

# Les domaines de compétence de ALL et leur estimation

Présentation simplifiée des cadres de références et des principes de l'estimation des scores dans l'enquête internationale sur les compétences des adultes (ALL)



La série «Statistique de la Suisse»  
publiée par l'Office fédéral de la statistique (OFS)  
couvre les domaines suivants:

- 0** Bases statistiques et produits généraux
- 1** Population
- 2** Espace et environnement
- 3** Vie active et rémunération du travail
- 4** Economie nationale
- 5** Prix
- 6** Industrie et services
- 7** Agriculture et sylviculture
- 8** Energie
- 9** Construction et logement
- 10** Tourisme
- 11** Transports et communications
- 12** Monnaie, banques, assurances
- 13** Protection sociale
- 14** Santé
- 15** Education et science
- 16** Culture, société de l'information, sport
- 17** Politique
- 18** Administration et finances publiques
- 19** Criminalité et droit pénal
- 20** Situation économique et sociale de la population
- 21** Développement durable et disparités régionales et internationales

# Les domaines de compétence de ALL et leur estimation

Présentation simplifiée des cadres de références  
et des principes de l'estimation des scores  
dans l'enquête internationale sur les compétences  
des adultes (ALL)

**Adult Literacy  
& Lifeskills Survey**



**Auteur** Philippe Hertig  
**Office fédéral de la statistique**

**Editeur** Office fédéral de la statistique

**Editeur:** Office fédéral de la statistique (OFS)

**Complément d'information:** Jean-Christophe Zuchuat, division Santé, éducation et sciences  
E-mail: jean-christophe.zuchuat@bfs.admin.ch

**Rédaction et réalisation:** Philippe Hertig

**Diffusion:** Office fédéral de la statistique, CH-2010 Neuchâtel  
Tél. 032 713 60 60 / fax 032 713 60 61 / e-mail: order@bfs.admin.ch

**Numéro de commande:** 938-0800-05

**Prix:** 6 francs (TVA excl.), impression à la demande

**Série:** Statistique de la Suisse

**Domaine:** 15 Education et science

**Langue du texte original:** Français

**Page de couverture:** Rouge de Mars, Neuchâtel

**Graphisme/Layout:** OFS

**Copyright:** OFS, Neuchâtel 2008  
La reproduction est autorisée, sauf à des fins commerciales,  
si la source est mentionnée

**ISBN:** 978-3-303-15446-5

# Table des matières

<b>Introduction</b>	<b>5</b>	<b>B – L'évaluation des compétences</b>	<b>16</b>
<b>A – Les compétences et les tests de ALL</b>	<b>6</b>	<b>La théorie des réponses aux questions</b>	<b>16</b>
<b>Les objectifs</b>	<b>6</b>	– Généralités	16
<b>Les instruments: le questionnaire et les tests</b>	<b>7</b>	Une confirmation graphique idéale de la théorie	17
<b>Définition des compétences</b>	<b>8</b>	Une infirmation graphique de la théorie	18
– Généralités	8	De la théorie à la réalité	19
– La littératie	8	– Le modèle de Rasch et les modèles IRT	19
– La numératie	8	<b>L'estimation des paramètres</b>	<b>22</b>
– La résolution de problèmes	9	– Le maximum de vraisemblance	22
<b>Le fonctionnement des tests</b>	<b>10</b>	– Le procédé de recouvrement des questions	23
– La conception des tests	10	– Valeurs manquantes et imputations multiples	25
– Les facteurs de difficulté des tâches	11	<b>ANNEXE B</b>	<b>26</b>
Généralités	11	<b>Généralités sur l'imputation de valeurs aux données manquantes.</b>	<b>26</b>
Se faire une idée du problème	11	<b>Bibliographie</b>	<b>27</b>
Exécuter les opérations	12		
<b>ANNEXES A</b>	<b>14</b>		
<b>1 – L'approche de numératie</b>	<b>14</b>		
<b>2 – Les niveaux caractérisés de difficulté des tâches.</b>	<b>15</b>		



# Introduction

Qu'a-t-on mesuré dans ALL? Que signifient littératie et numératie? Comment mesure-t-on des compétences? Qu'est ce que l'«Item Response Theory»? Comment interpréter les scores de compétences et pourquoi, à chaque répondant, en a-t-on attribué cinq plutôt qu'un? Et comment se peut-il que l'on ait attribué des scores à des répondants pour des tests qu'ils n'avaient pas passés?

A ces questions deux ouvrages apportent des réponses<sup>1</sup>.

Statistique Canada a publié en 2005 un ouvrage exhaustif<sup>2</sup> sur les cadres de travail des domaines de compétence étudiés dans les travaux préparatoires de l'enquête. Cet ouvrage traite donc aussi bien les domaines qui ont été retenus pour la collecte des données que ceux qui ne l'ont pas été. Il s'agit là de l'ouvrage officiel de référence de l'enquête.

Il n'a pas été publié de manuel technique pour ALL parce qu'un tel manuel avait été publié pour IALS<sup>3</sup> et que cet ouvrage s'applique également à ALL. Le manuel technique de IALS constitue un ouvrage de référence très complet, de l'enquête.

Ces deux ouvrages sont copieux et le second est d'une approche difficile, réservée aux spécialistes.

La présente publication est destinée aux profanes, curieux du contenu de l'enquête «Adult Literacy and Lifeskills – ALL» et qui n'auraient pas le loisir de s'attacher aux deux ouvrages cités. D'une approche facile et brève il comporte un premier chapitre qui se veut une synthèse de la partie «*Theoretical Frameworks for Specific Domains Included in ALL*» de l'ouvrage «*Measuring Adult Literacy and Life Skills: New Frameworks for Assessment*». Ce premier chapitre décrit donc les domaines de mesure de l'enquête avec le fonctionnement de leurs tests. Le deuxième chapitre expose les rudiments des principes de psychométrie qui fondent le traitement des réponses aux tests, traitement aboutissant à l'attribution, à chaque personne soumise aux tests, de cinq scores de performance, (dites valeurs plausibles), dans les quatre domaines de compétence.

---

<sup>1</sup> Voir la bibliographie en fin d'ouvrage

<sup>2</sup> *Measuring Adult Literacy and Life Skills: New Frameworks for Assessment*.

<sup>3</sup> *Technical Report on the First International Adult Literacy Survey*

# A – Les compétences et les tests de ALL

## Les objectifs

Pour l'Organisation de Coopération et de Développement Economique (OCDE), la mesure des compétences des adultes est un objectif stratégique, l'auxiliaire d'une politique qui veut «**investir dans les compétences pour tous**» et associer progrès social et progrès technologique. Sachant que la prospérité d'une collectivité doit beaucoup aux connaissances, aux talents et aux savoir-faire de ses membres, il s'agit qu'en tous pays les gens sachent s'informer, se déterminer et agir avec à propos dans leur vie quotidienne, qu'ils sachent remplir avec bonheur des rôles multiples et évolutifs – étudiant, apprenti, travailleur, citoyen, parent, consommateur.

*Definition and Selection of Competencies* (DeSeCo, 1999–2003), une étude théorique multidisciplinaire sous l'égide de l'OCDE, a identifié un ensemble exhaustif de trois catégories de compétences clés: **agir de façon autonome, fonctionner dans des groupes hétérogènes, se servir d'outils de manière interactive**. Dans cette dernière catégorie, on relève l'actualité des compétences millénaires associées au langage, à l'écriture et au calcul et l'on reconnaît que leur besoin, loin de s'atténuer, s'accroît des exigences nouvelles nées de l'émergence des technologies électroniques de l'information et de la communication.

