comparaison à une population de référence, c'est à dire une population dont les individus ressemblent à la personne qui vient de passer le test

Utiliser un étalonnage va permettre de classer les individus les uns par rapport aux autres en fonction de la performance qu'ils ont atteint.

« Dans le cadre d'une évaluation normée, tester des sujets consiste toujours à les comparer, à les distinguer entre eux. Sans référence au résultat d'autres sujets, les notes brutes d'un individu à un test donné sont sans signification précise. En effet, d'un test à l'autre, la nature et la difficulté des items varie. Sur la base d'un score brut, nous ne pouvons donc déterminer si un sujet est faible ou brillant. Pour pouvoir interpréter les résultats, il est nécessaire de faire correspondre les notes brutes à celle d'une échelle de référence qui possède une valeur normative. »

Laveault, Grégoire, 2014

Le quantilage prend un nom singulier pour certaines valeurs :

- Quartilage : 4 classes de 25% des effectifs totaux
- Quintilage : 5 classes de 20% des effectifs totaux
- Décilage : 10 classes de 10% des effectifs totaux (exemple ci-dessus)
- Centilage : 100 classes de 1% des effectifs totaux

2.1.1 Lecture de la table

Si la personne que j'évalue obtient un score de 38, je peux dire qu'elle se situe en quantile Q10 (ou au 10ième décile). Et donc qu'elle obtient un score dont seul 10% de la population est capable. Ou encore que 90% de la population obtient un score plus faible que le sien.

2.2 Les échelles réduites :

Lorsque la distribution des scores s'approche d'une loi normale, les quantilages ne conviennent plus. Pour vérifier si la distribution respecte les propriétés de la distribution normale, on utilisera le test statistique de **Kolmogorov-Smirnov** . Ici, les limites des catégories sont fondées sur les écarts à la moyenne.

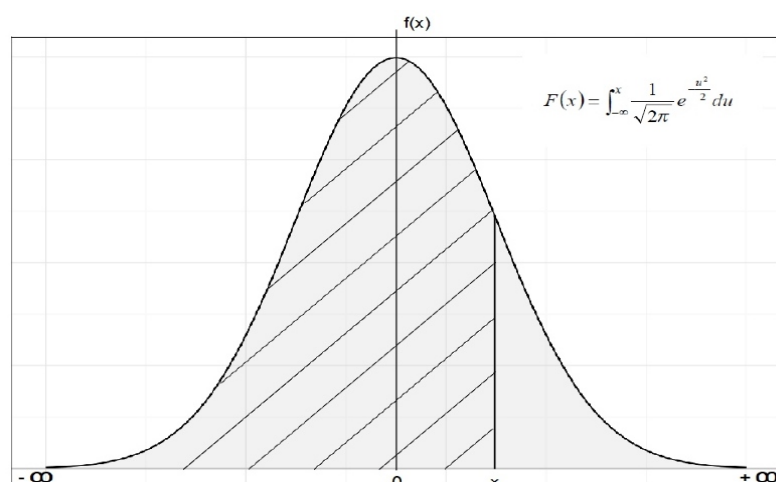
Le nombre de classes est habituellement impair (5, 7, 9, 11) et la classe centrale est centrée sur la moyenne. La « taille » ou l'**étendue d'une classe** est fonction du nombre de classe et est fixée par convention. Par exemple, pour un étalonnage en 7 classes, l'étendue est de 0,667 x l'écart-type de la distribution.

