



HAL
open science

Des modèles en éducation aux analyses multivariées

Angela Barthes, Michel Larini

► **To cite this version:**

Angela Barthes, Michel Larini. Des modèles en éducation aux analyses multivariées. *Traité de méthodologie de la recherche en sciences de l'éducation et de la formation*, 2022. hal-03679676

HAL Id: hal-03679676

<https://amu.hal.science/hal-03679676>

Submitted on 26 May 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Des modèles en éducation aux analyses multivariées

Angela Barthes, Michel Larini

Les chercheurs des SEF conduisent leurs travaux soit à partir de données brutes recueillies sur une diversité de terrains d'enquête selon une démarche inductive, soit par une démarche déductive à partir de modèles théoriques de référence reconnus dans la discipline. Ils arrivent souvent que les deux démarches soient articulées à des moments différents de l'enquête. Les statistiques multivariées peuvent s'adapter aux exigences de l'une ou l'autre démarche.

Ce chapitre présente dans un premier temps les bases de la statistique bi- et multi- variée illustré par un exemple pris sur la notation d'élèves. Une deuxième partie présente une démarche inductive classique à partir de données brutes en éducation. Une troisième partie utilise ces techniques d'analyses multivariées sur des modèles existants en SEF selon une démarche déductive. La capacité des SEF à tester et préciser des modèles en éducation leur permet de rendre plus robuste de tels modèles en les confrontant à des données quantitatives objectivables.

1- Les bases de la statistique bi- et multi- variée

Les variables formulées à partir du modèle peuvent être de nature quantitative (âge, taille, poids, notes obtenues, etc.) ou qualitative (genre, nationalité, situation de famille, catégorie-socio professionnelle, lieu d'habitation, comportement, etc.). Les variables quantitatives et qualitatives permettent de catégoriser les *individus*. Dès que le nombre d'individus et/ou de variables sont importants, les informations doivent être traitées pour devenir intelligibles, dans un premier temps en utilisant des indicateurs tels que moyenne, écart Type ou corrélation entre deux variables, puis en considérant le lien entre les variables deux à deux, puis en utilisant une *Analyse en Composante Principale* (ACP) qui permet d'évaluer l'incidence de plusieurs variables sur l'ensemble de la population.

1.1 – L'analyse bi-variée

Dans l'analyse bi-variée, les variables sont prises deux à deux.

Les notes moyennes obtenues par 9 élèves dans quatre disciplines (maths, physique, français, anglais) sont comparables mais les notes de français sont les plus dispersées alors que celle d'anglais sont les plus homogènes, les écart-type en témoignent (figure 1, ci-dessous). Les notes de maths et de physique sont très corrélées alors que les notes de maths et de français ne le sont pas, les deux coefficients de corrélation en témoignent. On peut en conclure que, le plus souvent, les élèves ont obtenu des notes en physique proches de celles obtenues en maths, alors que ce n'est pas le cas pour les notes de français. Pour obtenir une représentation visuelle, centrée sur les individus, deux systèmes d'axes (MATHS, PHYS) et (MATHS, FRAN) ont été superposés et chaque individu a été positionné avec pour abscisse la note de maths et en ordonnées, d'une part, la note de physique et, d'autre part, la note de français. Deux nuages de points très différents apparaissent. Concernant le couple (MATHS, PHYS), les élèves se retrouvent pratiquement alignés ce qui signifie, comme le laissait prévoir le coefficient de corrélation qu'il y a une relation entre les notes de maths et de physique. Concernant les variables (MATHS, FRAN) le nuage de point est très dispersé ce qui reflète la faible valeur du coefficient de corrélation entre maths et français. Il est possible d'identifier chaque individu et à l'aide d'une variable catégorielle, différencier les individus appartenant à un même groupe, fille ou garçon par exemple.

Elèves	MATHS	PHYS	FRAN	ANGL
E1	6	6	5	5,5
E2	8	8	8	8
E3	6	7	11	9,5
E4	14,5	14,5	15,5	15
E5	14	14	12	12,5
E6	11	10	5,5	7
E7	5,5	7	14	11,5
E8	13	12,5	8,5	9,5
E9	9	9,5	12,5	12
Moyenne	9,66	9,83	10,22	10,05
Ecart Type	3,37	2,99	3,47	2,81