ALL adopte la perspective de DeSeCo qui réserve le terme de compétence à la capacité de réagir de manière appropriée à une sollicitation. Dans cette perspective, la compétence est déterminée par la nature de la sollicitation, sans considération de ses composantes éventuelles. C'est une variable intensive, un trait latent observable seulement par ses manifestations.

ALL a un précédent: l'«International Adult Literacy Survey» (IALS), la première enquête internationale jamais entreprise sur la littératie des adultes. La collecte des données de IALS s'est effectuée en trois vagues successives, entre 1994 et 1998, dans plus de vingt pays<sup>4</sup>. Cette enquête et d'autres travaux ont montré que la seule littératie, une compétence liée à la lecture, avait un pouvoir explicatif certes central mais limité dans l'analyse des facteurs favorisant une vie accomplie. Il importait donc de considérer d'autres compétences pour une explication plus complète. Ainsi s'est manifestée la volonté d'évaluer avec ALL, en plus de **la littératie**, les domaines de **la numératie**, une compétence liée au calcul, de **la résolution de problèmes analytiques**, une partie de la compétence de résolution de problèmes, de **l'aptitude à coopérer**, de **l'intelligence pratique** et de **la familiarité avec les outils des technologies de l'information et de la communication**<sup>5</sup>. Tous ces domaines avec leurs contours ont été définis et des tests s'y rapportant ont été construits. Les tests élaborés à propos des trois derniers domaines cités n'ont pas obtenus des résultats convaincants lors d'examens préliminaires et on a renoncé à mesurer **l'aptitude à coopérer** et **l'intelligence pratique**. Quant à **la familiarité avec les outils des technologies de l'information et de la communication**, l'intérêt pour ce domaine est tel qu'il a été conservé en se contentant d'interroger les gens sur leurs habitudes et leurs modes d'utilisation des équipements informatiques, en premier lieu de l'ordinateur.

Finalement l'enquête ALL mesure ce que les gens comprennent des écrits auxquels ils sont quotidiennement confrontés, comment ils se débrouillent avec les nécessités de manipuler tous les jours chiffres, quantités

<sup>4</sup> Allemagne, Australie, Belgique (Flandres), Canada Chili, Danemark, Etats-Unis, Finlande, Hongrie, Irlande, Italie, Norvège, Nouvelle-Zélande, Pays-Bas, Pologne, Portugal, République Tchèque, Royaume Uni, Slovénie, Suède, Suisse.

<sup>5</sup> En anglais: Literacy, Numeracy, Problem Solving, Teamwork, Practical Cognition, ICT Literacy



et grandeurs, comment ils s'orientent parmi les multiples possibilités d'aborder et de résoudre un problème<sup>6</sup> et, enfin, comment ils perçoivent et utilisent l'ordinateur et l'Internet. L'enquête apporte ainsi des éléments de réponse, des pistes de réflexion, à nombre de questions qui touchent à l'organisation de nos sociétés et à leurs défis politiques.

Dans sa conception puis dans sa légitimation, l'enquête a été subordonnée à quelques objectifs principaux.

Un premier objectif consiste à rendre compte du niveau des compétences de base choisies pour l'enquête, puis de voir comment ces compétences se distribuent et se comparent entre pays, entre aires linguistiques et culturelles. Il s'agit de voir aussi dans quelle mesure ces compétences sont singulières ou corrélées. Parce que l'enquête s'intéresse aux gens de 16 à 65 ans ses résultats renseignent sur les compétences à tous les âges de la vie active, ils témoignent donc de la durabilité des connaissances acquise à l'école et, parallèlement, des apprentissages de la vie active. Ce faisant, ils contribuent à doter d'une base empirique le choix des objectifs de l'éducation et de la formation continue.

Un deuxième objectif est de valider les deux domaines nouvellement introduits qui n'ont jamais été testés à grande échelle, la **numératie** et la **résolution de problème**.

Un troisième objectif n'intéresse que les pays qui ont participé à IALS et à ALL. Il concerne la possibilité de comparer les résultats de ces deux enquêtes grâce à l'exploitation des tests communs. Cette comparaison utilise deux perspectives: une perspective transversale, qui met en lumière l'évolution globale des performances de la population entre 1994 et 2003, et une perspective longitudinale qui observe l'évolution des performances d'une même cohorte<sup>7</sup>. La première perspective répond à la question: la population actuelle est-elle plus ou moins compétente que celle d'il y a neuf ans? Et la deuxième: avec l'âge, devenons nous plus ou moins compétents?

Les résultats publiés jusqu'ici montrent que ces trois objectifs ont été atteints.

*Le questionnaire sous-jacent de ALL:*

1. *Comment se compare le niveau de notre capital humain à ceux d'autres pays?*
2. *L'école enseigne-t-elle des savoirs utiles et durables?*
3. *Dans quelle mesure notre bien-être, individuel et collectif, dépend-il de nos compétences?*
4. *Y a-t-il parmi nous des illettrés? Qui sont ils et pourquoi?*
5. *Quels objectifs faut-il privilégier dans la formation des adultes?*
6. *L'écrit est-il une forme de communication efficace?*

## Les instruments: le questionnaire et les tests

Les données sont recueillies au cours d'interviews face à face. Chaque interview dure environ une heure et demie et se déroule généralement au domicile du répondant. Elle comporte deux parties : l'administration d'un **questionnaire** puis l'administration d'une batterie de **tests**. Le questionnaire est lu à haute voix par l'enquêteur qui y reporte les réponses du répondant. Puis un livret de tests est remis au répondant qui le lit et le remplit en présence de l'enquêteur, mais sans son concours<sup>8</sup>.

**Le questionnaire** relève les caractéristiques principales des répondants ainsi qu'une série de variables que l'on suppose en relation avec les performances aux tests; il porte donc sur les caractéristiques socio-démographiques des répondants et sur quelques aspects de leur existence au cours des douze derniers mois. A savoir:

1. Le suivi éventuel de cours, d'un ensemble de cours ou d'une formation continue, la pratique d'apprentissages sur le tas ou en autodidacte.
2. L'activité professionnelle, la profession, la branche d'activité et la position professionnelle.
3. Les modes et les fréquences d'utilisation des outils de l'information et de la communication (essentiellement l'ordinateur).
4. Les habitudes de lecture et de calcul, à la maison et au travail; l'inclination à la lecture, le partage des loisirs entre lecture et télévision; l'aisance ou la difficulté éprouvée en calcul, en lecture et écriture.
5. L'engagement social et associatif.
6. Les états subjectifs de santé et de bien être.

<sup>6</sup> Une compétence du domaine du raisonnement analytique dont il est fait ici une évaluation pour la première fois à cette échelle

<sup>7</sup> Les gens nés en même temps (dans un intervalle d'une ou de quelques années)

<sup>8</sup> Il est impératif que l'interviewer n'intervienne en aucune manière.

