

# L'utilisation de l'analyse implicative et cohésive pour exploiter un test standardisé de mathématique

Antoine Bodin • Université de la Méditerranée - antoinebodin@mac.com  
Maria Lucia Giovannini • Università di Bologna - marialucia.giovannini@unibo.it  
Liliana Silva • Università di Bologna - liliana.silva@unibo.it

## Use of Implicative and Cohesive Analysis for exploiting a mathematical standardized test

La recherche présentée dans cet article concerne une analyse de données issues d'un test diagnostique standardisé de mathématiques passé au début de la première année de l'école moyenne. L'analyse implicative et cohésive a permis de mettre à plat les résultats de ce test d'une façon qui nous semble à la fois originale et complémentaire des analyses classiques. L'article analyse les relations entre les questions du test et entre ces questions et la perception des élèves et apporte des éléments susceptibles d'intéresser la formation des enseignants. L'article présente succinctement la méthode d'analyse implicative, son application à l'ensemble du test, des analyses partielles d'éléments composant le test ainsi qu'une analyse de certaines des habilités définies par les concepteurs du test. La méthode utilisée permet de mettre en évidence la structure fine du test et l'organisation cognitive des élèves qui l'ont passé.

**Mots-clés :** test diagnostique, test d'entrée, analyse implicative, complexité cognitive, mathématiques, collège

The research presented in this paper concerns a data analysis of a standardized diagnostic test used in 2008 at grade 6 level in the Bologna area (Italy). The implicative and cohesive analysis has been used to take a close look to the results of this test in a way that we think as both original and complementary to other analysis. The paper brings some information on the links among test items and between these items and some students' perceptions. Therefore, several points of interest for teachers' development are suggested. The paper briefly explores the item complexity issue, the difference between complexity and difficulty, the use of the implicative analysis for the whole test as well as for some partial analysis, particularly concerning the abilities defined by the test developers. The method allows to put in light the intimate structure of the test as well as the cognitive structure of the students.

**Keywords:** assessment, diagnostic test, implicative analysis, cognitive complexity, mathematics, middle school

83

ricerche

\* Bien que l'article soit le résultat d'un travail partagé, les paragraphes 1 et 2 ont été rédigés par Liliana Silva, les paragraphes 3 et 5 par Maria Lucia Giovannini et le paragraphe 4 par Antoine Bodin.

# L'utilisation de l'analyse implicative et cohésitive pour exploiter un test standardisé de mathématique

## 1. Introduction

Les données analysées dans cet article sont issues d'une évaluation diagnostique du domaine mathématique effectuée en collaboration avec 21 établissements scolaires de première année de l'enseignement secondaire (élèves de 11-12 ans). Ce niveau scolaire a été choisi parce que l'école moyenne (*scuola media o scuola secondaria di 1° grado*) est considérée en Italie comme étant «le maillon faible» du système éducatif: en elle sont enracinées les lacunes et les difficultés que de nombreux étudiants trouvent dans les deux premières années de l'enseignement secondaire du second degré (Fondazione Agnelli, 2011). En outre, le fait que la transition du primaire au secondaire soit souvent difficile pour de nombreux élèves est une source de préoccupation pour les familles.

Les élèves de la première année se trouvent devant des enseignants qu'ils ne connaissent pas et qui très souvent utilisent une démarche d'enseignement différente de celle de l'école primaire. De leur côté les enseignants ne connaissent ni les élèves ni leurs expériences d'apprentissages. En début d'année, dans certaines matières on utilise souvent des tests pour vérifier les connaissances et les habilités des élèves ; cela pour mieux organiser les activités didactiques en tenant compte des connaissances effectives des élèves de la classe. Par rapport à de tels tests des réserves sont souvent exprimées aussi bien en ce qui concerne le format fermé des questions que l'utilisation faite des résultats pour noter et classer les élèves.

Conscients de la complexité de l'évaluation des apprentissages y compris celle de début d'année et convaincus que les interventions didactiques mises en actes par les enseignants doivent s'appuyer sur ce que les élèves savent déjà, certains choix ont dus être faits. Le test a été construit de façon à ce que soient respectés les principes de planification, de rédaction et d'administration dans un nombre important de classes, mais aussi de façon à rendre possible une analyse des résultats recueillis pour permettre une utilisation diagnostique au service des élèves.

L'analyse implicative et cohésive présentée dans cet article est destinée à étudier dans quelle mesure cette méthode est de nature à compléter les analyses habituelles des résultats d'un test (analyses statistiques classiques, analyse des réponses aux items) et aussi les informations de nature qualitative. L'analyse statistique implicative développée par Gras (1979) et ses élèves, initialement dans le cadre de la didactique des mathématiques, est une méthode d'analyse qui permet d'étudier l'organisation d'un ensemble de données et particulièrement d'étudier les dépendances orientées entre variables. En particulier le présent article explore la contribution que l'analyse implicative peut apporter à l'analyse des résultats d'un test de mathématiques en évitant qu'ils ne débouchent sur des notes et des classements tout en favorisant des processus d'enseignement-apprentissage efficaces.



## 2. Cadre théorique

L'évaluation des apprentissages constitue indubitablement une question complexe qui pose de nombreux problèmes, dont celui des fonctions et des significations qu'on lui attribue.

Un point important concerne l'*évaluation diagnostique* au service du soutien du processus d'enseignement-apprentissage, une fonction à considérer dans toutes les phases de la gestion du test. En fait les réponses données par les élèves sont relatives à des connaissances et habilités mathématiques de base pour faciliter l'*hétérorégulation de l'enseignant* et l'*autorégulation des élèves*. Nous avons donc envisagé que les résultats de ce test peuvent aider les enseignants au niveau de la pratique d'enseignement individualisé mais aussi de dépassement de l'*autoréférentialité*. L'un des principes du groupe qui a planifié et mené le test consistait à considérer nécessaire une *restitution* à chaque enseignant participant à l'étude des résultats de sa classe (par types de compétences, ainsi que pour la globalité des classes).

Nous savons que l'interprétation des résultats des élèves peut être favorisée par la disponibilité de données relatives aux perceptions des élèves et des enseignants par rapport à l'étude des mathématiques mais aussi aux *perceptions des élèves* sur des aspects spécifiques d'une épreuve donnée. Par exemple sur le niveau de difficulté générale du test et leur niveau de réussite par rapport au test, sur la difficulté de chacun des items, et sur la possibilité que le test leur donne de mobiliser leurs habilités mathématiques, ainsi que sur la nouveauté pour eux de certains exercices présents dans le test.