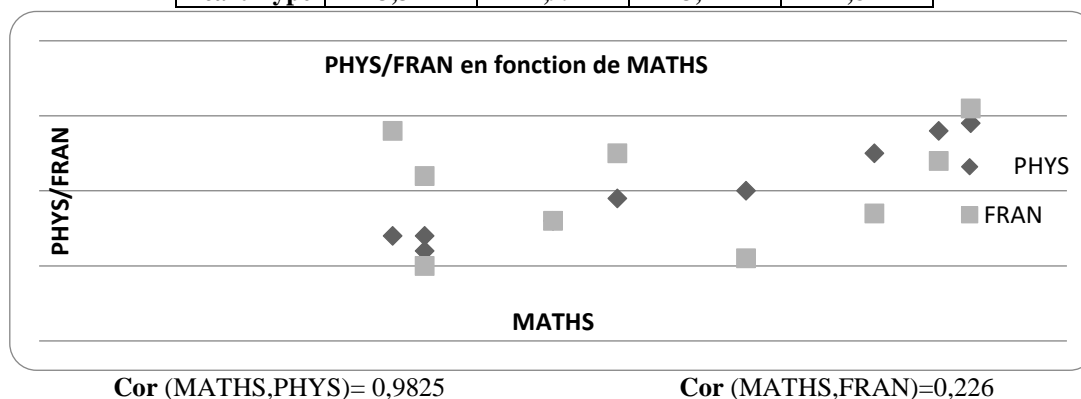


Figure 1 : Exemple de statistique bi-variée

Pour approfondir l'analyse et vérifier l'incidence des 4 variables sur la population, il convient de réaliser une *analyse en composante principale* (ACP). Dans le cas général, il est possible d'avoir un nombre quelconque de variables.

1.2 - Les fondements de l'analyse en composante principale

Si les bases mathématiques de l'ACP sont complexes (Benzeckri, 1980), la simplicité de la mise en œuvre avec un logiciel adapté¹ est avérée et les potentialités d'adaptation aux SEF très importantes. Pour présenter les fondements de l'ACP, la démonstration est réduite à 3 variables (V1, V2, V3 et N individus) car une représentation graphique des diverses étapes est ainsi possible. La démarche peut être généralisée aux cas de plus de 3 variables, sauf la représentation graphique.

En suivant une démarche comparable à l'analyse bi-variée, il est possible de positionner chaque individu, mais cette fois, dans un espace à trois dimensions défini par les trois axes du repère trirectangle portant respectivement les variables V1, V2, V3 (fig. 2, ci-dessous). L'ensemble des individus forment alors un nuage de points. La représentation des individus dans un espace à 3 dimensions est impossible à lire. Pour remédier à cela, le logiciel cherche le plan le *plus proche* du nuage de point. Il utilise la méthode des *moindres carrés* qui lui permet de déterminer les axes F1 et F2 qui définissent ce plan le *plus proche* (F1, F2). Un axe F3 est alors défini de telle sorte que le repère (F1, F2, F3) soit un repère trirectangle direct. Une fois ce plan déterminé, le logiciel projette orthogonalement tous les individus sur le plan (F1, F2) dit *plan principal*. L'ensemble des individus projetés forment un nuage de point sur ce plan, et le problème est ainsi ramené au cas précédent. Le nuage projeté devient l'objet de l'étude. A noter deux aspects : 1) le nuage de points projeté sur le plan a perdu une partie des informations contenues dans le nuage initial en 3 dimensions, par conséquent, suivant le pourcentage d'information conservé dans le plan la qualité de l'ACP sera plus ou moins bonne ; 2) si chaque point projeté sur le

¹ Liste non exhaustive : Excel Stat, SPHINX, SPSS, Modalisa, SAS et R.

plan principal a pour coordonnées f_1, f_2 sur les axes F_1 et F_2 alors f_1 et f_2 sont une combinaison des coordonnées initiales V_1, V_2, V_3 .