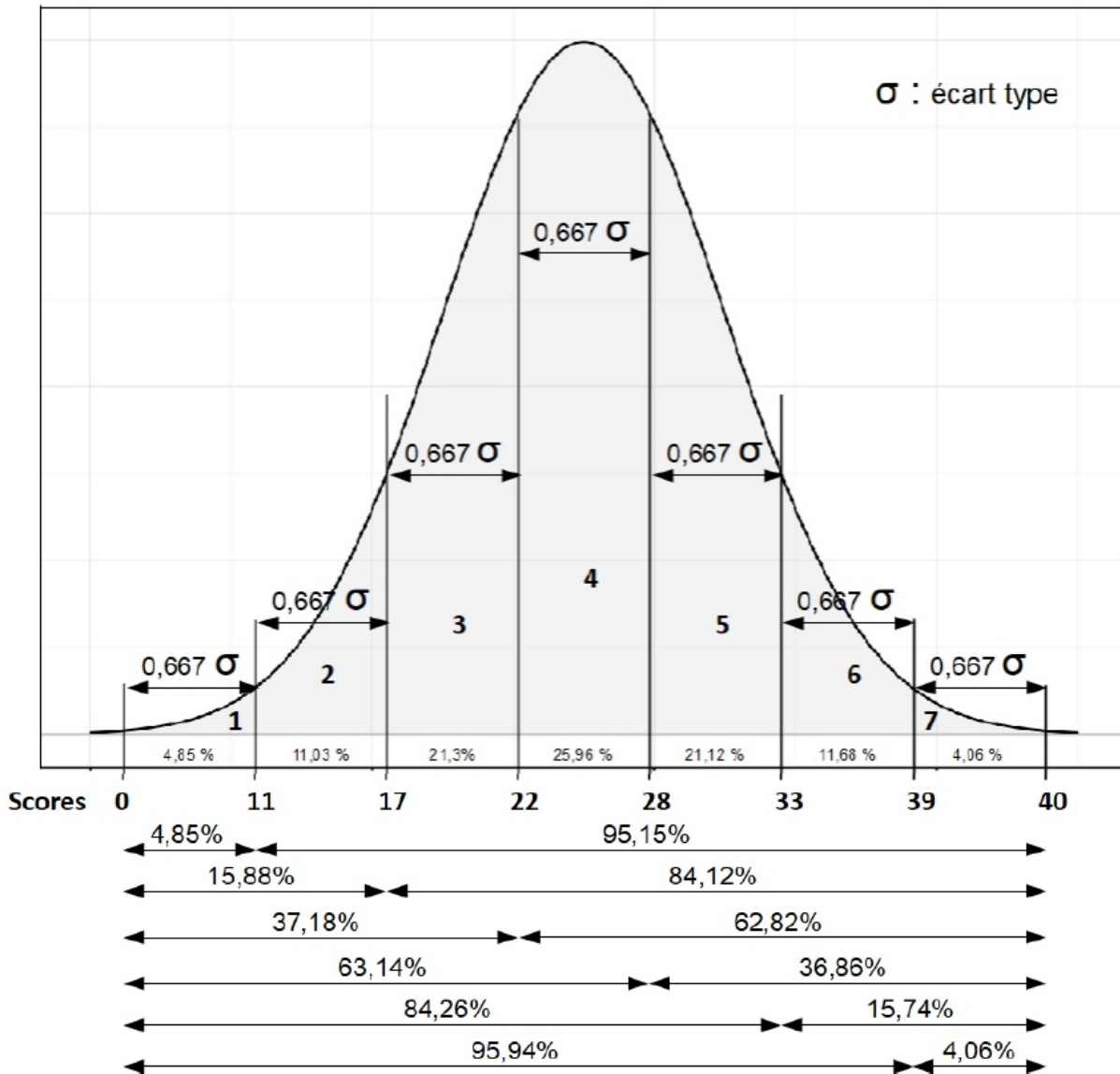
Exemple : avec la même distribution des scores obtenus par 200 personnes sur une échelle qui s'étend de 0 à 40. Le test de Kolmogorov-Smirnov indique un écart maximum de 0,05344 et une valeur critique de KS de 0,0960. H0 n'est donc pas rejetée et on peut considérer que la distribution des scores suit une loi normale.

- Propriété de la distribution : $m=25,185$ & Ecart-type = 8,119
- Etendue de classe (pour un étalonnage en 7 classes) : $0,667 \times 8,119 = 5,415$
- Pour mémoire, la surface sous la courbe représente 100% de la population.
- Calcul des bornes de la classe centrale (4) :
 - Borne inférieure : moyenne moins $(0,667 \times 8,119) / 2$
 - Borne supérieure : moyenne plus $(0,667 \times 8,119) / 2$
 - Toutes les autres bornes sont calculées avec un décalage de 5,415, c'est à dire $0,667 \times 8,119$

Notons que pour cette distribution, la valeur de symétrie est de -0,5199. C'est à dire que la distribution des scores est décalée vers la droite, donc vers les scores les plus élevés. En effet, la valeur de la moyenne est de 25,185 et non 20 qui correspondrait au score le plus élevé divisé par 2 ($40/2$). C'est la raison pour laquelle on trouve des scores de 0 à 11 en classe 1 (étendue de 11 points) alors que l'on ne trouve que des scores de 39 à 40 en classe 7 (étendue de 1 point).