Les tests servent à la mesure des performances des répondants. Ceux-ci doivent examiner une série de stimuli et répondre, à leur propos, à quelques questions de difficulté variable. Les questions posées contiennent les **informations données** tandis que les stimuli contiennent les **informations demandées**. Les stimuli sont tous extraits ou inspirés de documents banals, – un article de presse, une information des autorités, un contrat de bail, un formulaire d'inscription, le mode d'emploi d'un appareil, un prospectus publicitaire, etc. –, qui peut emprunter toute forme de communication, – la description, la narration, l'argumentation, l'instruction, etc. –, et concerner tout aspect de la vie quotidienne, – la famille, les loisirs, la santé, la consommation, la sécurité, la vie politique et associative, le travail, etc. Ces aspects sont tous exploités, dans le but d'observer les compétences à l'œuvre dans leurs multiples modalités d'application.

## Définition des compétences

### Généralités

ALL considère que la compétence est une grandeur continue dont aucune valeur ne distinguerait, même arbitrairement, des gens compétents de gens incompétents. Dans les pays de l'OCDE, quasiment tout le monde a au moins appris des rudiments de lecture et de calcul et il s'agit, avec les tests de ALL, de rendre compte dans quelle mesure la population adulte maîtrise ces pratiques.

L'analyse de la totalité des réponses aux tests permet d'inférer à chaque question un degré de difficulté. Dans chaque domaine, l'ensemble de ces degrés définit une échelle de compétence. Les questions correctement répondues par une personne, ou une population, constituent (avec leur degré de difficulté) un score de performance qui situe cette personne, ou cette population, sur l'échelle de compétence. L'analyse de tous les scores repère sur les échelles des seuils caractérisés de difficulté qui délimitent des niveaux de compétence<sup>9</sup>: cinq en littératie et numératie, quatre en résolution de problème. Ce sont les **niveaux de compétence** que l'on retrouve tout au long des exposés des résultats de ALL. On estime que les gens confinés aux niveaux 1 et 2 ont des compétences critiques pour se mouvoir à l'aise dans nos sociétés techniciennes et démocratiques et l'on estime que tous ceux qui ont au moins suivi l'école obligatoire de bout en bout devraient atteindre le niveau 3.

<sup>9</sup> Les échelles sont arbitrairement graduées de 1 à 500

### La littératie

La communication écrite est massivement présente dans tous les aspects du quotidien et la compétence de déchiffrer, d'interpréter et d'utiliser efficacement l'écrit est nécessaire à tous. On appelle littératie cette compétence que composent diverses connaissances et divers savoir-faire. Au-delà du simple déchiffrement la littératie couvre donc les aspects de la compréhension d'un texte, de la capacité à le résumer, à le critiquer, à le référencer, à en tirer parti, etc. On l'exploite pour une grande variété de motifs et dans une grande variété de contextes qui, chacun, occasionne une manière de lire différente.

*La littératie est la capacité d'utiliser l'écrit pour fonctionner dans la société, atteindre ses objectifs, parfaire ses connaissances et accroître son potentiel.*

A fin d'investigation et d'analyse, on évalue séparément la **littératie de textes suivis** et la **littératie de textes schématiques**. Les textes dits suivis sont constitués d'un enchaînement de phrases, regroupées ou non en paragraphes, avec ou sans titres et sous-titres. Les textes dits schématiques, sont constitués d'un ou de plusieurs documents synoptiques<sup>10</sup> – listes, tableaux, formulaires, graphiques, cartes, plans, dessins, – accompagnés ou non de textes – commentaires, explications, instructions, etc. –. Les deux types de tests sont fortement corrélés mais les stratégies mises en oeuvre à leur égard ne sont pas identiques.

### La numératie

Face aux décisions quotidiennes, la vie met chacun de nous en devoir de calculer, évaluer, mesurer, comparer, pour faire des achats, construire un meuble, demander un prêt, etc. Avec la notion de numératie on tente de circonscrire les modes d'application opportunistes des connaissances et des savoir-faire de caractère mathématique propres aux adultes.

Chez les adultes, la pratique des mathématiques acquise à l'école est, selon l'âge et la biographie, plus ou moins ancienne et plus ou moins estompée. En revanche, exposés aux contraintes du réel et confrontés dans leur quotidien à une grande diversité de problèmes, les adultes développent souvent des méthodes de résolution

<sup>10</sup> Un document synoptique ... donne par sa disposition une vue générale d'un ensemble que l'on peut ainsi embrasser d'un seul coup d'œil. (In «Trésor de la langue française»). Plutôt que «Littératie de textes schématiques» il aurait été plus judicieux d'utiliser «Littératie de documents synoptiques».

originales. Ils apprennent par eux-mêmes comment faire le calcul d'un pourcentage ou celui d'un complément à un nombre lorsqu'il s'agit de rendre de l'argent par exemple. On comprend donc que la numératie va au-delà du savoir scolaire et de la maîtrise des opérations arithmétiques. Personnalisée par l'expérience, elle s'alimente de méthodes originales, de cheminements pragmatiques reliant des problèmes concrets à des connaissances abstraites. Partant d'une perspective fonctionnelle on a donné de la numératie la définition suivante:

***La numératie rassemble les connaissances et savoir faire nécessaires pour traiter avec pertinence les aspects mathématiques des problèmes de la vie courante.***

Idéalement l'évaluation de la numératie devrait donc nous apprendre comment et dans quelle mesure tout un chacun se débrouille pour résoudre les problèmes de calcul au quotidien, qu'il s'agisse d'atteindre un objectif personnel ou de répondre à une invite. Cependant, comme il n'est possible d'observer que des comportements, et non les connaissances et savoir-faire qui les sous-tendent, on s'est attaché à définir une «approche de numératie» que manifeste la compétence latente.

***On adopte une «approche de numératie» lorsqu'il faut résoudre un problème pratique de caractère mathématique. Cette approche suppose la compréhension de divers langages et concepts mathématiques, elle mobilise des connaissances idoines et met en action des savoir faire appropriés.***

### La résolution de problèmes<sup>11</sup>

Il y a problème lorsque quelque chose est à faire mais qu'on ne sait pas exactement quoi, ni comment procéder. On s'aperçoit dans ce genre de situation que les gens font preuve d'une efficacité variable. Cependant la mise en évidence d'une compétence singulière, qui rendrait compte de cette efficacité, est et reste délicate, voire hasardeuse. On a constaté, par exemple, que cette efficacité était, au moins en partie, redevable d'une certaine familiarité avec le contexte et le thème du problème. Dans l'enquête ALL, on a cherché à neutraliser cette source de biais possibles par le choix de thèmes issus de situations courantes et jouissant de la plus large familiarité, tout au moins au sein des pays de l'OCDE.

Dans le but de fonder et guider la mesure de la résolution de problèmes, on en a formulé la définition suivante:

***La résolution de problèmes met en œuvre une séquence d'actions réfléchies, orientée vers un but inaccessible par une démarche de routine. Celui qui doit résoudre le problème a un but plus ou moins défini mais il ne sait pas d'emblée comment l'atteindre. L'inadéquation des modes opératoires ordinaires au but poursuivi constitue le problème. La compréhension de la conjoncture du problème puis sa transformation par étapes dûment planifiées constituent le processus de résolution de problèmes.***

Les problèmes exposés dans le test sont communiqués aux répondants sous une forme écrite, donc abstraite, et leur description n'est qu'une schématisation d'une portion de réalité. Partant, on ne peut s'attendre à ce qu'ils induisent chez le répondant la mobilisation de toutes les facettes d'un comportement en situation réelle. De plus, la nécessité de la lecture implique une part de littératie difficile à séparer de la compétence mesurée. Cette dernière toutefois, dont l'essentiel est identifiée comme une compétence en résolution de problèmes analytiques, se situe au coeur même de la résolution de problèmes définie plus haut. Les tests ont été conçus sur le modèle de l'approche projet, une approche qui fait appel à un modèle d'action complète.