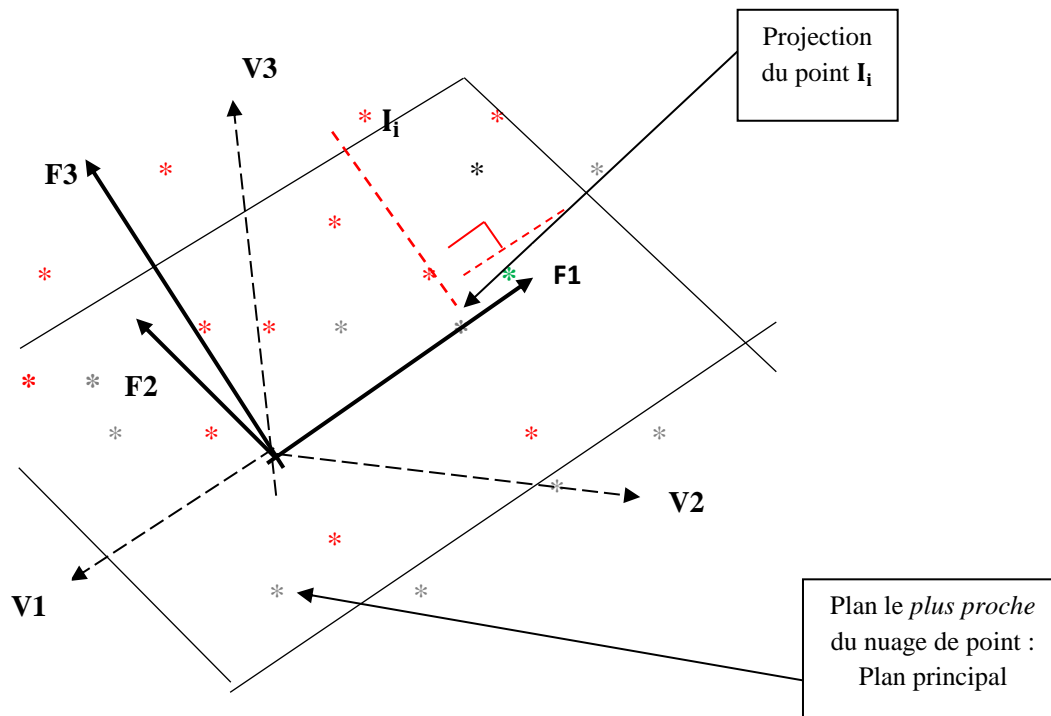


Figure 2 : Représentation 3 D

S'il y a plus de 3 variables initiales, le logiciel doit effectuer successivement plusieurs projections pour passer du nuage initial dans un espace à $L > 3$ dimensions au nuage sur le plan principal F_1, F_2 .

2 - Cas d'étude de démarche inductive à partir de données brutes

L'incidence des variables prises deux par deux venant d'être étudiée, il s'agit d'observer l'incidence simultanément de plusieurs variables sur la population.

Quatre variables sont testées : V_1 (MATHS), V_2 (PHYS), V_3 (FRAN), V_4 (ANGL), le calcul est lancé pour réaliser l'ACP³ (tableau 1 à 5, ci-dessous).

	F1	F2	F3	F4
Variabilité (%)	71,892	27,992	0,089	0,026
% cumulé	71,892	99,884	99,974	100,000

Tableau 1 : Variance expliquée (% informations contenues par chacun des axes factoriels)

² Remarque : le logiciel transforme les variables initiales V_1, V_2, V_3 en variables centrées réduites $\widehat{V}_1, \widehat{V}_2, \widehat{V}_3$ telles que : $\widehat{V}_i = \frac{V_i - \bar{V}_i}{S(V_i)}$ Où \bar{V}_i est la valeur moyenne de la variable V_i et $S(V_i)$ son écart type. Le but de ce changement de variable est de faire en sorte que les variables soient centrées dans le repère initial et qu'elles aient des normes comparables. Cela mérite d'être dit mais ne modifie en rien l'approche et la compréhension de la méthode.

³ Utilisation du logiciel *Excel Stat*.

Lors des projections successives jusque-là, le nuage projeté dans le plan principal (F1, F2) a perdu une partie de l'information contenue dans le nuage initial 4D, or les résultats obtenus (tableau 1, ci-dessus) renseignent sur cette perte d'information. La 4D permet de visualiser 4 axes factoriels. Les deux premiers contiennent 71,892 % et 27,992 % de l'information contenue dans le nuage initial. Ainsi, le plan principal (F1, F2) contient 99,884% de l'information du nuage initial ce qui est remarquable. Le plan (F1, F2) présente donc un degré de fiabilité et, compte-tenu du peu d'informations contenues par les axes F3 et F4, tous les autres plans factoriels pourront être ignorés.

	F1	F2	F3	F4
MATHS	0,478	-0,552	0,203	0,652
PHYS	0,532	-0,407	-0,441	-0,597
FRAN	0,444	0,621	-0,532	0,365
ANGL	0,540	0,379	0,693	-0,290

Tableau 2 : Contribution des variables initiales aux variables factorielles

Les coordonnées f_{1i} et f_{2i} d'un individu i projeté sur le plan principal (F1, F2) représentaient des fonctions des variables initiales (MATH, PHYS, FRAN, ANGL), les résultats obtenus (tableau 2, ci-dessus) renseigne sur cette dépendance.

$$f_{1i} = 0,478 \times \text{Maths}_i + 0,532 \times \text{PHYS}_i + 0,444 \times \text{FRAN}_i + 0,540 \times \text{ANGL}_i$$

$$f_{2i} = -0,552 \times \text{Maths}_i - 0,407 \times \text{PHYS}_i + 0,621 \times \text{FRAN}_i + 0,379 \times \text{ANGL}_i$$

Ce résultat permet de donner du sens à chacun des quadrants du plan factoriel de la représentation. Pour l'axe F1, f_1 croit quand chacune des variables initiales croit, ainsi plus f_1 est important et plus les élèves sont performants. Pour l'axe F2, les contributions des variables (MATHS et PHYS) sont négatives alors les contributions des variables (FRAN et ANGL) sont positives. Ainsi, si f_2 est positif les élèves sont plus performants dans les disciplines littéraires et si f_2 est négatif ils le sont davantage dans les disciplines scientifiques. On peut alors donner du sens aux différents quadrants du premier plan factoriel (F1, F2), la démarche conduit à la construction d'une autre représentation (fig. 3, ci-dessous).