Les pourcentages de population sont calculés à partir des valeurs exactes des bornes de classe et de la densité de probabilité de trouver une valeur inférieure, ou supérieure, par rapport à la loi normale centrée réduite.





2.2.1 Lecture de la table

Si la personne que j'évalue obtient un score de 27, je peux dire qu'elle se situe en classe 4. Il y a donc 25,96% de la population de référence qui obtient le même score. 37,18% de la population obtient un score inférieur au sien et 36,86% de la population obtient un score supérieur au sien. Elle fait partie des 62,82 % des personnes qui obtiennent un score supérieur à 22

Notons que ce type d'échelle n'est pas beaucoup utilisée car on lui préfère souvent les échelles normalisées.

2.3 Les échelles normalisées

Bien souvent la distribution des scores bruts est proche de la loi normale, mais pas suffisamment pour que l'on puisse fabriquer une échelle réduite. Il convient alors de construire une échelle normalisée. Dans le cas de scores bruts à un test psychologique, c'est bien souvent à ce troisième cas que l'utilisateur du test va être confronté.

Pour construire la table d'étalonnage en échelle normalisée, on va se référer à une distribution normale théorique. A partir des scores classés par ordre croissant, on va définir les classes en fonction du pourcentage de population qu'elle doit contenir.

3 Remarques générales

	Type de distribution	Avantage	Inconvénient
Quantilage	Toutes formes	Fonctionne avec tout type de distribution	Problème de finesse discriminative lorsque la distribution s'approche d'une loi normale
Echelle réduite	Distribution normale	Parfaite pour les distributions normales et échelles d'intervalles	Parfaite pour une distribution « presque normale »
Echelle normalisée	Distribution normale et quasi normale	La distribution doit être conforme à la distribution normale	« Transforme » la réalité de la distribution

Tableau récapitulatif des différents types d'étalonnage et leurs caractéristiques

Pour fabriquer un étalonnage, il convient de faire attention à ces 3 points :

- ✓ Le nombre de sujets doit être au moins 10 fois plus grand que le nombre de classe. Et à minima, un nombre de 300 sujets est indispensable pour être statistiquement significatif
- ✓ Le nombre de scores possibles doit être 3 à 4 fois plus grand que le nombre de classe.
- ✓ Plus le nombre de classe est grand et plus la discrimination est forte ; donc, plus la fidélité du test doit être grande.

4 Réponses aux questions fréquentes

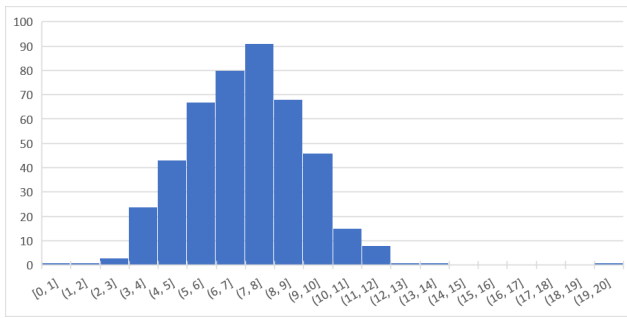
Foire aux questions fréquentes de mes étudiants, alimentée au fur et à mesure des promotions devant lesquelles j'interviens

4.1 Question N°1 : « A quoi sert un étalonnage ? »

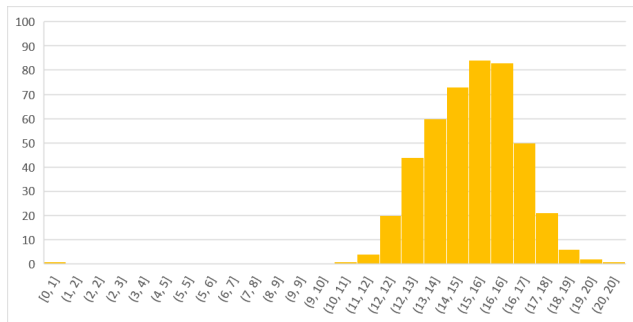
Voici les résultats obtenus par 2 groupes de 450 étudiants d'université. Ils ont suivi le même cours sur la psychologie générale (même contenu, même durée). A la fin du semestre, ils sont appelés à répondre au même examen, le même jour et chacun dans son amphithéâtre respectif. La seule différence concerne l'enseignant qui assure le cours :

