

La courbe normale en psychométrie et en éducatrice : postulat, constante naturelle ou erreur scientifique?

Michel Trahan et Clément Dassa

Volume 4, numéro 1, hiver 1978

URI : <https://id.erudit.org/iderudit/900066ar>

DOI : <https://doi.org/10.7202/900066ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Revue des sciences de l'éducation

ISSN

0318-479X (imprimé)

1705-0065 (numérique)

[Découvrir la revue](#)

Citer cet article

Trahan, M. & Dassa, C. (1978). La courbe normale en psychométrie et en éducatrice : postulat, constante naturelle ou erreur scientifique? *Revue des sciences de l'éducation*, 4(1), 73–80. <https://doi.org/10.7202/900066ar>

Résumé de l'article

Après avoir établi que la méthode classique de construction des tests, basée sur les principes psychométriques, postule la distribution normale des traits mesurés, la possibilité de démontrer que la normalité des traits psychologiques constitue une constante naturelle est écartée. Ensuite, après avoir contesté l'utilisation du postulat de normalité pour certaines formes d'évaluation en éducation, des distinctions sont établies entre principes « éducatrices » et principes psychométriques, entre tests critériés et tests normatifs, entre différences inter-individuelles et différences intra-individuelles. Ces distinctions permettent de voir qu'un instrument élaboré selon les principes psychométriques classiques, postulant la normalité, servira à amplifier les différences inter-individuelles et ne pourra pas être qualifié d'instrument basé sur une erreur scientifique. Toutefois, il est reconnu que certaines utilisations des principes psychométriques en éducation sont inacceptables et que les principes éducatrices sont généralement mieux adaptés aux besoins de l'enseignement et de l'apprentissage.

La courbe normale en psychométrie et en édu-métrie : postulat, constante naturelle ou erreur scientifique?

Michel Trahan et Clément Dassa *

RÉSUMÉ

Après avoir établi que la méthode classique de construction des tests, basée sur les principes psychométriques, postule la distribution normale des traits mesurés, la possibilité de démontrer que la normalité des traits psychologiques constitue une constante naturelle est écartée. Ensuite, après avoir contesté l'utilisation du postulat de normalité pour certaines formes d'évaluation en éducation, des distinctions sont établies entre principes « édu-métriques » et principes psychométriques, entre tests critériés et tests normatifs, entre différences inter-individuelles et différences intra-individuelles. Ces distinctions permettent de voir qu'un instrument élaboré selon les principes psychométriques classiques, postulant la normalité, servira à amplifier les différences inter-individuelles et ne pourra pas être qualifié d'instrument basé sur une erreur scientifique. Toutefois, il est reconnu que certaines utilisations des principes psychométriques en éducation sont inacceptables et que les principes édu-métriques sont généralement mieux adaptés aux besoins de l'enseignement et de l'apprentissage.

Les problèmes de mesure et d'évaluation suscitent régulièrement de nombreuses discussions sur l'utilisation des tests en éducation et, plus particulièrement, sur la pertinence de la courbe normale en éducation. Certains prétendent que la courbe en cloche constitue une erreur scientifique et associent nettement cette dernière à une idéologie

* Trahan, Michel et Dassa, Clément : professeurs, Université de Montréal.

bourgeoise. D'autres, assez nombreux parmi les praticiens et les spécialistes de la mesure et de l'évaluation en psychologie et en éducation, ne jurent que par la courbe normale et vont parfois jusqu'à prétendre qu'elle constitue une constante naturelle. Finalement, la courbe normale est considérée par plusieurs autres comme un outil scientifique dont l'utilisation dépend de choix méthodologiques et circonstanciels.

De telles prises de position et les discussions qui en découlent contribuent souvent à entretenir une certaine confusion entre ce qui relève de la politique et ce qui relève de la science. Nous tenterons, tout en maintenant la discussion à un niveau très simple, de réduire cette confusion en clarifiant, du point de vue scientifique, la fonction de la courbe normale en mesure et évaluation.