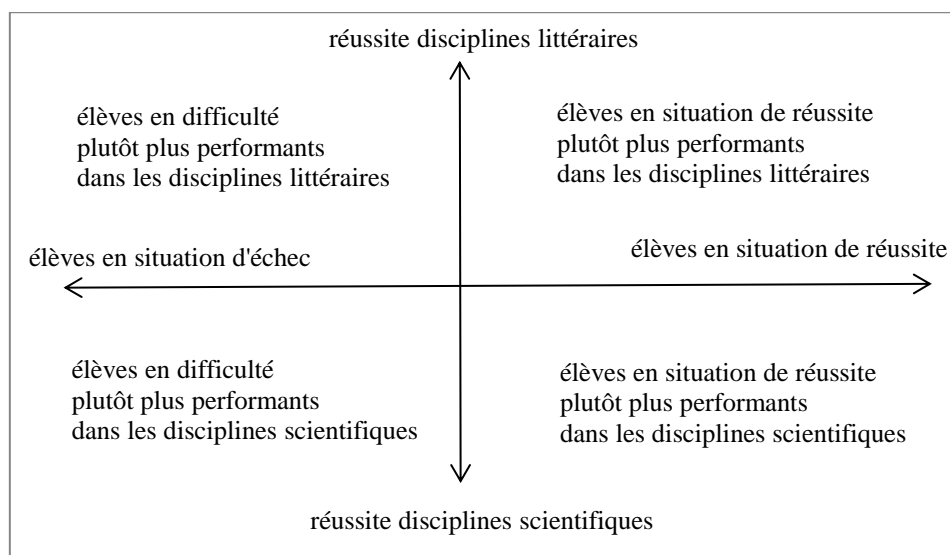


Figure 3 : Signification des cadrans du plan principal (F1, F2)

La représentation produite par le logiciel (fig. 4, ci-dessus) permet de visualiser la position de chaque individu dans le plan principal.

L'élève (4) est celui qui a obtenu les meilleurs résultats et que ses notes dans les domaines scientifiques et littéraires sont équilibrées car il est proche de l'axe F1. En revanche, l'élève (1) est celui qui a obtenu les plus mauvais résultats. Concernant l'élève (8) il a des résultats plutôt bons en moyenne mais avec surtout des bonnes notes dans le domaine scientifique.

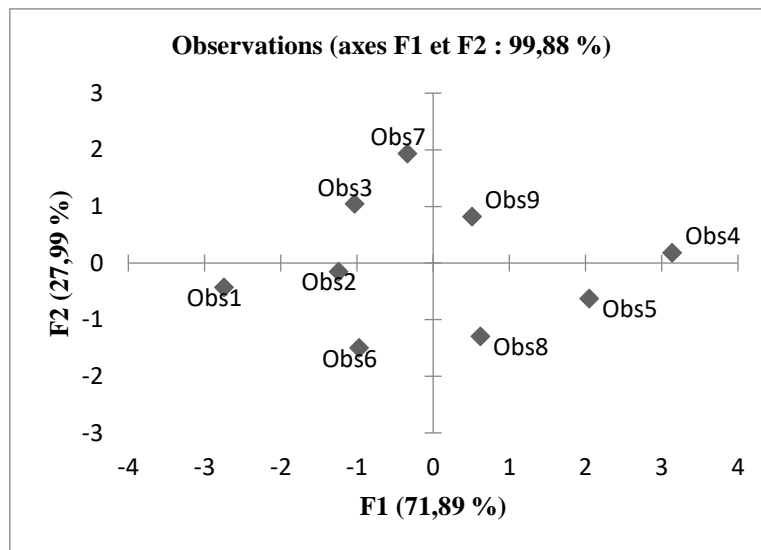


Figure 4 : Représentation des individus dans le plan principal

La représentation produite par le logiciel (fig. 5, ci-dessous) sur laquelle sont représentées quatre variables projetées dans le plan principal à l'intérieur d'un cercle de rayon unité avec pour origine le centre du cercle. Cette figure contient et résume plusieurs informations importantes. Les quatre variables ont leurs extrémités sur le cercle, cela signifie que les variables sont très bien représentées dans le plan principal. Au contraire, plus les extrémités sont éloignées du cercle de rayon 1, moins l'ACP est de bonne qualité.