Toujours d'un point de vue théorique, nous avons aussi retenu que l'utilisation d'un test planifié étalonné de façon rigoureuse, validé sur un échantillon important, permet non seulement d'obtenir *une mesure fiable et valide des habilités considérées*, définie de façon claire par rapport à une cadre de référence théorique, explicite et bien articulé, mais aussi celui de pouvoir interpréter les résultats obtenus dans une classe à la lumière des résultats obtenus pour l'ensemble des classes impliquées. Toutefois, une référence de ce type nécessite l'utilisation de procédures standardisées en relation avec les différentes étapes, depuis l'administration du test jusqu'à celles de la correction et de l'interprétation des résultats, mais aussi de la vérification de la qualité des mesures (Gattullo, 1967; De Luca et Lucisano, 2011). Le test dont nous analysons les résultats est défini comme un test standardisé présenté de la même façon à tous les élèves et avec une détermination a priori des réponses correctes. Ces avantages de la démarche définissent aussi les limites de ce type de test. Ainsi que le précisait Gattullo en 1978, les tests d'apprentissages «représentent exclusivement un instrument relativement plus fonctionnel pour la mesure de certains aspects des apprentissages, mais ne construisent pas un système "scientifique" d'évaluation» (Gattullo, 1978, p. 247). Il est aussi nécessaire de considérer le type de connaissance et d'habilité mesurées, la conscience que d'autres qui ne sont pas prises en considération demandent différentes modalités de recueil d'information pour préserver le principe de validité de l'évaluation. Pour pouvoir diagnostiquer la "situation de départ" on privilégie le principe de procéder en une première phase pour «effectuer des évaluations sur des points essentiels pour tous les élèves» (Gattullo et Giovannini, 1989, p. 15), et pour approfondir ensuite les autres aspects et spécificités des situations. Cela se base sur le principe que les objets d'une évaluation en entrée concernent non seulement les connaissances et les habilités disciplinaires mais aussi d'autres aspects ; on a donc besoin d'utiliser divers outils et procédures pour recueillir les données et aussi pour analyser les résultats.

Pour la présente recherche, nous avons choisi d'utiliser l'analyse implicite parce



que, cette méthode de traitement et de visualisation des données permet une mise à plat des questions d'un test et des relations de dépendances qu'elles entretiennent entre-elles. À partir de cette mise à plat, on peut repérer des graphes et des chaînes qui permettent d'analyser les réussites et les échecs des élèves. On obtient aussi des arbres orientés qui fournissent des vues d'ensemble sur l'organisation d'une épreuve. Graphes et arbres sont complémentaires. À côté d'autres méthodes telles que les analyses de corrélations, la méthode implicative a la particularité de mettre en évidence des dépendances non symétriques entre variables, ce qui, comme nous le verrons, est particulièrement bien adapté à l'étude d'un test (Bodin, 1997; Couturier e Pazmiño, 2016; Gras, 1979; Regnier et al., 2011).

### 3. L'organisation de la recherche

#### 3.1 Objectifs de la recherche



La présente recherche est de type exploratoire plutôt que confirmatoire. Cela en relation aux potentialités de l'analyse implicative et aux contributions qu'elle peut offrir à l'amélioration des épreuves d'évaluation, dans notre cas à celles de type diagnostique de début d'année scolaire.

Nous avons alors cherché à répondre aux questions suivantes:

- a) Quelle information peut-on tirer d'une analyse de la complexité des questions?
- b) Quelles dépendances observe-t-on entre les questions du test?
- c) Quelles relations observe-t-on entre les habilités et dans quelle mesure le classement proposé pour les habilités tient devant l'analyse?
- d) Quelles relations observe-t-on entre le ressenti des élèves et le niveau de complexité des exercices?

#### 3.2 L'échantillon

L'échantillon étudié est un échantillon de convenance (Lucisano et Salerni, 2002) tiré en 2008-2009 de l'ensemble des établissements de l'enseignement moyen de la province de Bologne. Tout en étant conscient qu'un échantillon de ce type n'était pas nécessairement représentatif et qu'il limitait la possibilité de généraliser les résultats à la population cible par inférence statistique, cette procédure a été choisie par rapport aux objectifs et aux hypothèses d'une recherche plus large sur la valeur ajoutée<sup>1</sup>. Il est important de considérer que les choix des écoles ont été faits selon les critères suivants: a) mise en place dans des zones et des contextes différents de

1 La recherche évaluative en début d'année scolaire pour les élèves de première année de l'enseignement secondaire considérée ici fait partie d'une recherche plus large sur la valeur ajoutée en matière de compréhension de texte et de mathématiques en tant que mesure de l'efficacité de l'enseignement; elle a été menée sur une période de trois années scolaires (2008-09, 2009-2010, 2010-2011) sur les mêmes élèves en collaboration avec les 21 établissements scolaires, sous la coordination de M.L. Giovannini. Un indicateur du niveau d'implication des écoles est leur totale coopération pendant trois années complètes et le fait qu'aucune école ne se soit retirée de l'enquête.

sorte que le bassin de recrutement des établissements soit aussi hétérogène que possible du point de vue des caractéristiques socio-économiques et culturelles des étudiants; b) la disponibilité d'au moins trois classes de première année secondaire, pour permettre une analyse de la variance des indicateurs de performance non seulement au niveau de l'école, mais aussi au niveau de la classe. Après avoir réparti toutes les écoles de la province de Bologne<sup>2</sup> en fonction de leur situation géographique, on a sélectionné celles qui avaient au moins trois sections. Là où il y avait, dans la même région, plus d'une école répondant à ce critère, on a procédé par tirage au sort ; on a fait de même dans les écoles où il y avait plus de trois classes du niveau considéré.

L'échantillon est finalement composé de 21 collèges, 64 classes et 1344 élèves. L'analyse présentée dans cet article porte sur 1219 élèves. Quelques élèves étaient en effet absents le jour des épreuves et d'autre part les élèves ayant des difficultés particulières d'apprentissage ainsi que les étrangers maîtrisant mal la langue italienne ont été sortis de l'échantillon. Ces élèves ont été soumis à des épreuves simplifiées préparées à leur intention. Compte tenu du nombre réduit d'élèves concernés, il n'a pas été possible de les inclure dans la présente étude.

### 3.3 Le test de mathématique

Le test de mathématique dont nous analysons ici les résultats est divisé en deux cahiers qui reprennent une partie des items utilisés dans une enquête menée quelques années plus tôt en fin de scolarité élémentaire dans 13 écoles de la province de Bologne (Giovannini et Tordi, 2009).

Le test est divisé en deux fascicules (arithmétique; géométrie, probabilité et statistiques) dont chacun, à la fin comporte un questionnaire relatif à la perception que les élèves ont eu du test ; un questionnaire plus général destiné aux élèves a aussi été utilisé, mais cette contribution n'est pas utilisée pour cet article.

Le test a été passé pendant la première semaine d'octobre 2008<sup>3</sup>. L'administration a été conduite par des jeunes chercheurs de l'équipe de recherche, formés de manière à préserver l'homogénéité de la passation. Chacun(e) d'entre eux a dû tenir un journal pour noter toutes les questions des élèves. La répartition du temps alloué à l'administration du test a été le suivant: a) 10 minutes pour la présentation des objectifs de la recherche évaluative, b) 50 minutes pour la passation de l'épreuve d'arithmétique c) 10 minutes pour la pause, d) 50 minutes pour la passation de l'épreuve de géométrie probabilités et statistique, e) 10 minutes pour la passation du questionnaire élève.