Rappelons d'abord que les psychologues ont grandement contribué à développer la méthode des tests en sciences de l'homme. Les efforts de Binet et Simon pour identifier et mesurer les différents facteurs du fonctionnement intellectuel ont contribué substantiellement à l'élaboration d'une méthodologie de la mesure qui, depuis les six dernières décades, a pour ainsi dire envahi tout le secteur de la mesure en psychologie et en éducation. Ce faisant, on a accordé une place de choix à la distribution normale pour des raisons qui débordent les considérations méthodologiques.

En effet, en psychologie, on considère généralement qu'un grand nombre de traits humains se répartissent selon la courbe normale. D'ailleurs, au-delà des considérations méthodologiques, certains ont parfois eu tendance à considérer cette normalité des traits comme une caractéristique naturelle, une constante qui reflétait l'ordre et la symétrie de l'univers. Cette croyance, que nous considérons comme une conclusion hâtive du point de vue scientifique, trouve son origine essentiellement dans deux constatations : premièrement, certaines variables physiques ont une distribution presque normale et, deuxièmement, les résultats aux tests psychologiques ont généralement une distribution presque normale.

De fait, on sait depuis longtemps que certaines variables physiques telles que la taille des nouveau-nés, le diamètre des trous obtenus à l'aide d'une foreuse, la température des êtres humains au repos, la force musculaire, etc., ont toutes une distribution presque normale. Puisqu'un grand nombre de caractéristiques physiques des êtres humains sont distribués normalement et que les résultats aux tests psychologiques ont tendance à se distribuer normalement aussi, certains en ont déduit qu'il en était de même pour les caractéristiques mentales. Ceci explique qu'on ait eu tendance à considérer la normalité comme étant une constante naturelle des mesures psychologiques.

En fait, on oubliait que les traits psychologiques ne pouvant être mesurés directement, il devenait nécessaire d'utiliser des tests construits à partir de principes basés sur une approche différentielle qu'on appellera pour les fins de cet article « principes psychométriques ». Ces principes peuvent être simplifiés comme suit : les êtres humains sont différents, leurs potentialités sont différentes, les tests doivent donc reproduire ces

différences, et la distribution normale permet de bien faire ressortir ces différences en les amplifiant.

On peut voir que la deuxième constatation, à savoir que les résultats aux tests psychologiques ont tendance à se distribuer normalement, résulte de choix méthodologiques. En effet, en choisissant d'élaborer des tests de façon à obtenir des scores distribués normalement, il ne faut pas se surprendre d'obtenir des distributions normales. On ne doit surtout pas en conclure que les traits latents (potentialités) se distribuent normalement.

Il importe de remarquer que ces principes psychométriques ne sont qu'une spécification des principes généraux de mesure, communément appelés principes de « scaling », qui visent à obtenir des échelles de mesure ayant le plus de propriétés métriques, c'est-à-dire dont les scores se prêtent au plus grand nombre d'opérations mathématiques possible. De ce point de vue, on cherche donc à construire un instrument dont les scores sont distribués sur une échelle à intervalles égaux : l'unité de mesure d'une telle échelle étant constante tout au long du continuum, elle permet de comparer des scores distincts, de faire ressortir des différences.

La normalité n'est donc qu'un des postulats à partir desquels un modèle théorique permettant la mesure des différences individuelles est construit. Dans ce sens, l'attitude de certains praticiens et spécialistes en mesure et évaluation qui exigent que les résultats ressemblent à la courbe normale avant qu'on s'en serve est conforme au modèle théorique qu'ils utilisent.