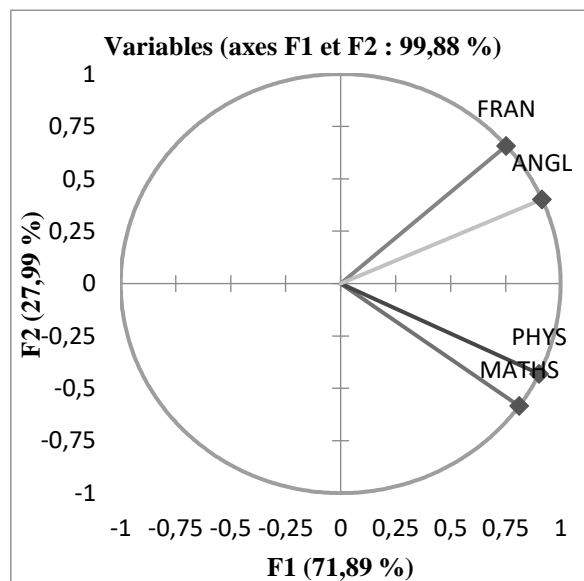


Figure 5 : Représentation des variables dans le plan principal (F1, F2)

Les quatre variables initiales (MATHS, PHYS, FRAN, ANGL) sont disposées de telle manière dans le cercle que le plan principal contient plus de 99% de l'information initiale. Les variables MATHS et PHYS sont proches et font entre elles un angle relativement faible, il en va de même des variables FRAN et ANGL mais les deux couples sont éloignés. Cela exprime que les variables MATHS et PHYS d'une part et les variables FRAN et ANGL d'autre part sont très corrélées mais que les couples se distinguent nettement.

Ainsi, lorsque deux variables sont très corrélées, elles forment entre elles un angle faible alors que si deux variables sont peu corrélées elles forment entre elle un angle important pouvant aller jusqu'à 90°. Ces résultats peuvent être vérifiés en calculant les coefficients de corrélation entre les variables à partir du tableau 2. Le sens des *vecteurs variables* renseigne sur la signification des 4 cadrans.

Les 4 variables sont orientées dans le sens des F1 croissant positifs ce qui suggère que plus le point se déplace vers la droite, plus les élèves sont performants. Par ailleurs, les variables FRAN et ANGL sont orientées vers le haut alors que les variables MATHS et PHYS sont orientées vers le bas. Ainsi, plus les élèves se trouvent placés vers le hauts, plus ils sont performants dans les disciplines littéraires et plus ils sont positionnés vers le bas, plus ils sont performants dans les disciplines scientifiques.

Dans un cas aussi simple l'ACP est de très bonne qualité. Si la qualité est moins fiable, en particulier avec l'augmentation du nombre de variables, l'interprétation peut alors devenir plus délicate et demande quelquefois de travailler sur d'autres plans que le plan principal (F1, F2).

3 - Démarche déductive sur les modèles d'éducation et formation par les statistiques multivariées

Plutôt que d'approcher un phénomène sur des bases empiriques, il s'agit davantage dans ce cas de tester le modèle théorique en le confrontant à des données quantitatives objectivables. Passer d'un modèle d'analyse en éducation à un recueil de données permettant d'en tester la robustesse suppose la définition d'un protocole de recherche en cinq étapes (fig. 1, ci-dessous).

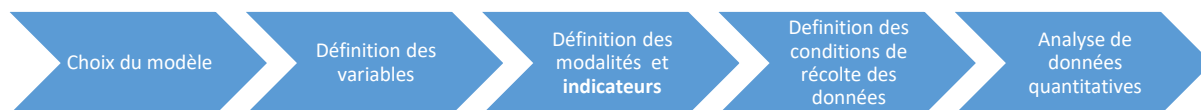


Figure 6 : Du modèle au traitement des données

Issu d'un état de l'art et/ou du cadre théorique, le choix du *modèle d'analyse* correspond au besoin d'explicitation d'une situation éducative. Il s'agit donc dans un premier temps d'explicitier le modèle afin d'objectiver et de définir les *variables* les plus pertinentes permettant de caractériser au mieux le phénomène dont il rend compte. Il s'agit alors de définir des *indicateurs* qui permettent d'attribuer une valeur numérique, un score, à chacune des variables. A partir de cette étape, il est possible de concevoir une *méthode de récolte des données* permettant d'apporter une réponse à toutes les modalités de toutes les variables. L'ensemble des données recueillies sont rassemblées dans un tableau qui présente la réponse des individus (au sens statistique du terme) pour chaque variable. Ce tableau est généralement illisible en l'état. Il convient alors d'en produire une analyse bi-, puis multi-variée.