Indépendamment des objectifs propres d'une recherche en milieu scolaire, nous pensons que les instruments mis au point par les chercheurs et les données recueillies devaient avoir une utilité pédagogique et didactique pour les enseignants et les écoles concernés: ainsi, chaque enseignant a eu connaissance des ré-

- 2 Dans la province considérée, les niveaux primaire et secondaire moyen (le collège français) sont le plus souvent organisés dans les mêmes établissements (*istituti comprensivi*, que nous appelons ici écoles).
- 3 Un essai avait été effectué en Septembre 2008 dans deux écoles des provinces voisines (Modena et Forlì-Cesena) pour vérifier les caractéristiques métrologiques (indices de difficulté et de discrimination) et les temps nécessaires à l'administration des items.



sultats de ses élèves ainsi que de la synthèse des résultats de l'ensemble de l'échantillon. En outre, les chercheurs en associant les enseignants, ont conduit des discussions avec les élèves après la passation des épreuves de façon à faire ressortir les raisons qui les avaient conduits à répondre de telle ou telle façon. Des conseils leur ont été donnés pour qu'ils puissent mener ces discussions sur le mode d'entretien en miroir dans lesquels l'enseignant n'intervient pas et où le conflit socio-cognitif entre élèves peut s'exprimer pleinement.

S'agissant d'un test diagnostique et formatif de début d'année, les questions portent sur les habilités de base considérées comme essentielles pour affronter le curriculum mathématique de la première année de l'enseignement secondaire (cf. Tab. 1).

Le terme habilité désigne ici « la capacité d'appliquer les connaissances et d'utiliser les savoir-faire (know-how) pour mener à terme des tâches et pour résoudre des problèmes ; les habilités sont décrites comme cognitives (usage de la pensée logique, intuitive et créative) et pratiques (qui impliquent l'habilité manuelle et l'utilisation de méthodes, de matériel et d'instruments) » (Pellerey, 2010, p. 200; traduction personnelle). Notons que les tests de notre étude se limitent aux habilités cognitives.

Les concepteurs du test souhaitent faire le point sur les habilités mises en jeu par l'enquête. Plus précisément, ils ont défini des habilités a priori, en référence au guide officiel italien et ont construit leur test autour de ces habilités.

Pour bien situer l'étude, précisons ici les intitulés des habilités visées, le nombre d'items qui sont attachés à chacune d'elle dans l'enquête et le taux moyen de réussite observé de chacune d'elles (Tab. 1).



		Nombre d'items	Score moyen
Habilité 1	Écriture des nombres entiers	21	75%
Habilité 2	Compétence calculatoire et reconnaissance des relations entre les nombres	36	67%
Habilité 3	Vitesse de calcul	14	49%
Habilité 4	Résolution de problèmes	6	45%
Habilité 5	Calcul de fractions, fractions décimales, équivalence de fractions	19	62%
Habilité 6	Connaissance des figures géométriques	34	56%
Habilité 7	Calcul de longueurs, périmètres et aires	10	30%
Habilité 8	Évaluation de probabilités et statistique	19	73%
Habilité 9	Capacités logiques et ensembles	8	57%

**Tab. 1 : Les neuf habilités**

### 3.4. *Choix de l'analyse implicative et son organisation*

Dans notre cas, nous avons privilégié l'analyse statistique implicative et cohésive (ASI) de façon à pouvoir préciser les relations observées d'un point de vue statistique entre les items et entre les questions du test, entre leurs niveaux de complexité et de difficulté des questions, ainsi qu'entre les habilités évaluées. Nous voulions aussi étudier les relations que les sujets qui ont répondu au test avaient entretenues avec les questions posées.

Avant d'aller plus loin il est nécessaire de préciser rapidement quelques aspects

théoriques. Pour cela, considérons par exemple un couple (A, B) de questions d'un test cognitif passé par une population P de sujets et supposons que la question A soit moins bien réussie que la question B ; ce que nous notons  $p(A) < p(B)$ . On peut se demander si les sujets qui ont réussi A ont aussi réussi B. Si cela est le cas, la réussite à A implique la réussite à B *au sens de la logique classique*. Si maintenant l'on considère  $n$  questions ( $A_1, A_2, \dots, A_n$ ) rangées dans l'ordre de réussite croissante et si, pour toute valeur de  $i$  telle que  $1 < i < n-1$ , la réussite à  $A_i$  implique la réussite à  $A_{i+1}$ , nous avons affaire à une chaîne de Guttman. Dans ce cas, les relations entre les questions apparaissent de façon évidente. Cependant, chacun sait que, dans la pratique, ce cas est très exceptionnel et, en général, la chaîne ( $A_1, A_2, \dots, A_n$ ) s'éloigne plus ou moins d'une chaîne de Guttman.

L'ASI est une méthode qui permet de repérer dans un test des chaînes de questions qui ne s'éloignent pas trop d'une chaîne de Guttman. La théorie de l'ASI, de nature probabiliste, permet de mesurer et de prendre en compte cet éloignement. Plus précisément, en nous limitant à deux événements A et B, par exemple la réussite à deux items A et B (cf. Fig. 1), et alors que nous visons l'inclusion de A dans B, c'est à dire que les élèves qui réussissent l'item A réussissent aussi l'item B, on se demande dans quelle mesure, sous l'hypothèse d'indépendance de ces deux événements, ce qui est observé en termes de réalisation conjointe de A et de non B peut ou non être le fruit du hasard.

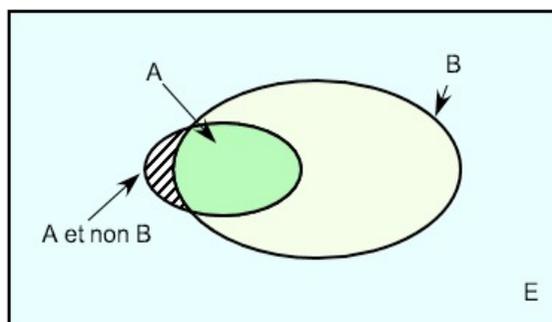


Fig. 1: Schématisation des événements A et B

La probabilité d'avoir A et non B dans ces conditions définit l'indice d'implication de A vers B. Un indice d'implication de A vers B supérieur à 0,95 signifiera alors que la probabilité d'obtenir aussi peu d'éléments de A qui ne soient pas dans B est inférieur à 0,05 ( $1 - 0,95$ ).

L'un des intérêts de la méthode réside dans son caractère non symétrique : à un seuil d'implication donné (défini par la théorie), si A implique B, en général B n'implique pas A. Un autre intérêt réside dans le fait que la méthode permet d'organiser les questions de façon à obtenir des cartes qui sont des mises à plats des questions et de leurs relations, cartes qui reflètent la structure cognitive des sujets par rapport au questionnement.

L'ASI est associée à une taxonomie de la complexité (taxonomie Gras-Bodin), issue de la taxonomie de Bloom mais adaptée aux mathématiques et validée, pour les mathématiques, par de nombreuses études empiriques (Gras, 1979).

Précisons simplement les intitulés des niveaux de complexité: niveau A: Connaissance et reconnaissance; Niveau B: Compréhension; Niveau C: Application; Niveau D: Créativité; Niveau E: Jugement. Par définition, la taxonomie est

hiérarchisée, c'est à dire que l'on s'attend à ce qu'une question de niveau A soit mieux réussie qu'une question de niveau C. En effet classer une question au niveau C suppose qu'elle met en jeu des connaissances et suppose la compréhension de ces connaissances et de leurs conditions d'utilisation. Cette propriété doit cependant être nuancée. En effet, elle est d'autant plus facilement vérifiée que les tâches concernées relèvent d'un même champ conceptuel (Vergnaud, 1994). Dans la présente étude, nous utilisons la taxonomie pour analyser l'équilibre du questionnement par rapport aux différents niveaux de complexité des tâches.