Les problèmes proposés aux répondants sous l'appellation de **projets** sont au nombre de quatre et bâtis autour des thèmes suivants: l'achat d'une bicyclette, la recherche d'un appartement, l'organisation de la visite d'un chœur étranger, l'organisation d'une fête sportive. Le projet est introduit sur la première page du livret par une brève description de ses conditions et du rôle virtuel que le répondant doit y tenir, ainsi que par une liste de quatre ou cinq étapes de réalisation à parcourir. Au verso, la première étape est décrite avec ses conditions ainsi qu'une liste d'actions possibles. La page en regard comporte une question amenant le répondant à opérer un ou plusieurs choix dans la liste des actions. ... et ainsi de suite jusqu'à l'épuisement des étapes.

<sup>11</sup> La résolution de problèmes dont il est question ici n'a rien à voir avec la résolution de conflits ou de problèmes émotifs.

**Le modèle d'action complète**

- |                           |  |
|---------------------------|--|
| 1. Définir les buts:      | Identifier les objectifs, les classer par ordre d'urgence et d'importance, choisir les buts et motiver les choix.  |
| 2. Analyser la situation: | Recenser les informations disponibles, acquérir et ordonner les informations manquantes, déceler et évaluer les conditions et les contraintes.                                   |
| 3. Planifier la solution: | Envisager les étapes, élaborer des plans sous forme de séquences ordonnées d'actions appropriées, coordonner les tâches et les délais, comparer les plans et en sélectionner un. |
| 4. Exécuter le plan:      | Exécuter les actions planifiées et intégrer leurs effets, contrôler les délais, prendre les mesures correctives si nécessaires.  |
| 5. Evaluer le résultat:   | Evaluer l'atteinte des objectifs, analyser les dysfonctionnements constatés et apprécier leurs conséquences, en tirer les enseignements.   |

Il va de soi que le contexte de l'enquête, à savoir la visite de l'enquêteur au domicile du répondant, limite le domaine à la résolution de problèmes analytiques. Les aspects de l'action complète dépendants des relations interpersonnelles (sociabilité, leadership, compétence de communication), de la gestion de la durée (continuité, persévérance, anticipation) et des imprévus (intégration des modifications, changements de cap) sont notamment absents de l'épreuve. C'est-à-dire que le test ne simule que les étapes 1 à 3 du modèle d'action et, très partiellement, l'étape 4.

**Le fonctionnement des tests****La conception des tests**

Le but de toute enquête est de relever des données probantes permettant de porter des jugements éclairés. Dans ce sens, les tests doivent fournir des résultats **pertinents, comparables et interprétables**. S'agissant de la mesure des compétences ils doivent:

- Révéler la vérité des faits et refléter la pluralité des populations et des typologies.
- Se comparer d'une enquête à l'autre, entre pays, cultures et langues différentes.

- Rendre compte des écarts entre individus et entre groupes d'individus de manière différenciée et explicable<sup>12</sup>.

Les tests doivent fournir un instrument de mesure fin et précis, sur l'ensemble de l'échelle de la performance. Ils doivent simuler la sollicitation de compétences, telle qu'on peut l'expérimenter dans toutes les sphères de l'existence: à la maison, dans la rue, au travail, à l'école. L'origine, les contextes, les thèmes et les formes d'expression des stimuli doivent refléter l'éventail des documents usuels des pays de l'enquête afin d'éviter toute discrimination culturelle. Pour réduire au minimum l'effet éventuel de connaissances préalables, ils ne doivent pas reproduire des documents spécialisés.

Pour construire des tests répondant au besoin d'évaluer les compétences dans leurs multiples manifestations et sur une échelle étendue de compétence, on s'est astreint à suivre les règles suivantes:

- Les tests doivent comporter une grande variété de tâches inspirées de situations que tout un chacun peut devoir affronter dans sa vie quotidienne.
- Les tâches doivent être plutôt brèves et nombreuses, afin de couvrir l'étendue des difficultés et les dimensions des domaines retenus pour l'enquête, tout en gardant à l'interview une durée convenable.
- Une partie des questions doit être très facile et cependant capable d'une fine discrimination dans la gamme des basses performances.
- Les questions ne doivent pas paraître farfelues, ni moralement ou politiquement incorrectes.
- Il faut minimiser le chevauchement des domaines et notamment prévoir des stimuli de numératie avec un minimum de texte.
- Les stimuli doivent être raisonnablement familiers à toutes les populations des pays de l'enquête, ne nécessiter des connaissances préalables que de faible niveau et très communément répandues.
- Les stimuli et les questions doivent pouvoir se traduire et s'adapter dans toutes les langues et tous les systèmes d'unités en usage dans les pays de l'enquête, tout en gardant les mêmes caractéristiques psychométriques.

<sup>12</sup> Classer ne suffit pas; il faut encore pouvoir mesurer et caractériser les écarts.

- Les stimuli de numératie doivent avoir un aspect rigoureusement identique dans toutes les versions et les chiffres qui y figurent doivent être toujours les mêmes<sup>13</sup>.

### Les facteurs de difficulté des tâches

#### Généralités

L'instrument qui mesure la compétence est constitué d'un ensemble de questions dont les réponses demandent l'accomplissement de tâches plus ou moins difficiles. La difficulté d'une tâche relève de sa transparence – **qu'est-ce qu'il faut faire?** – et de sa nature – **comment le faire?** La difficulté de se faire une idée de la tâche relève de la confrontation de la question au stimulus et des caractéristiques de ce dernier.

La difficulté globale de la tâche est ainsi la somme de difficultés partielles variables, dont les différentes combinaisons permettent l'élaboration d'une gamme de tâches de difficulté étendue, apte à rendre compte d'une grande variation de compétence avec un pouvoir précis de discrimination. La connaissance des variables de difficulté met en mesure de prévoir un classement des tâches. Cependant, la vraie difficulté des tâches se déterminera *a posteriori*, par l'analyse des résultats que nous verrons plus loin. Dans ce paragraphe, l'analyse des conditions des tests et de la nature des tâches permet une détermination *a priori* de leur difficulté approximative. Lorsque les résultats confirment les prévisions, les variables utilisées dans la construction des tests se voient validées et cette validation ouvre des perspectives sur l'interprétation du fonctionnement des compétences. Elle permet aussi d'enrichir le savoir dans le domaine de la construction des tests.

Rappelons qu'un test comporte un stimulus et quelques questions s'y rapportant. Dans la question se trouvent les **informations données**, qui président à l'examen du stimulus et conduisent à en extraire ou à y générer les **informations demandées** répondant à l'exigence de la question.

Afin de distinguer les variables qui contribuent à la difficulté des tâches, distinguons deux étapes dans le processus qui mène à l'élaboration de la réponse:

1. Se faire une idée du problème – Qu'est-ce qu'il faut faire?
2. Exécuter les opérations – Comment le faire?

Certaines variables sont communes aux trois domaines de test, d'autres sont spécifiques à l'un ou l'autre des domaines.

#### 1. Se faire une idée du problème

Comprendre le problème à l'aide des informations données dans la question et par une appréhension globale du stimulus constitue la première étape. La plupart des questions sont formulées de façon à poser un minimum de difficulté. Elles sont brèves et directes. Certaines cependant peuvent être plus longues, incluant soit des conditions qui rendent la tâche plus opaque soit des précisions qui la clarifient. Ces conditions et/ou précisions peuvent être plus ou moins explicites, mentionner ou taire par exemple le nombre des informations demandées ou des opérations à effectuer.