Le modèle des *rapports aux savoirs* (Charlot, 1997) conduit à explorer l'articulation des parcours singuliers et des rapports sociaux. Ce travail est souvent opéré afin de comprendre les situations scolaires de réussites, d'échec, voire de décrochage scolaire. Il s'agit donc dans l'étude des *rapports au savoir* de comprendre les articulations d'une signification sociale (en général en lien avec un but) avec un sens plus individuelle. Les rapports sociaux se construisent dans une histoire collective et dans une relation personnelle à ces modèles sociaux. Par ailleurs, considérer les rapports aux savoirs, implique d'appréhender l'activité de leur appropriation. Dans l'éducation, il s'agit d'un rapport à un processus d'apprentissage, plus largement, c'est le fait de posséder la maîtrise d'une activité de réflexion en situation. Ainsi donc, selon le champ socio-anthropologique des *rapports aux savoirs*, il y a trois dimensions à considérer : la relation de sens du sujet vis-à-vis de son objet de savoir ; la relation aux modèles sociaux ; la relation au processus d'appropriation. Dans le vocable des *rapports aux savoirs*, il s'agit respectivement des dimensions dites *identitaires* ; *sociales* ; *épistémiques*.

Les méthodologies de l'étude des *rapports aux savoirs* sont traditionnellement axées sur des enquêtes (appelées bilans de savoirs) investiguant l'individu (sexe, parents, origine ethnique, histoire antérieure, relations sociales, image de soi, identité personnelle et sociale, compétences, modèle de références, projection dans le futur, etc.) et leur réponse à une situation sociale donnée - qui fait sens ou non - dans une situation précise (très souvent les études tournent autour du rapport au monde et à la forme scolaire).

Les bilans de savoirs ne consistent pas à évaluer ce qui a été appris, mais plutôt ce qui fait sens pour le sujet dans ce qui a été appris. Les bilans de savoirs sont très complets mais d'une part ils sont extrêmement chronophages à conduire, d'autre part, ils permettent l'analyse des situations, mais sont peu reproductibles car contenant une très grande part de subjectivité analytique par exemple dans les cas comparatifs ou évolutifs.

Une méthode a donc été développée pour approcher et positionner dans un ensemble reproductif les dimensions des *rapports aux savoirs*. Chacune de ces trois dimensions ou variables définissent alors des facteurs d'analyses susceptibles d'être quantifiés. La seconde démarche consiste donc à exprimer le modèle en variables, indicateurs, modalités de façon synthétiques qui se rattache au contexte de l'étude menée.

Une étude a porté sur les rapports aux savoirs et aux territoires de proximité d'élèves de 3^{ème} afin de vérifier s'ils influent sur leur prise en charge des enjeux et problématiques de durabilité (Morin, Lange, Barthes 2020). Il s'agit d'expérimenter une action d'éducation au développement durable permettant d'évaluer la manière dont les élèves problématisent les enjeux de développement de leur territoire de proximité. Il a été choisi volontairement des élèves dans les contextes sociaux et géographiques très différents, de la banlieue au rural.

Le modèle des *rapports aux savoirs* qui comprend 3 variables est donc décliné en 5 modalités qui comprennent des indicateurs permettant de rapporter une situation de terrain observée à une modalité d'une variable (tableau 3, ci-dessous).

Expressions du rapport au monde lors de la confrontation aux productions d'autres groupes.			
	variable 1, dimension identitaire	Variable 2, dimension sociale	Variable 3, dimension épistémique
n 1	La problématisation est déconnectée d'investissement identitaire .	Les questions de normes ne sont pas abordées.	Les apports d'informations nouvelles ne sont pas identifiés
n 2	Les participants expriment une facette de leur identité propre (dans les exemples cités).	Les participants indiquent « <i>ce qui se fait</i> », « <i>ce qu'il convient de faire</i> ».	Les participants constatent eux-mêmes des différences dans les informations mobilisées, sans modifier pour autant leur raisonnement.
n 3	Les participants mobilisent une/des facettes de leur identité propre dans leur problématisation.	Les participants soulignent des différences sur ce qui paraît « normal » à différents groupes.	Les participants intègrent des connaissances nouvelles dans leur îlot de rationalités (sans les remettre en question).
n 4	Les participants mobilisent dans leur problématisation des facettes de leur identité propre ET de l'identité du territoire dans lequel ils habitent.	Les participants s'interrogent sur ce qui paraît « normal » dans leurs différents contextes.	Les participants questionnent la fiabilité des différentes informations reçues.
n 5	Les participants justifient leur problématisation par des facettes de leur identité propre et /ou de l'identité du territoire dans lequel ils habitent.	Les participants discutent (et donc explicitent) leur conformité aux usages, traditions, et pratiques de leurs contextes socio-culturels.	Les participants discutent (et donc explicitent) les intérêts et limites, la cohérence ou la pertinence des diverses connaissances mobilisées dans les docs 1 et 2.