Par exemple, le graphe suivant (Fig. 2), issu de notre recherche, montre les liaisons implicatives entre six questions d'arithmétique. Les questions sont rangées de la gauche vers la droite par ordre de difficulté croissante (taux de réussite allant de 62% à 92%). Les indices d'implication sont tous supérieurs à .99.

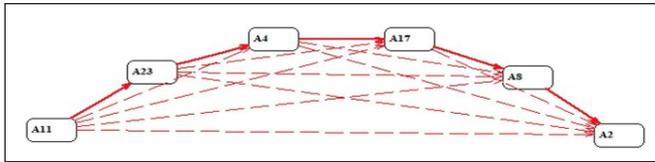


Fig. 2: Une chaîne implicative

Le coefficient d'implication est complété par un indice de confiance de l'inclusion de A dans B qui, dans le cas de variables binaires correspond à l'estimation de la probabilité conditionnelle de B sachant A (probabilité que nous écrivons en pourcentage). Un indice d'implication élevé peut en effet être accompagné d'un indice de confiance faible, ce qui serait de moindre intérêt dans le cas de questions d'évaluation. L'ajout de cet indice est particulièrement bien adapté aux enquêtes à grande échelle.

Constater que, d'un point de vue probabiliste la réussite à une tâche A s'accompagne d'une tendance à la réussite à une tâche B plutôt que de l'échec à cette même tâche est certes utile, mais constater que de plus, du moins dans la population observée, cela est vérifié par 80 ou 90% des élèves sera de nature à fournir des clés utiles à l'interprétation. Dans le cas ci-dessus, tous les indices de confiance sont supérieurs à 80%.

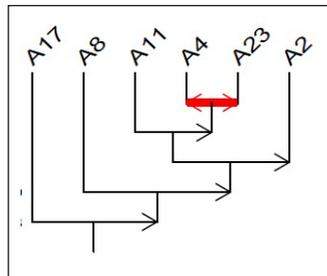


Fig. 3: Arbre cohésif des 6 questions

Dans l'arbre cohésif (Fig. 3) les regroupements se font à partir d'un indice d'implication entre groupes de questions construit à partir de l'indice d'implication entre items. Ici par exemple, A11 implique la classe (A4, A23) formée de questions équivalentes d'un point de vue de l'ASI et la classe (A11, A4, A23) implique la question A2.

## 4. Résultats de l'analyse implicative

Les données de notre étude portent sur 1219 élèves. Ces élèves ont répondu à 167 items du domaine cognitif (arithmétique, géométrie, probabilités et statistiques) et à 18 items portant sur le test (perception de la difficulté des questions, etc.).

Le test complet est constitué de 47 questions comportant chacune plusieurs items. Le codage est fait par item ce qui permet d'avoir des scores par question et des scores par item. Nous avons mené l'analyse au niveau des items et au niveau des questions.

### 4.1 L'analyse de la complexité

Parmi les qualités attendues d'un test figure évidemment la qualité du recouvrement des contenus que l'on veut évaluer. Dans le cas du test étudié ici, ce recouvrement est vérifié par sa construction même. Figure aussi un niveau d'appel diversifié aux différentes fonctions cognitives : remémoration, compréhension, etc. Niveaux qui sont synthétisés dans notre taxonomie de la complexité. Un test dont les questions se limiteraient au niveau A n'aurait en effet pas la même valeur diagnostique qu'un test dont les questions se répartiraient uniformément sur les 5 niveaux de la taxonomie.

Nous commençons donc par étudier le test de ce point de vue. Le tableau (Tab. 2) résume à la fois la distribution des items selon les niveaux et les domaines et les scores moyens obtenus pour chacune des catégories.



Domaine	Arithmétique				Géométrie				Statistiques et probabilités		
	A	B	C	D	A	B	C	D	A	B	C
Niveau de complexité											
Nombre d'items	30	32	27	7	26	10	8	5	6	15	2
Pourcentage d'items	31%	33%	28%	7%	53%	20%	16%	10%	26%	65%	9%
Score moyen	0,73	0,73	0,52	0,39	0,63	0,60	0,26	0,26	0,84	0,70	0,41

Tab. 2 : Distribution des item, des niveaux de complexité et des scores moyens

On observe que les items se répartissent sur les quatre niveaux A, B, C, D de la taxonomie. Les items se concentrent sur les niveaux A, B et C de la taxonomie et il n'y a pas d'item de niveau E (créativité), ce qui est assez normal pour un test portant sur les acquisitions de base.

### 4.2 L'analyse implicative et cohesive du test et ses compléments en termes de confiance

Le nombre d'items (167) est trop important pour que le test puisse être analysé de façon globale à ce niveau. Le nombre de questions (47) est lui même trop important pour que l'on puisse présenter le graphe complet. Cependant l'étude de ce graphe sur écran a contribué à orienter notre recherche. En plus du graphe, nous avons travaillé sur les matrices d'implication et de confiance qui elles aussi mettent en évidence des pistes qu'il s'agit ensuite d'explorer.

À partir de ces matrices, nous avons étudié en particulier la densité d'implications du test. Nous utilisons pour cela l'indicateur de densité d'implication au seuil  $\lambda$  d'un ensemble de  $n$  variables comme étant le rapport exprimé en pourcentage entre le nombre d'implications observées au seuil  $\lambda$  et le nombre de couples  $n \times n$ , qui correspond au maximum du nombre d'implications possibles:

$$d(n, \lambda) = \frac{\text{Nombre d'implications supérieures au seuil } \lambda}{n \times n} \times 100$$

Selon le cas étudié, cet indicateur peut prendre une valeur très faible (i.e. 10%) ou très forte (i.e. 60%). Une valeur faible correspond au cas où les questions sont peu liées les une autres. En termes de corrélation, on dira que les questions sont peu corrélées entre elles; cas qui peut être souhaité pour un test de nature sommative. À l'opposé, une valeur forte correspond au cas où les questions sont fortement corrélées entre elles; cas qui peut être préférable pour un test de nature diagnostique. Pour un test de nature sommative ou pour un examen, une valeur élevée mettrait en évidence des redondances dans le questionnement et une valeur informative faible du test.

Le tableau 3 ci-dessous présente les densités obtenues pour les deux parties du test objet de la présente étude:



Indicateur de densité	Partie arithmétique	Partie géométrie et statistique
Seuil $\lambda = 99\%$	60%	42%
Seuil $\lambda = 95\%$	75%	57%

**Tab. 3 : Distribution des intensités d'implication**

Ces valeurs sont très élevées<sup>4</sup>, en particulier pour l'arithmétique, ce qui, nous l'avons dit, signale une forte redondance d'un questionnement dont l'économie pourrait être améliorée. Cependant s'agissant d'un test d'entrée à visée diagnostique, il est possible de porter une telle redondance au bénéfice du test. D'un part, comme le montre le tableau 1, le test est facile, ce qui est souhaitable pour un test diagnostique, mais ce qui favorise une densité d'implications forte. D'autre part, en termes de diagnostique, la redondance permet à l'enseignant de mieux repérer les difficultés de chaque élève, sans confondre une erreur accidentelle (étourderie, fatigue, etc.) avec une erreur de nature conceptuelle, qu'il conviendra de traiter.