- **Professeur A** : très jeune, c'est sa première année d'enseignement. Il vient d'obtenir son diplôme. Il est autoritaire, directif. Sa méthode pédagogique est essentiellement descendante à partir d'un PPT. Il n'y a pas d'interaction, les étudiants ne sont pas autorisés à poser des questions. Les étudiants ne peuvent pas non plus le solliciter en dehors des heures de cours
- **Professeur B** : senior, c'est sa 32ième année d'enseignement. Il a beaucoup d'expérience. Il est à l'écoute de ses étudiants, il y a beaucoup d'interactions avec les étudiants qui lui posent des questions. Ses méthodes pédagogiques sont régulièrement actualisées pour mettre en œuvre les approches les plus récentes de la recherche en science de l'éducation. Et enfin, ses étudiants peuvent le solliciter par visio après les cours pour lui poser des questions, éclaircir un point mal compris ou même envisager des prolongements non abordés en amphithéâtre

Distributions des notes



Professeur A



Professeur B

Tables d'étalonnages des 2 distributions des notes des étudiants

Table d'étalonnage					
Classe	I	II	III	IV	V
Score	0-3	4-5	6-7	8-9	10-20
%pop.	6,7	24,2	38,2	24,2	6,7
%SUP	93,3	69,1	30,9	6,7	0
%INF	-	6,7	30,9	69,1	93,3

Professeur A

Table d'étalonnage					
Classe	I	II	III	IV	V
Score	0-12	13-13	14-15	16-16	17-20
%pop.	6,7	24,2	38,2	24,2	6,7
%SUP	93,3	69,1	30,9	6,7	0
%INF	-	6,7	30,9	69,1	93,3

Professeur B

L'étudiant du **professeur A** qui obtient la **note 12** est situé en classe V. Soit la plus haute classe de l'étalonnage. Il fait partie des 6,7 % d'étudiants qui ont la note **la plus haute** parmi les étudiants du professeur A. Et 93,3 % des étudiants de sa promo ont une note inférieure à la sienne. Il peut être considéré comme un **très bon** étudiant.

L'étudiant du **professeur B** qui obtient la **note 12** est situé en classe I. Soit la plus basse classe de l'étalonnage. Il fait partie des 6,7 % d'étudiants qui ont la note **la plus basse** parmi les étudiants du professeur B. Et 93,3 % des étudiants de sa promo ont une note supérieure à la sienne. Il peut être considéré comme un **très mauvais** étudiant.

Cet exemple est volontairement caricatural pour la démonstration. Dans la vraie vie, il n'existe pas de « bon » et « mauvais » professeurs. Tous les enseignants universitaires sont excellents. De plus, dans cette situation, le lecteur peut se passer d'étalonnage car il existe des conventions très largement partagées : les notes sont souvent sur une échelle de 0 à 20, la moyenne est située à 10, une note inférieure à 10 est considérée comme mauvaise et supérieure à 10 comme bonne.

Mais cela devient plus compliquée lorsque l'on intéresse aux dimensions psychologiques. En effet qu'en est-il de l'**introversion** par exemple ? Dans la mesure où l'introversion peut être évaluée par différents tests de personnalité (Neo-PI-3, BFI, PAPI, SOSIE, GOLDEN, MBTI, etc ...) et que chaque test a sa propre échelle de scores, il devient impossible d'interpréter un score d'une personne. Et deuxième difficulté, que signifie un score de 28 en introversion sur une échelle de scores de 0 à 40 ?