« Les résultats scolaires ou ceux des tests devant logiquement être parallèles aux potentialités des sujets, on a conclu que la répartition (...) des notes se faisait aussi selon une courbe (normale). »

(De Landsheere 1976, p. 345)

D'ailleurs, la prolifération des pratiques de normalisation des distributions observées reflète bien cette attitude. Toutefois, on ne peut transformer un postulat d'un modèle en constante naturelle. On peut difficilement prétendre démontrer que la normalité est « vraie » à l'intérieur d'un modèle qui postule la normalité :

« One cannot — as has happened — use the normal distribution of tests scores as proof that even psychological variables give normally distributed scores, i.e. that the assumption is correct. »

(Magnusson 1966, p. 18)

La normalité demeure un *postulat* qui peut s'avérer très utile dans certaines approches de la théorie des tests ; cependant plusieurs traits du comportement humain se prêtent à des mesures qui ne se distribuent pas normalement. Et cela, essentiellement, parce que ces mesures ne proviennent pas d'une approche psychométrique basée sur la différenciation des individus. On verra par la suite que certaines formes d'évaluation en

éducation ne nécessitent pas le postulat de normalité. Ceci est le cas notamment pour l'apprentissage individuel et l'enseignement par objectifs.

Une brève étude de certains principes utilisés pour l'élaboration de tests en psychologie devrait nous permettre de mieux comprendre l'importance de la courbe normale en psychométrie et, par le fait même, ses utilisations en éducation.

L'un des objectifs de la psychologie est de comprendre et d'expliquer le fonctionnement de la personne. Une démarche fréquemment adoptée pour atteindre cet objectif consiste à étudier les différences individuelles et à essayer de les expliquer. La psychologie différentielle part donc de la constatation qu'il existe des différences individuelles et cherche à approfondir la connaissance du fonctionnement des personnes en expliquant mieux en quoi et pourquoi les individus diffèrent entre eux. En général, les tests psychologiques sont des instruments qui servent à faire ressortir les différences individuelles.

Contrairement à l'opinion courante, il y a plusieurs façons d'élaborer un instrument de mesure comme le montrent, entre autres, Lord et Novick (1968) et Popham (1975). Le choix d'une méthode devrait dépendre de l'usage auquel on destine cet instrument. L'une de ces méthodes d'élaboration d'instruments de mesure — la plus répandue, en fait, permet de maximiser les différences individuelles. Selon cette méthode, les items sont choisis en fonction de leur contribution à la maximisation des différences individuelles. Ceci — comme il sera établi dans ce qui suit — entraîne une distribution normale du trait mesuré.

Dans cette méthode, les deux indices les plus utilisés dans une analyse d'items sont l'indice de difficulté p d'un item, qui désigne la fraction des répondants qui ont réussi l'item, et le coefficient de corrélation item-test. Habituellement, on sélectionne les items ayant une valeur p proche de 0.5 et une corrélation avec le score total au test la plus élevée. À titre d'illustration, examinons les raisons qui justifient le choix de p à l'intérieur de cette méthode.

Premièrement, les valeurs de p influencent la forme de la distribution des scores observés. En effet, lorsque tous les items ont une valeur p proche de 0.5, la distribution des résultats qui en découle est approximativement symétrique et, compte tenu de cette symétrie, elle a la variance totale la plus grande possible. Cette distribution s'approche de la normale : ceci assure une plus grande dispersion des scores observés et maximise donc les différences individuelles.

Deuxièmement, ce choix des valeurs de p tient à des considérations de mesure. La sélection des items qui ont une valeur p proche de 0.5 — ainsi que les corrélations item-test les plus élevées — permet en général, d'obtenir un coefficient maximal de fidélité du test. Ces constatations ont largement contribué à convaincre les concepteurs de tests de la quasi-nécessité d'obtenir une distribution normale. C'est pourquoi on a souvent

vu des praticiens de la mesure et de l'évaluation soutenir que la distribution des résultats à un test doit ressembler à une courbe normale avant de pouvoir se servir du test.

Nous n'avons fait que schématiser l'importance d'un élément parmi d'autres de l'analyse d'items classique, l'indice de difficulté. Il existe plusieurs indices mais tous ont pour effet, délibéré ou non, de mettre au point des instruments de mesure qui différencient les individus entre eux. De tels instruments jouent le rôle de microscopes pour le chercheur en psychologie différentielle. Il est donc tout à fait naturel et raisonnable que ce dernier choisisse d'élaborer un instrument selon une telle méthode. La courbe normale que l'on ne manquera pas d'obtenir résultera donc d'un *choix* et il serait dangereux de conclure que le trait se distribue réellement normalement puisqu'il s'agit d'un artifice psychométrique.