De son côté, l'arbre cohésitif des 47 questions met en évidence une cohésion forte de l'ensemble des questions. Les flèches indiquent le sens des implications et l'ensemble peut être interprété comme un flux implicatif s'écoulant de la gauche vers la droite.

4 Signalons que pour la littéracie mathématique de PISA 2012 nous avons trouvé une densité d'implication inférieure à 0,30 pour la France, l'Allemagne et le Japon tandis que cette densité s'élève à 75% dans le cas de l'Italie (seuil  $\lambda = 99\%$ ). L'étude en cours se poursuit avec d'autres pays mais nous faisons l'hypothèse que ces différences pourraient révéler des différences de nature curriculaire incluant des différences de pratique didactiques.

Précisons que si la cohésion était moins forte, l'arbre devrait se séparer en plusieurs branches non reliées entre elles.

Parmi ces 47 questions il y a des questions d'arithmétique, des questions de géométrie et des questions de statistiques et de probabilités. L'arbre (Fig. 4) met en évidence le fait que ces questions ne sont pas indépendantes. Leur analyse montre en effet que des questions de géométrie sont en fait des questions de mesure qui impliquent des calculs tandis que les questions de statistique appellent aussi des calculs.

Trois groupes se distinguent qui sans être totalement indépendants méritent d'être analysés séparément. L'analyse des questions nous conduit au classement suivant:

- Groupe A: exercices classiques: application directe d'une connaissance sans adaptation nécessaire
- Groupe B: exercices moins classiques pour lesquels une adaptation de la connaissance est nécessaire.
- Groupe C: exercices moins classiques nécessitant une adaptation moins évidente de la connaissance, avec éventuellement un changement de cadre.

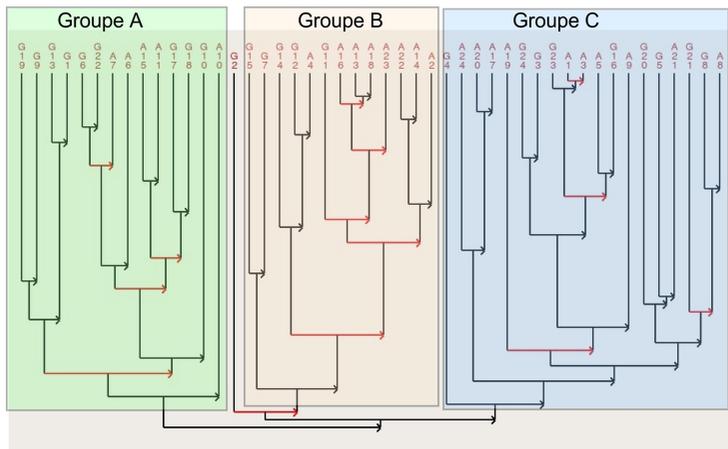


Fig. 4 : Arbre cohésif des 47 questions du test

#### 4.3 Principaux résultats sur les relations entre le ressenti des élèves et leur niveau de réussite

Une des particularités de l'enquête réside dans les questionnaires de ressenti accompagnant le questionnement cognitif. Le tableau (Tab. 4) présente les questions posées et les catégories utilisées pour l'analyse (certaines catégories ont été regroupées).

	Catégorie	1	2	3
D1	Comment as-tu trouvé cette épreuve ?	Très ou assez facile	Ni facile, ni difficile	Très ou assez difficile
D2	Comment penses-tu avoir réussi cette épreuve ?	Très ou assez bien	« <i>cosi cosi</i> »	Mal ou assez mal
D3	À ton avis, cette épreuve t'a-t-elle permis de faire émerger tes compétences en mathématiques ?	Beaucoup	Assez	Peu ou très peu
D4	Dans cette épreuve, y a-t-il des exercices que tu n'avais jamais rencontrés avant aujourd'hui ?	OUI	NON	

Tab. 4 : Questions relative au ressenti des élèves

Le même questionnaire a été passé après chacune des deux épreuves (A: arithmétique; G: Géométrie, probabilités et statistiques). Dans la suite de l'article les variables DAX\_Y se réfèrent à ce tableau. Par exemple DA3\_3 doit se lire: «*l'épreuve d'arithmétique m'a 'peu ou très peu' permis de faire émerger mes habilités en mathématiques*».

Compte tenu de la place dont nous disposons, nous ne traiterons ici que de l'épreuve d'arithmétique. Le même traitement a été fait pour l'épreuve de géométrie, statistique et probabilité.

Les scores de réussite à cette épreuve (scores moyens de réussite à l'ensemble des items) ont été répartis en 5 catégories selon le tableau 5.

	Score_A1	Score_A2	Score_A3	Score_A4	Score_A5
Intervalle	$[0 ; \mu - 1,25\sigma]$	$[\mu - 1,25\sigma ; \mu - 0,25\sigma]$	$[\mu - 0,25\sigma ; \mu + 0,25\sigma]$	$[\mu + 0,25\sigma ; \mu + 1,25\sigma]$	$[\mu + 1,25\sigma ; 1]$
Scores en %	0 à 41%	41% à 60%	60% à 69%	69% à 88%	88% à 100%
% d'élèves	14%	23%	19%	36%	9%

Tab. 5 : Répartition des scores de l'épreuve arithmétique

La répartition des scores en 5 classes a été choisie pour tenir compte du caractère très dissymétrique de leur distribution. Dans le tableau,  $\mu$  est la moyenne (64%) et  $\sigma$  est l'écart-type (16%).



Fig. 5 : Relations implicatives entre ressenti et scores – épreuve arithmétique (voir Tab. 4 la signification des variables DAX\_Y)

L'analyse implicative (graphe avec indice d'implication supérieur à .99 et indice de confiance supérieur à .50) montre des implications importantes entre le ressenti et les scores (Fig. 5). Le graphe est nettement partagé en deux sous graphes: l'un qui lie les variables de ressenti positif (épreuve facile ou assez facile, composée d'exercices de types déjà rencontrés, que les élèves estiment avoir bien ou assez bien réussie, leur ayant permis de mettre en valeur leurs compétences) et les scores les plus élevés (scores A5 et A4) ; l'autre lie les ressentis négatifs aux scores moyens ou faibles.

Ainsi, les élèves qui déclarent avoir rencontré ce type d'exercice au cours de leur scolarité primaire (DA4\_2) sont associés à un ressenti positif et à un haut niveau de réussite (graphe de gauche), tandis que les élèves qui déclarent ne pas avoir rencontré certains types d'exercices au cours de leur scolarité primaire sont associés à un ressenti moyen ou négatif et ont des résultats faibles.