Dans tous les cas l'appréhension du stimulus passe par la maîtrise de sa complexité formelle, une étape que les caractéristiques du stimulus rendent plus ou moins ardue.

La mise en page, la quantité et la nature des informations sont les caractéristiques du stimulus. Celui-ci peut comporter des textes plus ou moins longs, des documents synoptiques<sup>14</sup> plus ou moins étendus et riches ou encore un mélange des deux. Ces caractéristiques sont en gros analogues dans les trois domaines de tests avec quelques particularités propres à l'un ou à l'autre.

Les variables relatives aux textes ont trait à l'architecture du texte et à son style. L'architecture désigne la longueur du texte et sa composition – explicite en paragraphes, avec titres et sous-titres, ou implicite. Le style désigne les aspects littéraires – phrases courtes ou longues, syntaxe simple ou compliquée, vocabulaire commun ou spécialisé – et les aspects typographiques – grosseur et type des caractères (uniques ou changeant pour mettre en évidence certaines informations).

<sup>13</sup> Ces deux conditions ne sont pas toujours conciliables avec la logique. Des stimuli de numératie présentent des objets avec un prix d'achat. A cause des différences de pouvoir d'achat entre devises il a fallu adapter les objets pour garder un même montant aux prix indiqués.

<sup>14</sup> On désigne ici par document synoptique une liste, un tableau, un graphique, un diagramme, un formulaire, un plan, une carte, une image etc. On trouve des documents synoptiques dans les stimuli de littératie de textes schématiques, de numératie et de résolution de problèmes. Un document synoptique ... donne par sa disposition une vue générale d'un ensemble que l'on peut ainsi embrasser d'un seul coup d'œil. (In «Trésor de la langue française»)

Les variables relatives aux documents synoptiques ont trait au type du document ainsi qu'à la nature – mots, nombres, formules, symboles, textures, etc. – et à l'abondance de son contenu, - nombre de catégories et d'éléments par catégorie, densité de l'information, complexité des documents graphiques –. Dans les types de documents synoptiques on recense les documents matriciels – liste simple, liste plurielle, tableau à deux dimensions, tableau à trois dimensions ou listes imbriquées –, les documents graphiques – diagrammes, schémas, images –, les documents cartographiques – plans et cartes – et les documents à remplir – formulaire, protocole, questionnaire.

En résolution de problèmes la tâche globale est introduite par une brève description du problème. Celle-ci induit la phase d'idéation du problème qui permet d'aborder la tâche. Dans la description, l'objectif du projet, les conditions à observer, les différentes étapes à ordonner peuvent être plus ou moins explicites, plus ou moins clairement exposés. A ceci s'ajoute la difficulté liée à l'ampleur du problème, soit le nombre des étapes et de leurs relations conditionnelles, le nombre des tâches et leurs difficultés propres.

Les variables relatives aux formats de réponse sont particulières à la résolution de problèmes et se rapportent aux trois formats possibles – réponse à choix multiple, réponse à choix dans un tableau à double entrée, réponse ouverte.

## 2. Exécuter les opérations

Cette étape aboutit à identifier ou générer l'information demandée, en éludant les distracteurs

La difficulté des opérations est influencée par

- le type d'information demandée,
- le type et le nombre d'opérations à effectuer,
- la présence et la plausibilité des distracteurs.

### Le type d'information demandée

La difficulté de repérer l'information demandée tient à la nature de l'objet concerné. Cette difficulté va croître avec le degré d'abstraction de l'objet. Un animal, un ustensile, un lieu sont faciles à repérer. Un objectif, une intention, une condition, une relation de cause à effet, sont des notions plus abstraites donc plus difficiles à cerner, et plus encore, par exemple, un jugement relatif.

*Les degrés de difficulté du type de l'information demandée*

1. Une personne, un animal, un lieu, une chose
2. Une quantité, une qualité, une durée, une action, une position
3. Un procédé, un objectif, une intention, un attribut
4. Une cause, un effet, un mécanisme, une similarité
5. Une équivalence, une différence, un thème.

### Le type d'opération en littératie

L'opération type de littératie est dite «d'appariement». Il s'agit du processus de rapprochement qui consiste à rechercher dans le stimulus l'information qui «s'apparie» avec l'information donnée dans la question. Il implique souvent un va et vient entre la question et le stimulus.

On distingue quatre types d'appariement<sup>15</sup> de difficulté croissante: l'identification, le regroupement, l'intégration et la formulation.

*L'identification* consiste à repérer une information qui est formulée soit dans les termes mêmes de la question soit à l'aide d'un ou de plusieurs synonymes.

*Le regroupement* consiste à repérer tous les éléments d'information de la catégorie formulée dans la question. Cette dernière peut spécifier ou taire le nombre d'éléments.

*L'intégration* consiste à repérer tous les éléments d'informations qui satisfont une relation donnée dans la question (par exemple: semblable, différent, contraire, plus petit, etc.).

*La formulation* consiste à générer la catégorie ou la relation qui caractérise l'information à repérer. Dans le cas le plus difficile la formulation se déduit de notions abstraites, voire conditionnelles, et peut, à l'extrême, nécessiter l'apport d'informations que le répondant doit générer lui-même.

### Le type d'opération en numératie

En numératie, outre un possible appariement, la tâche requiert une ou plusieurs opérations de type mathématique, des opérations que l'on peut classer en:

- élémentaires*, comme le dénombrement, le classement, la mesure d'une droite, un calcul élémentaire, la reconnaissance d'une forme géométrique simple
- assez simples*, comme la division, l'application d'une formule simple, la lecture d'un graphique
- assez difficiles*, comme le calcul avec des proportions, la lecture d'un graphique compliqué, l'interprétation de données statistiques

<sup>15</sup> En anglais «Type of match»

*difficiles*, comme la comparaisons de notions abstraites, la construction d'un algorithme, la déduction, la justification mathématique.

La difficulté de la tâche dépend en outre du nombre des opérations à effectuer ainsi que des formats des nombres. Ceux-ci peuvent être entiers ou fractionnaires, comprendre peu ou beaucoup de chiffres caractéristiques, être, dans le cas des pourcentages, très communs (10%, 25%, etc.) ou rares et moins directement éloquentes (23%, 61%, etc.).

#### **La présence et la plausibilité des distracteurs**

Les stimuli comprennent des distracteurs en nombre variable. Un distracteur est une information qui a l'apparence de la pertinence. Il répond à certains des critères de l'information demandée mais pas à tous. La **plausibilité** d'un distracteur désigne la proportion des caractéristiques qu'il partage avec l'information demandée. Un répondant a d'autant plus de chance de se laisser abuser par un distracteur que celui-ci est plus plausible, qu'il y en a un plus grand nombre et qu'ils sont plus proches de l'information demandée.