L'étalonnage permet donc de répondre à ces questions en « situant » le score de la personne parmi les personnes qui lui ressemblent ... comme les étudiants des professeurs A et B ci-dessus. Si la personne obtient un score brut de 28, je ne sais pas interpréter. Mais avec la table d'étalonnage :

- si le score de 28 est en classe V alors on peut dire que la personne apparaît beaucoup plus introvertie que l'ensemble de la population qui lui ressemble
- si le score de 28 est en classe I alors on peut dire que la personne apparaît beaucoup moins introvertie que l'ensemble de la population qui lui ressemble

4.2 Question N°2 : « Est-ce à moi, utilisateur, de calculer l'étalonnage ? »

Cette fiche pédagogique explique que l'étalonnage est calculé à partir d'une distribution de scores de personnes qui ont passé le test. Pour calculer un étalonnage, il faut à minima 300 passations. Ce serait très difficile pour un utilisateur de test de récolter 300 passations avant de pouvoir utiliser le test.

L'étalonnage (voire les étalonnages) est donc produit par l'éditeur du test. Et bien souvent, l'étalonnage est calculé à partir des données qui servent à la validation psychométrique du test.

4.3 Question N°3 : « Où est-ce que je trouve l'étalonnage du test ? »

Un test est toujours livré avec le manuel technique qui présente à minima les informations suivantes :

- une présentation de la (ou des différentes) dimension évaluée par le test
- une explication sur la méthode de construction du test
- la présentation des analyses psychométriques de validation du test
- **Le ou les étalonnages(s)**

Mais reconnaissons qu'aujourd'hui les tests papier/crayon ont quasiment tous disparus. Le temps où l'on commandait son test chez un éditeur qui livrait un paquet de cahiers de passation, un paquet de feuilles de réponses, une grille de correction et un manuel technique est révolu. Aujourd'hui, on achète un nombre de passations sur une plateforme en ligne. Et force est de constater que dans le maelström de la pression du quotidien, on ne prend plus le temps de s'intéresser au manuel technique. D'autant que ... la plateforme informatisée donne directement la note étalonnée sans devoir faire cet exercice de comparaison de la note brute avec la table d'étalonnage ... et c'est d'autant plus regrettable pour le développement et le maintien des compétences des professionnels qui évaluent

4.4 Question N°4 : « S'il y a plusieurs étalonnages dans le manuel technique, lequel dois-je utiliser ? »

Les étalonnages sont toujours accompagnés des caractéristiques de la population qui a été utilisée pour le calculer. Vous devez donc prendre la table d'étalonnage dont la population correspond à la personne que vous accompagnez.

Par exemple, pour la batterie multifactorielle NV7, il ne faudrait pas utiliser l'étalonnage de la population de collégiens du manuel technique pour trouver la note étalonnée d'un adulte de 42 ans en reconversion.

Autre exemple, le modèle RIASEC des intérêts professionnels est assez sensible aux différences de genre. Dans les analyses psychométriques de validation on constate des différences statistiquement significatives de scores entre les hommes et les femmes ou entre les garçons et les filles pour un public jeune. Donc si vous accompagnez une étudiante vous comparerez son score à un étalonnage réalisé sur une population composée de femmes et jeunes.

4.5 Question N°5 : « Quand je fais une évaluation des intérêts professionnels, est-ce que je dois utiliser les scores bruts ou bien les scores étalonnés ? »

Quand on accompagne une personne sur la découverte de son profil d'intérêts professionnels, on peut se retrouver devant deux situations : soit les scores bruts et étalonnés sont congruents, soit ils ne le sont pas.

Attention, le fait qu'ils ne soient pas congruents ne doit pas être interprété négativement. Il faut uniquement considérer cette non-congruence comme une information complémentaire.

- Exemple de scores bruts et étalonnés **congruents** :

Dimensions	Réaliste	Investigateur	Artiste	Social	Entreprenant	Conventionnel
Scores bruts	2	10	8	14	5	9
Scores étalonnés	4	5	4	5	4	5

Ici le score brut le plus élevé est "14" pour social. Il correspond également au score étalonné le plus élevé "classe 5". Les deux autres scores bruts les plus élevés, respectivement "10" et "9" correspondent également aux scores étalonnés les plus élevés "classe 5". Les scores bruts plus faibles "2", "8" et "5" correspondent aux scores étalonnés plus faibles également "classe 4".