L'analyse implicative permet de distinguer certaines variables en les plaçant en variables supplémentaires. De telles variables ne sont pas prises en compte pour la construction des arbres ou des graphes mais on a accès à la contribution que chacune d'elles apporte à la constitution de telle ou telle chaîne de variables. Cela permet aussi de distinguer les groupes d'élèves qui contribuent le plus à la construction du graphe de ceux qui s'en éloignent plus ou moins fortement et qui pourraient avoir besoin d'un traitement particulier.

Dans le cas présent, nous avons mis les variables de ressenti de l'épreuve d'arithmétique en variables supplémentaire et étudié la contribution de ces variables à la constitution du chemin présenté en 3.4, lequel est le plus long de cette épreuve : chemin de 9 questions dont les taux de réussite s'étagent de 60% à 92%. Les variables DA1\_1 et DA2\_1 contribuent avec un risque nul à la construction de ce chemin ; il s'agit là d'à peu près la moitié des élèves qui, à la fois, ont une attitude positive par rapport au test et ont réussi l'ensemble des questions de cette chaîne.

Une autre possibilité de l'ASI est la mise en évidence de la façon dont un ensemble de questions se situent par rapport à une question donnée: c'est la visualisation en mode cône du graphe (cf. Fig. 6). Compte tenu de sa position centrale dans les résultats du test, la question A7 a été choisie comme pivot. Les questions qui apparaissent dans la partie supérieure du graphe sont moins bien réussies que A7 tandis que les questions qui apparaissent dans la partie inférieure sont mieux réussies que A7. Il s'agit là d'un mode de présentation des données dont l'intérêt didactique est évident.

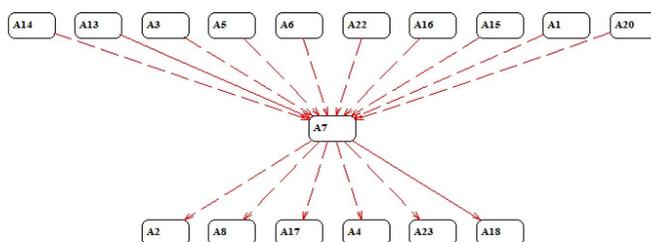


Fig. 6 : Graphe en mode cône des questions d'arithmétique

Les résultats obtenus pour les deux épreuves du test sont conformes aux attentes et de ce fait pourraient sembler peu intéressants. Pourtant ils mettent en

évidence la cohérence des résultats d'une façon qui reflète à la fois le sérieux avec lequel les élèves ont passé ces épreuves et la qualité de l'organisation de l'enquête. Ils sont aussi de nature à nous conforter dans la pertinence de l'analyse implicative pour analyser et donner du sens à ce type de données.

#### 4.4 Relations entre les exercices

Reprenons les 6 questions du questionnaire d'arithmétique utilisés dans le paragraphe 3.4 pour présenter dans la figure 7 la démarche et utilisons cette fois RCHIC (ASI sous R software) pour obtenir le graphe aux mêmes conditions (Indice d'implication .99 ; indice de confiance .80).

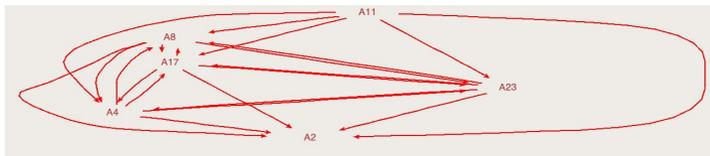


Fig. 7: Graphe implicatif des questions du paragraphe 3.4 (cf. Fig. 2)

Le logiciel RCHIC utilisé ici se comporte un peu différemment du logiciel CHIC utilisé jusqu'ici. Pour un couple (A, B) de variables, le logiciel CHIC ne présentait dans ses graphes, que l'implication d'indice le plus élevé dans la mesure où il était supérieur au seuil choisi (selon le cas, de A vers B ou de B vers A). Dans les mêmes conditions, RCHIC présente les deux implications lorsqu'elles sont toutes deux d'indice supérieur au seuil choisi.

Les mêmes liaisons implicatives apparaissent que dans 3.4, mais on voit aussi que les implications se font parfois dans les deux sens. Par exemple les élèves qui réussissent A4 ont une tendance forte à réussir A8, mais les élèves qui réussissent A8 ont aussi une tendance forte à réussir A4. Notons que la réussite va de A4 (78%) à A8 (88%).

Jusqu'à présent, dans cette étude, nous avons pris comme indicateur de réussite d'un exercice, la moyenne des réussites aux items le composant. Toutefois, le nombre d'items composant un exercice étant variable (de 14 items pour A1 à un item pour A8, par exemple), cet indicateur peut être trompeur quant à la qualité de la capacité globale mesurée par la question. Par exemple le score moyen de la question A1 est 49% tandis que le pourcentage de réussite conjointe aux 14 items n'est que 6%.

Pour une question donnée, les items la composant sont dans le même champ conceptuel (Vergnaud, 1994), mais de plus vérifient une même capacité, éventuellement à des niveaux de difficulté différents.

Voici un exemple:

Écris en chiffres, sur la ligne indiquée, le nombre formé de :
3 centaines et 20 unités _____
4 milliers et 3 dizaines _____
20 dizaines _____
3 milliers et 30 dizaines _____

Les taux de réussite des items de cette question sont respectivement 85%, 73%, 43%, 39%. La moyenne de ces réussites est 60%, mais le taux de réussite conjointe des 4 items n'est que 28%. Nous pouvons considérer que la capacité contrôlée par la question est maîtrisée par un élève s'il réussit tous les items composant cette question. Dans le cas étudié le taux de maîtrise est donc 28%.

Pour la suite, nous appelons indicateur de maîtrise d'une question le score conjoint aux items le composant.

Reprenons alors l'analyse implicite des 6 questions précédentes assorties de leurs indicateurs de maîtrise, nous obtenons alors le graphe suivant (cf. Fig. 8) où M signifie maîtrise. On voit que la maîtrise de A23 n'implique plus celle de A8 aux seuils d'implication et confiance précédents. Cela s'explique par le fait que, pour A23, la moyenne de réussite aux 4 items le composant est de 78% tandis que la réussite conjointe aux items (maîtrise) n'est plus que de 55%. L'exercice A8 ne comportant qu'un item, les valeurs des deux scores (score moyen et indicateur de maîtrise) sont donc égales (88%). L'ordre des réussites dans ce cas n'est pas modifié, mais cela n'est pas toujours le cas. Cette distinction entre score moyen et score (ou indicateur) de maîtrise est importante et peut conduire à des conclusions différentes, en particulier dans le cas d'un test diagnostique. Nous retrouverons cette question à propos des habiletés.

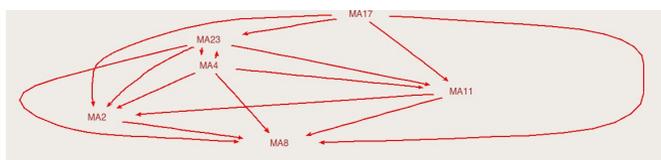


Fig. 8 : Graphe implicatif des maîtrises des questions de la figure 2

#### 4.5 Relations entre le ressenti des élèves et le niveau de complexité des exercices

Nous avons présenté en 4.1 l'analyse de la complexité des questions et la répartition des questions selon les niveaux de complexité. Il était donc intéressant d'étudier les relations entre les scores de réussite aux divers niveaux. Dans le graphe de la figure 9, Arith\_X est la variable score de réussite des questions de niveau X de l'épreuve d'arithmétique, GEOM\_X est le score de réussite des questions de niveau X de la seconde épreuve.