C'est dans ce contexte de *recherche* que s'est développée la méthode des tests en psychologie. Toutefois, l'on a eu vite fait d'utiliser les tests à d'autres fins : l'orientation scolaire et professionnelle, la sélection, le dépistage des troubles mentaux et des troubles d'apprentissage, etc. Il y a donc eu transposition, d'un contexte de recherche à un contexte de décision concernant la vie d'une personne. Cependant, même si les tests étaient utilisés à d'autres fins, on a généralement conservé la même méthode d'élaboration de tests : celle par laquelle on situe systématiquement un individu par rapport aux autres.

Le postulat de la courbe normale étant un *choix*, implicite ou explicite, il déborde le domaine des faits scientifiques et s'avère discutable tant sur le plan méthodologique que philosophique ou politique. L'application de cette méthode d'élaboration des tests dans le domaine de l'éducation aurait donc dû susciter une étude sérieuse de la validité de la transposition des postulats de la dite méthode. Au lieu de cela, l'extension des principes psychométriques dans le domaine de l'éducation s'est faite automatiquement sans remettre en cause ce choix. Il en est résulté une situation pour le moins boiteuse où l'éducateur en milieu scolaire, ayant en général besoin d'instruments lui permettant d'améliorer son enseignement et l'apprentissage individuel, n'a eu à sa disposition que des instruments mettant en valeur les différences entre individus. Il ne faut donc pas s'étonner du peu d'efficacité de tests scolaires construits de cette manière et du peu d'informations directement utilisables par les enseignants pour améliorer ou corriger l'apprentissage individuel. D'où généralement un désenchantement assez grand, voire une certaine méfiance, face aux tests « traditionnels ».

Ce problème est l'objet, depuis quelques années, de recherches assez extensives. Il semble à présent qu'on ait dépassé le stade de la formulation : on connaît le domaine des *tests normatifs* et des *tests critériés*. Les tests normatifs se réfèrent aux différences inter-individuelles des traits mesurés : il s'agit de comparer les individus *entre eux*. Ces tests ont surtout été employés par les psychologues à des fins de sélection dans le milieu scolaire ou à des fins de classification. Cronbach (1971) nous le fait bien remarquer :

« Psychological test theory (...) has concentrated on comparative interpretations. For the psychologist, what the test measures has been defined as whatever causes some persons to get high scores and others to get low scores. If everyone scores the same, (...) the psychologist is likely to say unthinkingly that the test measures nothing. » (p. 446 et 447)

Cronbach (1971) a également observé que l'intérêt spécial de la psychométrie pour ce type de mesure apparaît dans les manuels techniques tels que *Standards for Educational and Psychological Tests* (American Psychological Association, 1966) ainsi que dans les divers « coefficients » utilisés pour valider ces tests.

Les tests critériés s'attachent à faire ressortir ce que l'individu a appris. Ils mesurent donc les différences intra-individuelles : ils servent à comparer chaque individu à *lui-même*. Ces tests sont de plus en plus employés en éducation car ils fournissent des mesures directement interprétables en termes de tâches accomplies lors de l'apprentissage plutôt qu'en termes de traits reliés au groupe ; la différenciation entre individus intéresse moins que la différenciation entre ceux qui atteignent un critère ou pas. La recherche bat son plein dans ce domaine et toute une méthodologie de validation de ces tests est en cours d'élaboration. Ainsi, le concept de fidélité a moins d'importance que la validité de contenu par exemple. Il faut donc s'attendre à ce que les standards des tests en éducation, qui déjà subissent des changements, s'écartent de plus en plus des standards des tests en psychologie.