- Exemple de scores bruts et étalonnés **non-congruents** :

Dimensions	Réaliste	Investigateur	Artiste	Social	Entreprenant	Conventionnel
Scores bruts	9	12	17	2	3	0
Scores étalonnés	7	5	6	1	4	1

Ici le score brut le plus élevé est "17". Mais il ne correspond pas au score étalonné le plus élevé puisqu'il est "classe 6". Le score brut "9" vient en 3ème position après le 17 et le 12. Par contre, ce 9 brut correspond lui, au score étalonné le plus élevé "classe 7".

- Analyse

Dans le cas des scores **congruents** ci-dessus, on peut choisir une approche idiosyncrasique. C'est à dire en analysant les scores bruts entre eux sans comparaison avec les scores étalonnés sur la population de référence. Dans ce cas, on peut partager avec la personne le profil "Social – Investigateur" pour confirmer que la personne se retrouve bien dans les descriptifs des deux lettres S et I.

Dans le cas des scores **non-congruents** ci-dessus, il convient d'ajouter dans l'analyse des scores, l'approche nomothétique de comparaison à la population de référence. En effet, si l'on s'en arrête à l'approche idiosyncrasique alors le profil serait "Artiste – Investigateur", respectivement scores bruts 17 et 12. Alors que si l'on tient compte des scores étalonnés, le profil devient "Réaliste – Artiste", respectivement classe 7 et 6.

Les scores ci-dessus correspondent à un profil féminin. Sur le modèle RIASEC, les femmes ont habituellement un score brut en REALISTE beaucoup plus faible que les hommes. Ainsi, un score brut de 9 en REALISTE pour une femme correspond à un score très élevé comparativement à la population de référence féminine et donc correspond à la classe la plus élevée 7. On observe également sur le modèle RIASEC que les femmes ont habituellement des scores très élevés sur la dimension ARTISTE. Donc malgré un score de 17, le score étalonné ARTISTE est une classe en dessous du score étalonné REALISTE. Dans ce cas de scores non-congruents, il convient donc de tenir compte des deux types de scores, bruts et étalonnés, pour accompagner la personne dans la découverte des 3 lettres A, I et R et l'aider à identifier le profil dans lequel elle se retrouve le mieux.

4.6 Question N°6 : « Je n'ai toujours pas compris à quoi sert l'étalonnage !! »

Alors essayons cette métaphore : imaginez une classe de 25 élèves de seconde qui répondent à une interrogation de mathématiques. Pierre obtient une note de 11/20. Que peut-on dire de cette note ? C'est une note moyenne car elle est juste au-dessus de la moyenne ? C'est une bonne note car elle permet de décrocher son bac ? C'est une mauvaise note car Pierre ambitionne de candidater en 1^{er} année de Psychologie à l'université et il faut un très bon niveau en math pour être admis en filière psychologie ?

En l'état nous ne pouvons rien dire sur cette note. Nous ne pouvons pas la qualifier. Alors je vous donne une information supplémentaire : les 24 autres élèves de la classe de Pierre ont tous obtenu 20/20. Cette fois, nous pouvons dire que la note de Pierre n'est pas bonne. En effet, il est le seul à n'avoir pas obtenu le 20/20 et en plus, il est même bien en deçà.

Imaginons l'autre cas de figure : les 24 autres élèves de la classe de Pierre ont tous obtenu 0/20. Cette fois, Pierre a une note extrêmement bonne car non seulement, il est au-dessus de la moyenne de 10/20 mais en plus il est le seul à avoir décroché des points à cette interrogation de math.

L'étalonnage sert à cela : situer un score comparativement à une population de référence. Pour Pierre, la population de référence ce sont les 24 autres élèves qui ont fait le même exercice (l'interrogation de math) dans les mêmes conditions. Quand une personne répond à un test psychométrique, la population de référence ce sont toutes les autres personnes qui ont répondu au même test, dans les mêmes conditions et qui lui ressemblent : même genre, même qualification, sensiblement le même âge, etc ...

Dans l'exemple de Pierre et si on utilise un étalonnage en 7 classes alors :

- dans le 1^{er} cas : il obtient 11/20 et tous les autres 20/20, Pierre serait en score étalonné 1 (classe I) et les 24 autres en classe VII
- dans le 2^d cas : il obtient 11/20 et tous les autres 0/20, Pierre serait en score étalonné 7 (classe VII) et les 24 autres en classe I