# ANNEXES A

## 1 – L'approche de numératie

**L'approche de numératie est la disposition mentale qui permet de:**

- 1) Résoudre un problème quantitatif issu de multiples contextes, comme:
  - *la vie privée (famille, loisirs)*
  - *la collectivité (association, travail bénévole, engagement politique)*
  - *la profession*
  - *l'apprentissage (études, formation continue)*
- 2) en reconnaissant des informations relatives à
  - *des grandeurs, des quantités et des nombres*
  - *des formes et des dimensions*
  - *des fonctions, des relations et des variations temporelles*
  - *des collections de données et des probabilités*
- 3) exprimées par
  - *des objets et des images d'objets*
  - *des nombres et des symboles*
  - *des formules et des équations*
  - *des diagrammes et des graphiques*
  - *des plans et des cartes*
  - *des tableaux*
  - *des textes*
- 4) amenant à enchaîner des opérations
  - *de repérage des données, (recherche des données)*
  - *de traitement des données pour obtenir des résultats, (dénombrement – ordonnancement – estimation-calcul – mensuration – modélisation)*
  - *de jugement des résultats et (cohérence – plausibilité – adéquation-pertinence)*
  - *de communication des résultats (documentation – rédaction – argumentation – justification)*
- 5) en mobilisant des connaissances, des savoir-faire et des modèles d'interprétation tels que
  - *des connaissances en mathématiques*
  - *des techniques de résolution de problèmes et des compétences de lecture*
  - *des présomptions et des points de vue.*



## 2 – Niveaux caractérisés de la difficulté des tâches

Niveaux et scores	Littérature de textes suivis et de textes schématiques	Numératie	Résolution de problèmes
1 0–225	Repérer une information unique, dans un texte suivi court ou dans un texte schématique simple.	Faire une opération arithmétique simple, un classement, une mesure ou un dénombrement.	Appliquer un raisonnement pratique, maîtriser des tâches familières simples propres à la nature du problème ainsi qu'exposé.
2 226–275	Repérer une information unique, dans un texte ou un document synoptique simple comprenant quelques distracteurs. Relier et/ou comparer quelques informations éparses faciles à trouver.	Comprendre les notions mathématiques rudimentaires. Effectuer des calculs avec des pourcents ou des fractions en écartant des distracteurs simples. Interpréter un graphique.	Raisonnement systématiquement et confronter les informations de la question et du stimulus. Évaluer plusieurs solutions satisfaisant des contraintes explicites. Le but est simple et bien défini.
3 276–325	Trier et relier plusieurs éléments d'informations à repérer dans un texte suivi assez long ou plusieurs textes schématiques. Faire une déduction simple. Écarter de nombreux distracteurs. Rédiger une réponse.	Comprendre diverses formes d'informations mathématiques. Interpréter des proportions ou des statistiques insérées dans des textes parsemés de distracteurs. Effectuer des opérations par itération.	Poursuivre un raisonnement par itération. Ordonner des objets selon des critères donnés, déterminer une séquence d'actions en tenant compte de contraintes opaques et/ou interdépendantes. Le but est multiple et partiellement défini.
4/5 326–500	Faire une interprétation correcte d'un texte suivi long et difficile ou de l'examen conjoint de plusieurs textes schématiques compliqués. Faire des déductions de conditions multiples. Tirer parti de connaissances spécialisées. Rédiger plusieurs réponses.	Comprendre des informations mathématiques abstraites enchâssées dans des textes complexes. Effectuer un enchaînement d'opérations compliquées. Appliquer des formules données, éventuellement à produire. Donner des justifications fondées mathématiquement.	Envisager un système complet d'états et de solutions possibles. Juger de l'exhaustivité, de la cohérence, de l'interdépendance des critères. Pouvoir évaluer le problème et l'approche qu'on en fait de manière critique et expliquer le pourquoi et le comment de la solution choisie.

# B – L'évaluation des compétences

## La théorie des réponses aux questions (Item Response Theory IRT)

### Généralités

ALL avait pour objectif de mesurer l'intensité de quelques compétences de base dans un échantillon de population de plusieurs pays. Ce fut réalisé à l'aide de tests ad hoc inspirés des tâches de la vie quotidienne. Mais, pour constituer un instrument de mesure, il fallait que ces tâches virtuelles procèdent d'un dessein, puis, pour exploiter l'instrument, il fallait disposer d'un ensemble de méthodes avérées pour déduire des tâches exécutées des valeurs de compétence. La première phase de ce processus a été exposée dans le chapitre A où nous avons vu comment ont été définis les domaines de tests, opérés le choix des stimuli et des questions. Le chapitre B se penche ici sur la deuxième phase du processus.

Une fois obtenues les données brutes des tests, on les a traités pour en déduire des scores de compétences. Voyons ci-dessous, de manière très simplifiée, quels on été ces traitements qui ont abouti à attribuer aux répondants des scores de compétence.

Un test mobilise diverses ressources qui, par leur combinaison, constituent la compétence propre à réussir le test. Dans le cas de ALL, les tests sont fait de questions et, pour mesurer la compétence sollicitée, ces questions doivent constituer une épreuve, au sens où la probabilité d'y répondre correctement dépend étroitement de la compétence que l'on veut mesurer. Les questions de ALL sont presque toujours des questions ouvertes qui réclament du répondant des réponses écrites, souvent sous forme de phrase<sup>16</sup>. Ces dispositions préservent plutôt bien de l'éventualité d'obtenir une réponse juste par hasard.

Un premier traitement des réponses consiste à les juger justes ou fausses<sup>17</sup>. Reportées dans un tableau, les réponses donnent à voir une multiplicité de «1 (= juste)» et de «- (= faux)» éparpillés. Pour en tirer un enseignement il faut trouver le moyen d'y mettre de l'ordre et, pour être signifiant, cet ordre doit procéder d'une théorie qui rende compte de la confrontation d'une compétence d'intensité variable (selon les individus) à des questions de difficulté variable.

Une telle théorie est connue sous le nom de «Item Response Theory».

Item désigne traditionnellement, en anglais comme en français, une position, ou un article, dans une liste. A propos d'un test, item désigne généralement l'élément de sollicitation qui constitue l'instrument du test. Dans ALL cet élément est toujours une phrase qui appelle une réponse et c'est pourquoi nous préférons ici le terme de question à celui d'item. Nous parlerons alors de «la théorie des réponses aux questions». En quoi consiste cette théorie.

La théorie énonce que le niveau d'un trait latent<sup>18</sup> d'une personne, en l'occurrence l'intensité<sup>19</sup> d'une compétence, peut se déduire des réponses que cette personne fournit aux questions d'un test ad hoc. L'existence de traits, comme variables continues, est donc une hypothèse explicite de la théorie. Ceci signifie en pratique que le plus haut degré de difficulté des questions auxquelles une personne répond correctement peut être tenu pour significatif de l'intensité de sa compétence à répondre. Ce qui fera dire aussi que la compétence et la difficulté des questions appartiennent au même espace unidimen-

<sup>16</sup> A l'exception des tests de résolution de problème qui proposent des listes à choix avec une seule réponse correcte.

<sup>17</sup> Les critères et la procédure de jugement sont très délicats à établir, d'autant plus que les tests sont passés dans plusieurs langues. Les critères ont fait l'objet d'un manuel de codification précis et une permanence d'experts, tout au long de la phase de codification, a examiné et tranché les cas problématiques soulevés par les pays participants. Une fois tranchés, ces cas ont pris force exécutoire et ont été communiqués à tous les participants.

<sup>18</sup> Caractéristique individuelle (compétence, aptitude, habileté, etc.) non observable directement mais que l'on peut appréhender par le biais de ses manifestations.

<sup>19</sup> La littérature générale utilise souvent «niveau». Je renonce ici à ce terme pour éviter la confusion avec les «niveaux de compétence» tels que les organisateurs de ALL les ont voulu et utilisés dans leurs rapports d'étude.

sionnel, ou encore que l'intensité de la compétence et le degré de difficulté des questions prendre place sur une même échelle.