Carver (1972, 1974) propose de distinguer les tests conçus pour mesurer les différences inter-individuelles de ceux qui mesurent les différences intra-individuelles et appelle « psychométrique » et « édumétrique », respectivement, la dimension dont relèvent ces deux types de tests. Il est important de noter que la plupart des tests relèvent de ces deux dimensions en même temps mais que les méthodes de construction de tests se rattachent généralement à l'une ou à l'autre des dimensions.

Quel est l'effet des principes édumétriques sur la distribution des items d'un test ? Obtient-on une courbe normale ? Il existe plusieurs indices de sélection des items d'un test édumétrique. Afin de faciliter la comparaison avec les tests psychométriques, considérons l'indice de difficulté p . On a vu préalablement que le choix de p proche de 0.5 a pour effet de faire apparaître une distribution approximativement normale des scores, ce qui facilite la comparaison inter-individuelle. Or, dans le cas de tests centrés sur les différences intra-individuelles, il convient de sélectionner les items qui reflètent le mieux les changements (généralement des « gains ») individuels. Ainsi, dans l'approche « avant et après traitement » décrite par Carver (1974),

« the best edumetric items would have p values approaching .00 prior to the treatment condition and p values approaching 1.00 subsequent to the treatment condition. » (p. 513)

Dans ces conditions, il n'y a plus aucune raison pour que les scores de tests éduométriques aient une distribution normale. Plus souvent qu'autrement, on observera des distributions biaisées, voire rectangulaires ou autres.

De tels tests sont non seulement acceptables mais souhaitables. D'ailleurs, les enseignants qui sont acquis à « l'enseignement par objectifs » et qui préparent des « tests maison » font des tests de type éduométrique en se rapprochant plus ou moins de la procédure fort complexe des tests critériés, le critère se réduisant ici à une liste de tâches d'apprentissage à accomplir ou « objectifs ».

En conclusion, la courbe normale est un effet qui découle du choix d'un postulat. Que ce choix soit conscient et délibéré ou non, il implique qu'on accepte les prémisses de la psychologie différentielle. Un instrument élaboré selon les principes psychométriques servira à amplifier les différences individuelles et ne pourra sûrement pas être qualifié d'instrument basé sur une « erreur scientifique ». En effet, si l'on accepte de faire de la sélection en éducation, l'utilisation de ces principes peut s'avérer très adéquate dans certaines situations. Cependant, l'utilisation d'un tel instrument dans une situation où l'on ne privilégie pas l'accentuation des différences individuelles constitue, à tout le moins, une application inadéquate. De telles applications inadéquates peuvent avoir des conséquences sérieuses sur le développement des institutions et des individus concernés.

Il se trouve qu'en éducation, trop souvent, les instruments de mesure sont conçus selon des principes psychométriques qu'il faudrait remplacer par des principes mieux adaptés aux besoins de l'enseignement et de l'apprentissage, à savoir les principes éduométriques. Les résultats des tests construits selon ces principes ne se distribuent généralement pas suivant la courbe normale.

BIBLIOGRAPHIE

- Carver, Ronald P., « Reading Tests in 1970 Versus 1980 : Psychologic Versus Edumetric », *Reading Teacher*, v. 26, 1972, 229-302.
- Carver, Ronald P., « Two Dimensions of Tests 'Psychometric and Edumetric' », *American Psychologist*, v. 29, 1974, 512-518.
- Cronbach, Lee J., « Test Validation », *Educational Measurement*, deuxième édition, R.L. Thorndike, éditeur, American Council on Education, Washington, 1971, 443-507.
- De Landsheere, G., *Introduction à la recherche en éducation*, quatrième édition, Éd. Armand Colin-Bourrelier, Paris, 1976, 403 p.
- Lord, Frederic M., et Novick, Melvin R., *Statistical Theories of Mental Test Scores*, Addison-Wesley Publishing Company, Reading, 1968, 568 p.
- Magnusson, D., *Test Theory*, Addison-Wesley Publishing Company, New York, 1966, 270 p.
- Popham, James W., *Educational Evaluation*, Prentice-Hall, Inc., Englewood Cliffs, 1975, 328 p.