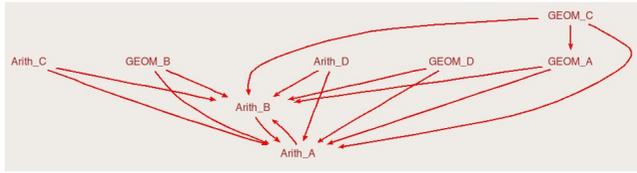
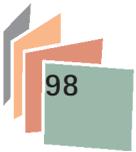


Fig. 9 : Relations implicatives entre scores selon les niveaux de complexité cognitive des questions

La réussite aux questions de géométrie de niveau C implique la réussite des questions de géométrie de niveau A, laquelle implique la réussite aux questions d'arithmétique de niveau A.

Il est intéressant de noter la convergence de toutes les variables de complexité des deux épreuves vers le niveau A de l'épreuve d'arithmétique. La réussite aux questions d'arithmétique de ce niveau, avec l'opérationnalisation qui en est donnée dans l'épreuve, apparaît ainsi comme une condition nécessaire à la réussite aux autres niveaux, et cela pour les deux épreuves.



#### 4.6 Relations entre les habilités

Ainsi que nous l'avons dit plus haut, l'enquête a été construite autour d'habilités définies a priori. Les 9 habilités ont été présentées dans le tableau 1 avec leurs taux de réussite respectifs.

Comme on le voit dans la figure 10, l'analyse implicative met en évidence des relations implicatives fortes entre plusieurs de ces habilités (relations orientées). Il est à noter que toutes les habilités convergent vers l'habilité 1, à savoir "Écriture des nombres entiers". Une analyse du contenu des questions qui opérationnalisent cette habilité permet de considérer qu'au delà de la simple écriture des nombres entiers il s'agit de la compréhension du nombre et de ses différents modes d'écriture et de la compréhension de la numération de position. Cette convergence remarquable permet, là encore, de penser que cette habilité serait une condition nécessaire à la maîtrise de toutes les autres.

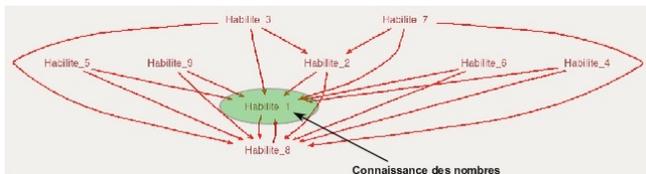


Fig. 10 : Graphe des implications liant les habilités

De même, toutes les habilités convergent vers l'habilité 8 (Évaluation de probabilité et statistique). Dans ce cas, ce qui n'était pas le cas pour la l'habilité 1, la facilité relative des questions et, sauf pour une question, leur limitation aux niveaux de complexité A et B, peut expliquer ce résultat. Il n'en reste pas moins que la maîtrise de cette habilité, au niveau des attentes manifestées par l'enquête apparaît aussi comme une condition nécessaire à la maîtrise des autres habilités.

Les habilités 3 et 7 (vitesse de calcul et calcul de longueurs, périmètres et aires)

impliquent en particulier l'habilité 2 (habilité calculatoire et reconnaissance des relations entre les nombres).

Précisons que, bien que pris en compte pour certaines analyses, les taux de réussite sont sans importance pour notre étude ; ce sont essentiellement les relations entre les niveaux de réussites qui nous intéressent. Il reste que ces relations pourraient être affectées par une opérationnalisation différente des habilités. Par exemple il aurait suffi de mettre plusieurs exercices relevant de l'habilité 8 aux niveaux de complexité C ou D pour que le sens de plusieurs flèches ait été inversé. Les conclusions que l'on peut tirer de l'analyse sont donc, pour une large part, le reflet des conceptions et des attentes des concepteurs des épreuves.

#### 4.7 Analyse des habilités

Par manque de place, nous ne développerons ici que l'analyse de l'habilité 1 : écriture des nombres entiers, nous limitant à indiquer quelques résultats relatifs à l'ensemble des 8 habilités.

L'habilité 1 est opérationnalisée par 24 items regroupés en 6 questions.

Le nombre relativement important d'items opérationnalisant cette habilité conduit à un graphe trop complexe pour être présenté ici. Disons simplement que les implications sont très fortes entre tous les items. Nous avons déjà signalé que cela était le signe d'une grande cohérence entre les items, de plus, le nombre important d'implications fortes dans les deux sens d'un item A à un item B indique que de nombreux items feraient double emploi s'il s'agissait d'un test d'évaluation. Ici le caractère exploratoire et diagnostique du test peut justifier ce fait et, ainsi que nous l'avons dit plus haut, l'information de la quasi équivalence de certaines questions n'est pas dépourvue d'intérêt.

Pour une part, ces équivalences pouvaient être prévues. C'est le cas d'items qui se suivent dans une même question et qui portent sur une même connaissance. Dans d'autre cas, l'analyse *a priori* ne laissait pas supposer une telle proximité. C'est le cas, par exemple, des items A.1.4.1 et A.1.9.3 qui sont tous deux réussis par 75% des élèves. Le premier concerne l'utilisation des symboles  $>$ ,  $<$ ,  $=$ , tandis que le second porte sur le sens des chiffres dans l'écriture des nombres entiers (écrire le plus grand nombre possible en utilisant une fois et une seule chacun des chiffres 5, 2, 6, 9) ; toutefois, de façon implicite, l'ordre des nombres est bien impliqué dans les deux cas.

Le graphe présenté dans la figure 11 porte sur les indicateurs de maîtrise des questions relevant de cette habilité ainsi que sur les points de vue des élèves par rapport à ces exercices. En effet, les questions concernant leur ressenti sont complétées par des questions leur demandant quel était, selon eux, la question, la plus facile, quelle était la plus difficile, ainsi que les questions qu'ils n'avaient jamais faites auparavant. Nous appelons point de vue les jugements des élèves par rapport aux questions. Ainsi, la variable DA5\_8 qui apparaît dans la figure 11 correspond aux élèves qui ont désigné la question A8 comme étant la plus facile. On ne s'étonne pas de voir cette variable implique fortement la réussite à la question A8. Par contre, on s'étonne davantage de constater que les autres variables du même type impliquent peu, voire très peu, la réussite aux questions correspondantes. Les élèves ne seraient peut-être pas toujours bons juges du niveau de difficulté des questions.