Source : Morin *et al.* (2020)

Tableau 3 : Traduction de l'épistémologie des rapports aux savoirs en variables et modalités

Le recueil des données a été réalisé par questionnaire et entretien, chaque question renvoyant à une modalité de la grille (tableau 3, ci-dessus). Chaque individu coche une modalité à laquelle un score est attribué (de 1 à 5), les scores les plus élevés correspondants aux adhésions les plus fortes (n5) à la variable. En l'occurrence, la variable « *le quotient d'autonomie de problématisation des élèves* » a été rajouté pour évaluer au fur et à mesure que des situations sont présentées aux élèves, là où ils en sont dans le processus de problématisation de l'enjeu de développement durable dont il est question dans leur territoire de proximité. Les élèves sont également catégorisés par CSP, sexe, niveau scolaire, etc, Les collèges et territoires selon leurs contextes sociologiques et géographiques. Il s'agit aussi de voir la part de reproduction sociale,

la part des *rappports aux savoirs* et la part des contextes dans la prise en charge par les élèves des enjeux de la durabilité.

Il s'agit donc de mettre en relation *les rapports aux savoirs* et la prise en charge de la problématisation dans une situation éducative donnée, puis par la suite d'intégrer les variables qualitatives pour porter une analyse. Pour cela, une *Analyse en Composantes Principales (ACP)*⁴ est appliquée aux 4 variables (tableau 4, ci-dessous).

Variabiles	Facteur 1	Facteur 2
rapport à. Identitaire	0,859	-0,341
rapport à. Social	0,781	-0,355
rappports à. Épistémique	0,549	0,794
Quotient autonomie de problématisation	0,853	0,157

Tableau 4 : Contribution des variables initiales aux variables factorielles

Puis nous calculons l'ACP et cela nous donne la représentation suivante :

Facteur 1 : 59,50%
Facteur 2 : 25,46%

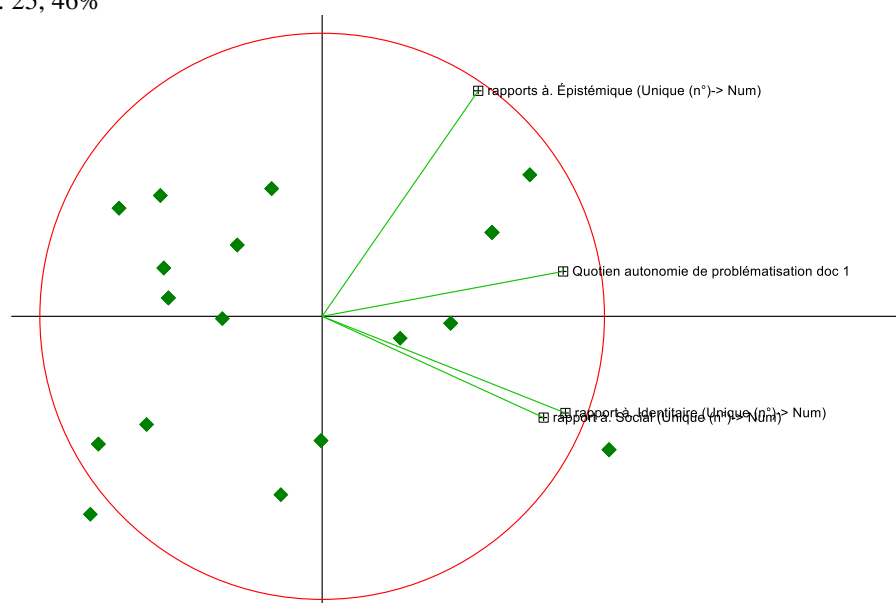


Figure 7 : Cercle des corrélations de l'étude de cas des rapports aux savoirs

L'ACP réalisée est de très bonne qualité car le *plan principal* contient $59,50\% + 25,46\% = 84,96\%$ des informations contenues dans le tableau de données initial. Les 4 variables sont bien représentées à l'intérieur du cercle des corrélations car leur extrémité est proche du cercle. Les angles que font les 4 vecteurs variables montre que V1 (dimension identitaire) et V2 (dimension sociale) sont très corrélées. En revanche V3 (dimension épistémique) est peu corrélée de V1 et de V2 et V4 (le quotient d'autonomie de problématisation) est partiellement corrélées aux trois autres variables. Il est possible d'en conclure que les dimensions des *rappports aux savoirs* sont fondamentales aux sujets pour problématiser des enjeux de développement durable. En poursuivant la démonstration sur les ACP, la contribution des variables initiales aux variables factorielles permet de donner du sens à chacune des zones du plan.

⁴ Utilisation du logiciel *Modalisa*.