Afin d'illustrer notre propos considérons un exemple simulant d'assez près la mesure des compétences de ALL et plus intuitif que les considérations sur la littératie ou la numératie. Cet exemple utilise l'habileté au tir (arc, arbalète, fusil, peu importe). Cette habileté, variable selon les individus, puise à plusieurs ressources. Elle est le résultat de dispositions innées (l'acuité visuelle, la qualité du schéma corporel, le contrôle moteur, etc.) et d'apprentissages (la patience, la précision des gestes, le contrôle du souffle, etc.). En réponse à l'invite du tir, le tireur mobilise ces ressources en un ensemble complexe pour produire une action au résultat simple et mesurable, l'impact d'un projectile à une certaine distance du centre de la cible. Un concours de tirs consistera à demander aux participants de tirer un nombre de coups assez grand pour que leurs performances moyennes les départagent sans ambiguïté. Le résultat de chaque tireur, son nombre de points par exemple, témoignera sans équivoque de son habileté au tir et d'elle seule. Si le concours consistait à toucher des cibles de différents diamètres il serait analogue aux tests de ALL avec leurs questions de différente difficulté.

*L'identification d'une compétence à la capacité de répondre à un ensemble de questions ad hoc nécessite des précautions, conceptuelles aussi bien que pratiques. Par exemple, chacun a, pense t'il, une plus ou moins bonne mémoire, mais une connaissance plus précise révèle une réalité bien plus complexe constituée d'une pluralité de fonctions mémorielles dispersées dans le cerveau, interagissant et affectées à des tâches de différente nature. Il serait ainsi déraisonnable d'espérer construire un test mesurant la mémoire comme un tout sur une seule échelle. Cependant, dans le cadre de conditions contrôlées on peut faire passer des tests pour mesurer un aspect particulier d'une de ces fonctions, comme la mémorisation à court terme par exemple. Ce n'est pas alors la mémoire dans son ensemble que l'on mesure mais l'une de ses fonctions qui se manifeste dans un contexte particulier, manifestation influencées de plus par des aptitudes comme l'attention et la concentration. La compétence testée sera de mémorisation, liée partiellement à une fonction mémorielle particulière.*

*Une confirmation graphique idéale de la théorie*

Le tableau ci-dessous illustre idéalement les réponses de 10 personnes soumises à un ensemble de 20 questions<sup>20</sup>. Le traitement des réponses a consisté ici à arranger de gauche à droite les questions qui ont reçu le plus de réponses correctes (notées 1), puis, de haut en bas, les personnes qui ont donné le plus de réponses correctes. Si ce traitement abouti à un résultat du genre exposé dans lequel les «1» occupent un espace compact, continu<sup>21</sup>, on est bien en présence d'un test qui répond aux critères de la théorie des réponses aux questions. L'intensité de la compétence et le degré de difficulté des questions se mesurent sur une même échelle.

Tableau 1	Questions de difficulté croissante ➡																					
Fréquences	10	9	9	9	8	8	8	8	8	7	6	5	5	4	3	2	2	1	1	1		
Répondants de compétence décroissante ↓	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	20	
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	17	
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	15	
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	14	
	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	-	13	
	1	1	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	-	13	
	1	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	11	
	1	1	1	1	1	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	9	
	1	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1

<sup>20</sup> Dans la réalité, les personnes interrogées sont beaucoup plus nombreuses que les questions

<sup>21</sup> Appelé schéma de Gutmann

Remarquons que le tableau ci-dessus implique que les questions sont de difficulté si bien calibrée que les répondants les moins performants répondent à une ou deux questions au moins et que les répondants les plus performants peuvent juste répondre aux questions les plus difficiles. Il implique une science consommée de la construction des questions. Si une grande proportion de répondants répondait à toutes les questions ou si une grande proportion de répondants ne répondait à aucune question le test ne serait que partiellement exploitable.

Remarquons encore que de la somme des 1 de chaque ligne on peut déduire une distribution des fréquences de réponses correctes dans l'échantillon et que de la somme des 1 de chaque colonne on peut déduire une distribution des fréquences de réponses correctes dans l'ensemble des questions.

La disposition compacte des réponses implique que chaque question pose un même problème à tous les répondants. Si le tableau comportait des 1 ou des paquets des 1, isolés, non contigus, ils désigneraient des questions hors de l'espace commun à la difficulté et à la compétence. Ces questions seraient non conformes à la théorie, car plus faciles à certain qu'à d'autres, et ceci

indépendamment de leur difficulté intrinsèque. Dans la pratique, de telles questions sont repérées lors des tests préliminaires et éliminées. Dans l'élaboration des tests de compétence de ALL, on a prêté une grande attention à éviter les questions susceptibles d'être plus faciles pour l'une ou l'autre des populations nationales des pays participants, ou pour une sous population de l'un d'eux<sup>22</sup>.

Enfin nous remarquons que si le tableau établit un classement sans ambiguïté des répondants (de compétence croissante) et des questions (par difficulté croissante) il ne permet pas de séparer ces paramètres. C'est-à-dire que la difficulté des questions n'y a de sens que par rapport à l'échantillon des répondants et, réciproquement, que la compétence de ceux-ci n'y a de sens que par rapport à l'échantillon des questions.

#### *Une infirmation graphique de la théorie*

Si l'aisance à répondre aux questions n'est pas le fait d'une seule compétence mais de plusieurs compétences (ou savoirs) aléatoirement distribuées dans l'échantillon, le tableau pourra avoir l'allure ci-dessous.

Tableau 2	Questions																				
Fréquences	6	5	4	4	4	4	4	4	4	3	3	3	3	3	3	3	3	3	2	1	
Répondants	-	1	1	1	-	-	1	1	-	1	1	-	1	1	1	1	-	1	1	-	13
	1	1	1	1	-	1	1	-	1	-	-	-	1	1	-	-	1	1	-	-	11
	1	-	1	-	-	1	-	-	1	1	1	1	-	-	-	1	1	-	-	-	9
	1	1	1	-	1	1	-	-	-	-	-	1	1	-	-	-	-	1	-	-	8
	1	-	-	1	1	-	-	1	-	-	1	-	-	1	-	-	1	-	-	-	7
	1	-	-	-	-	-	-	-	1	1	-	1	-	-	1	-	-	-	-	-	5
	-	1	-	1	1	-	-	1	-	-	-	-	-	-	1	-	-	-	-	-	5
	-	-	-	-	1	-	1	1	-	-	-	-	-	-	-	1	-	-	-	1	5
	1	1	-	-	-	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	4
	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	1	-

Malgré l'arrangement des réponses justes les plus fréquentes de gauche à droite et de bas en haut, le tableau ne révèle aucune structure. D'évidence le test n'entre pas dans le cadre de la théorie.

<sup>22</sup> Une controverse avait d'ailleurs surgi à ce sujet, la présence d'un biais culturel dans les tests, lors de la publication des résultats de IALS

Tableau 3	Question de difficulté croissante →																					
	Fréquences																					
	9	9	9	8	8	7	7	6	6	5	5	4	3	3	3	3	2	2	1	1		
Répondants de compétence décroissante ↓	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	-	1	1	1	-	17	
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	1	1	-	-	1	17
	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	-	1	-	1	-	-	15
	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-	1	-	-	-	-	-	14
	1	1	1	1	1	1	1	-	1	1	-	1	-	-	1	-	1	-	-	-	-	11
	1	1	1	1	1	1	1	1	1	-	1	-	1	-	-	-	-	-	-	-	-	9
	1	1	1	-	1	1	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	7
	1	1	1	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	5
	-	1	-	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	3
	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	2

### De la théorie à la réalité

Dans la réalité, un test conforme (à peu près) à la théorie ne donnera pas vraiment la disposition du tableau 1, mais plutôt quelque chose comme le tableau ci-dessus.