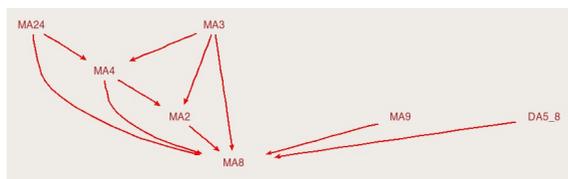


Fig. 11 : Graph des relations entre la maîtrise des question de l'habilité 1 et le point de vue des élèves

La chaîne MA24 (complexité niveau D), MA4 (niveau B), MA2 (niveau B), MA8 (niveau B), qui apparaît dans le graphe, ne manque pas d'attirer l'attention. Dans cet ordre, les scores de réussite sont respectivement de 28%, 53%, 68%, 87%. L'item A24 est un exercice portant l'utilisation du boulier chinois pour effectuer des calculs ; les items A4 et A2 portent sur le classement des nombres entiers, A8 sur l'écriture de position. L'analyse implicative fournit ici un éclairage sur l'organisation cognitive des élèves pris dans leur ensemble par rapport aux questions relatives à l'écriture des nombres.

L'arbre cohésitif est scindé ici en trois parties que l'on peut identifier comme relevant de sous-habilités relativement peu liées : l'une concerne l'écriture des nombres proprement dite, une autre concerne l'ordre dans l'ensemble des nombres, la troisième étant plus spécifique de la numération de position.

La même étude faite sur chacune des autres habilités confirme la forte cohérence interne de chacune des habilités. Toutefois, cela ne signifie pas que ces habilités soient indépendantes. Les questions de géométrie ou de statistiques supposent une certaine maîtrise de l'arithmétique, les capacités calculatoires de l'habilité 2 sont réinvesties dans l'habilité 5 relative aux fractions, la capacité à adapter une connaissance acquise à une situation en partie nouvelle se retrouve dans chacune des habilités, etc.

La présence de plusieurs parties non liées dans les arbres (non présentés ici) suggère que l'on pourrait aller plus loin dans l'analyse et suggérer des sous-habilités qui pourraient être entraînées séparément.

## 5. Conclusions

Dans cet article, nous avons utilisé l'analyse implicative et cohésitive pour tenter d'apporter des éléments de compréhension complémentaires aux données qui concernent 1219 élèves. Cette étude, menée en 64 classes de 21 collèges dans la province de Bologne, avait pour objectif de faire le point sur les acquis mathématiques des élèves à l'entrée du premier cycle de l'enseignement secondaire italien. Se plaçant loin des démarches d'évaluation traditionnelles, fussent-elles de nature formative, l'enquête avait pour ambition de favoriser l'expression des élèves en ce qui concernait le test, la difficulté et la facilité des questions, ainsi que d'organiser la discussion entre les élèves à son propos, pour permettre un repérage des conceptions erronées et des sources d'erreurs. Elle avait de ce fait un objectif de développement professionnel des enseignants.

L'analyse implicative n'est qu'une méthode parmi d'autres pour analyser les résultats d'un test. Elle a le mérite d'approcher l'organisation des connaissances des élèves, permettant ainsi aux enseignants de mieux situer et de mieux comprendre les difficultés des élèves. Elle peut de ce fait leur suggérer des pistes de réponse



pédagogique. Cela d'autant mieux que les analyses seraient étudiées et partagées lors d'actions de formation. Signalons que l'implantation de la démarche implicative dans un logiciel libre facilement maîtrisable (RCHIC sous R software) permet de faire les analyses en temps réel et de mettre sous les yeux des enseignants des graphes complexes qu'il n'est pas possible de reproduire dans un article du type de celui-ci, mais qui sont faciles à lire et à interpréter sur écran.

De plus, en complément des méthodes éducatives habituelles, l'analyse implicative peut contribuer au contrôle de qualité des tests : qualité et équilibre des questions, progressivité de la difficulté et de leur complexité. Ajoutons que dans le cadre de l'évaluation adaptative l'analyse implicative permet de mieux choisir les entrées et les parcours individualisés qui caractérisent cette forme d'évaluation (Bodin, 2010).

L'analyse implicative nous semble avoir apporté des pistes pour une meilleure compréhension du test lui-même, sur l'organisation de la connaissance des élèves, sur leur niveau de maîtrise, ainsi que sur les relations entre ces éléments et la façon dont les élèves considéraient le test et les questions qui le composaient.

Il reste à voir comment ces éléments pourront être utilisés pour la formation des enseignants mais notre expérience est que l'appui sur les éléments concrets tel qu'un test et ses résultats, traités par l'analyse implicative, est de nature à permettre aux enseignants de reconsidérer leurs conceptions tant sur l'évaluation que sur la didactique de la discipline.



## Bibliographie

- Bodin A. (1997). Modèles sous-jacents à l'analyse implicative et outils complémentaires. In A. Bodin, R. Gras, J. B. Lagrang (Eds.), *Implication statistique, prépublication* (pp. 97-32). IRMAR de Rennes.
- Bodin A. (2010). Vers un test adaptatif - critérié, combinant l'utilisation de l'IRT et de l'ASI, pour l'évaluation du socle commun de connaissances et de compétences. *5 Proceedings 5-7- November 2010. Quaderni di Ricerca in Didattica (Mathematics), n°20 suppl 1, 2010\_G.R.I.M. (Department of Mathematics, University of Palermo, Italy)\_A.S.I.*, 383-409.
- Couturier R., Pazmiño R. (2016). Use of Statistical Implicative Analysis in Complement of Item Analysis. *International Journal of Information and Education Technology*, 6(1), 39-43.
- De Luca A.M., Lucisano P. (2011). Item analisi tra modelli e realtà. *Giornale Italiano della Ricerca Educativa*, IV(7), 85-96.
- Fondazione Giovanni Agnelli (2011). *Rapporto sulla scuola in Italia 2011*. Bari: Laterza.
- Gattullo M. (1967). *Didattica e Docimologia. Misurazione e valutazione nella Scuola*. Roma: Armando.
- Gattullo M. (1978). *Voti, test, schede: ricerche sulla valutazione scolastica*. Firenze: La Nuova Italia.
- Gattullo M., Giovannini M.L. (Eds.) (1989). *Misurare e valutare l'apprendimento nella scuola media*. Milano: Bruno Mondadori.
- Giovannini M.L., Tordi C. (2009). Misura del valore aggiunto e miglioramento dell'insegnamento. Riflessioni da un'indagine empirica nelle scuole primarie bolognesi. In G. Domenici, R. Semeraro (Eds.), *Le nuove sfide della ricerca didattica tra saperi, comunità sociali e culture* (pp. 655-668). Roma: Monolite.
- Gras R. (1979). *Contribution à l'étude expérimentale et à l'analyse de certaines acquisitions cognitives et de certains objectifs didactiques en mathématiques*, Thèse d'Etat, Université de Rennes 1.
- Lucisano P., Salerno A. (2002). *Metodologia della ricerca in educazione e formazione*. Roma: Carocci.

- Pellerey M. (2010). *Competenze. Conoscenze, abilità, atteggiamenti. Il ruolo delle competenze nei processi educativi scolastici e formativi*. Roma: Tecnodid.
- Regnier J.C., Gras R., Spagnolo F., Di Paola B. (Eds.) (2011) *Analyse Statistique Implicative: Objet de recherche et de formation en analyse des données, outil pour la recherche multidisciplinaire. Prolongement des débats. QRDM Quaderni di Ricerca in Didattica - GRIM* ISSN on-line 1592-4424, Palerme: Université de Palerme.
- Vergnaud G. (1994). La théorie des champs conceptuels. *Recherches en Didactique des Mathématiques*, 10(2-3), 133-169.