Plus le point se déplace vers la droite et plus les 3 dimensions du *rapport au savoir* (V1 + V2 + V3) sont importants, de même que pour V3. Ainsi, plus les individus sont positionnés à droite et plus ils sont favorablement influencés par les 4 variables. Plus ils sont positionnés à gauche et moins ils sont influencés. D'un point de vue de l'étude de cas, les sujets qui sont les moins autonomes dans la problématisation des enjeux de développement durable de leur territoire de proximité sont ceux qui ont peu d'investissement identitaire et qui interrogent peu les normes. D'un point de vue éducatif, conduire les élèves à interroger les normes, en particulier les normes dominantes d'un point de vue critique permet une prise en charge citoyenne des enjeux de développement.

Plus les individus sont positionnés haut dans le plan et plus ils sont influencés, principalement, par la variable V3, c'est-à-dire que plus ils discutent et intègrent les nouvelles informations, plus *a priori* ils sont à même de problématiser les enjeux, même si la variable V4 n'est que plus modérément corrélée. Enfin, plus les individus sont positionnés bas et plus ils sont influencés par les variables V1 et V2, puisque la dimension épistémique, n'est pas corrélée aux deux autres.

L'examen des variables qualitatives des individus positionnés dans le cercle montre donc que ce ne sont pas spécialement les élèves les plus performants, ou ceux dont les parents sont CSP + qui comprennent le mieux les enjeux de durabilité, mais ceux dont les *rapports aux savoirs* se trouvent dans les modalités 4 et 5, donc ceux qui expriment une identité propre et celle de leur territoire en même temps qu'ils discutent les normes. La variable genre n'intervient pas ici. En réalité, les milieux urbains sensibles et les milieux ruraux isolés sont les plus autonomes dans la problématisation, ce qui renvoie bien, sans rentrer dans les détails, à l'adaptation indispensable des enseignements aux contextes entre autres.

A ce stade, il est possible de rechercher des régressions linéaires multiples entre les variables. Il est possible d'exprimer l'une d'entre elles (*variable expliquée*) en fonction d'une ou plusieurs autres (*variables explicatives*).

La variable V4 « *quotient d'autonomie de problématisation des élèves* » en fonction des trois variables V1 (rapport à l'identitaire), V2 (rapport au social), V3 (rapport à l'épistémique). Lors de la régression multiple, le logiciel renvoie le résultat : $V4 = 0,081 + 0,060 \times V1 - 0,003 \times V2 + 0,053 \times V3$.

Ainsi, il apparaît que V4 croît de façon sensiblement identique avec V1 et V3. La contribution de la variable V2 est négative et de plus pratiquement négligeable, ce qui signifie que l'autonomie de problématisation ne dépend pas des rapports aux normes sociales.

Le logiciel produit d'autres données essentielles qu'il n'y a pas lieu de présenter dans ce chapitre, mais qui permettent d'apprécier, à l'aide de tests statistiques, la qualité et la précision de la régression⁵. Il est vivement conseillé de les prendre en compte car, quelle que soit la dispersion des résultats, le logiciel renvoie toujours une régression (Bressoux 2010 ; Larini, Barthes, 2018).

Conclusion

Les ACP sont des traitements classiques robustes et puissants qui peuvent être utilisés de façon classique dans une démarche déductive, mais il est possible de consolider les modèles produits par les SEF pour gagner en robustesse et reproductibilité.

Références bibliographiques

⁵ Dans le cas présent, la p. value qui exprime la probabilité de se tromper est faible pour V1 et V3 (0.001), les résultats sont donc fiables, plus élevée pour V2.

- Benzeckri, J-P. (1980). *Analyse des correspondances, exposé élémentaire*. Paris : Dunod.
- Bressoux, P. (2020). *Modélisation statistique appliquées aux sciences sociales*. Bruxelles : De Boeck.
- Charlot, B. (1997). *Du rapport au savoir. Éléments pour une théorie*. Paris : Anthropos.
- Charlot, B., Bautier, E., Rochex, J-Y. (1992). *École et savoir dans les banlieues et ailleurs*. Paris : Armand Colin.
- Kalali, F., Charlot B. (2017). Rapports aux savoirs. In A. Barthes, J-M. Lange, N. Tutiaux-Guillon (dir.), *Dictionnaire critique des enjeux et concepts des éducations à* (p. 544-548). Paris : L'Harmattan.
- Larini, M., Barthes, A. (2018). *Quantitative and Statistical Data in Education: From Data Collection to Data Processing*. Hoboken, New Jersey: Wiley.
- Larini, M., Barthes, A. (2018). *Statistiques et traitement de données quantitatives en éducation*. Londres : ISTE Edition Ltd, Série Education.
- Morin, O., Lange, J-M., Barthes, A. (2020). Caractéristiques cognitives et sociales de l'engagement territorialisé des jeunes en contexte scolaire vis-à-vis des enjeux de la durabilité/soutenabilité. *Educations. International sciences and technical edition*, 20-2, 29-52.