Ce tableau met en évidence une zone dans laquelle les 1 sont prépondérants mais il montre aussi que tous les 1 ne sont pas contigus; une zone d'incertitude, sépare la zone bien marquée des réponses correctes de celle bien marquée des réponses fausses. Philosophiquement, ce flou manifeste l'imperfection inhérente à toute réalité par rapport à l'idéal théorique mais, plus pragmatiquement et au-delà du pur hasard, il manifeste l'incertitude, ou l'imprécision, du classement choisi, aussi bien quant à la compétence précise des répondants que quant à la difficulté précise des questions. Si le classement distingue avec une bonne certitude, pour un répondant donné, les questions faciles des questions difficiles, il n'en va pas de même lorsque on se situe entre deux, aux limites de la compétence du répondant, là où presque rien sépare le succès de l'échec.

Par ailleurs, avec un échantillon important, de nombreux répondants auront la même proportion de réponses correctes tout en présentant des schémas de réponses légèrement différents. Seule une connaissance très précise des difficultés des questions pourrait mieux les départager, c'est ce qu'il va s'agir d'établir si l'on veut connaître les valeurs vraies de l'intensité de la compétence des répondants.

Pour ce faire on va associer à la collection des réponses, correctes et incorrectes (dites données brutes), un modèle mathématique de la relation répondant-question qui va permettre d'estimer les valeurs «vraies» de l'intensité de la compétence de chaque répondant.

Pour rendre compte impeccablement de la compétence et considérer séparément celle-ci de la difficulté des questions le modèle mathématique doit remplir les deux conditions ci-dessous.

1. la probabilité d'une réponse correcte doit pouvoir s'exprimer en fonction de la difficulté des questions et de l'intensité de la compétence à l'aide d'une seule et même relation mathématique pour tout l'échantillon.
2. la réponse à une question donnée est indépendante des réponses aux autres questions (On dit que les réponses sont localement indépendantes).

#### Remarque:

*Il existe pour chaque test une incertitude à propos des performances individuelles des répondants, au même titre que les performances d'un sportif varient selon la «forme du jour». L'application d'un modèle probabiliste de la performance dans le cadre de ALL n'est pas là pour refléter ce phénomène et on peut par ailleurs considérer que, statistiquement, l'effet de la «forme du jour» s'annule dans tout groupe de quelque importance.*

### Le modèle de Rasch et les modèles IRT

Le modèle le plus simple est celui de Rasch<sup>23</sup>. Ce modèle relie la compétence (le trait latent) à la difficulté de la question en exprimant la probabilité d'une réponse correcte en fonction du paramètre (l'intensité de la compétence) du répondant et du paramètre (le degré de difficulté) de la question. C'est un modèle idéal dans le sens qu'il explicite les réponses de façon à en déduire

<sup>23</sup> **Georg Rasch** (1901–1980), mathématicien, statisticien, et psychométricien danois.

des valeurs. La compétence du répondant et la difficulté de la question y sont décrites chacune par un seul paramètre. Ainsi, pour un individu «n» de compétence  $\theta$  donnée confronté à une question «i» de difficulté  $\delta$  donnée, la réponse correcte,  $X_{ni} = 1$ , relève d'une loi de probabilité. Cette probabilité P est décrite par une fonction logistique de la différence entre le paramètre  $\theta_n$  du répondant et le paramètre  $\delta_i$  de la question. La différence sera nulle pour une probabilité de 0,5. En d'autres termes, la compétence du répondant sera mesurée par la difficulté de la question qu'il a 50% de chance de résoudre correctement.

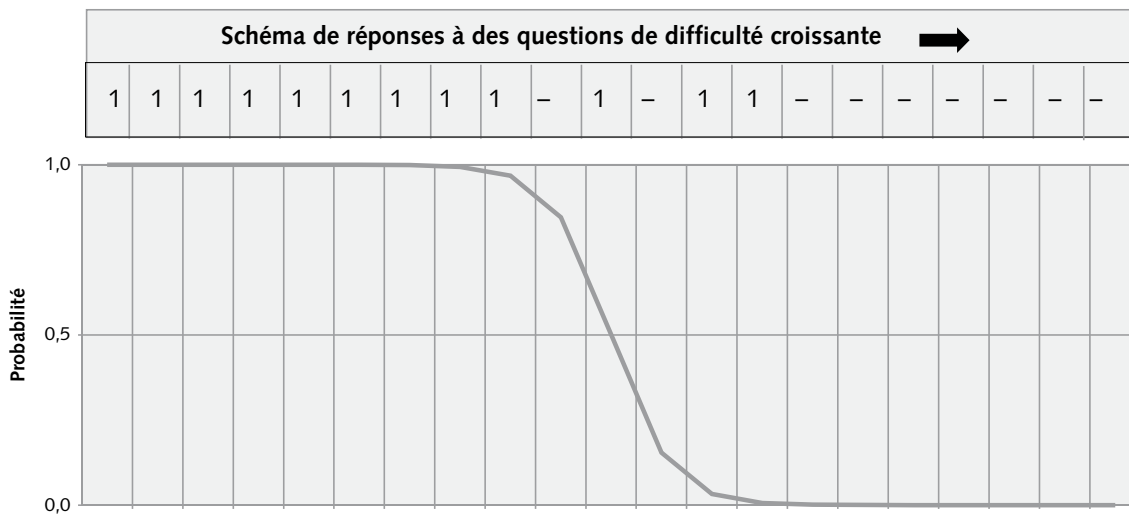
Un premier avantage de ce procédé est la considérable compression des données qu'il effectue puisque qu'il en réduit le nombre de (i X n) à (i + n).

Une coupe horizontale dans le tableau 4 révèle une image du genre de celle ci-dessous. La ligne de «1» et de «-» représente les réponses brutes d'un répondant. Selon l'équation ci-dessus nous avons tracé, au-dessous des réponses, une courbe possible de leur modélisation idéale exprimant la probabilité de réponses correctes.

Modèle de Rasch	
$P(X_{ni} = 1) = \frac{1}{1 + e^{-(\theta_n - \delta_i)}}$	$\theta_n$ = paramètre du répondant «n» $\delta_i$ = paramètre de la question «i»

Schéma observé de réponses d'un répondant donné à des questions de difficulté croissante et courbe de la probabilité de répondre correctement

G 1



© Office fédéral de la statistique (OFS)

Cette courbe montre comment, dans le modèle de Rasch, la probabilité P d'une réponse correcte varie en fonction du paramètre  $\delta$  des questions pour une valeur  $\theta$  donnée du paramètre du répondant. Cette courbe caractérise un répondant.

Inspirés du modèle de Rasch, deux autres modèles sont couramment utilisés. Ainsi les modèles à deux et trois paramètres. Celui dit «à deux paramètres» s'écrit comme ci-dessous.

Le paramètre  $\alpha_i$ , dit «de discrimination», représente la netteté avec laquelle la question «i» distingue les répondants qui ont la capacité potentielle de répondre à la question de ceux qui ne l'ont pas.

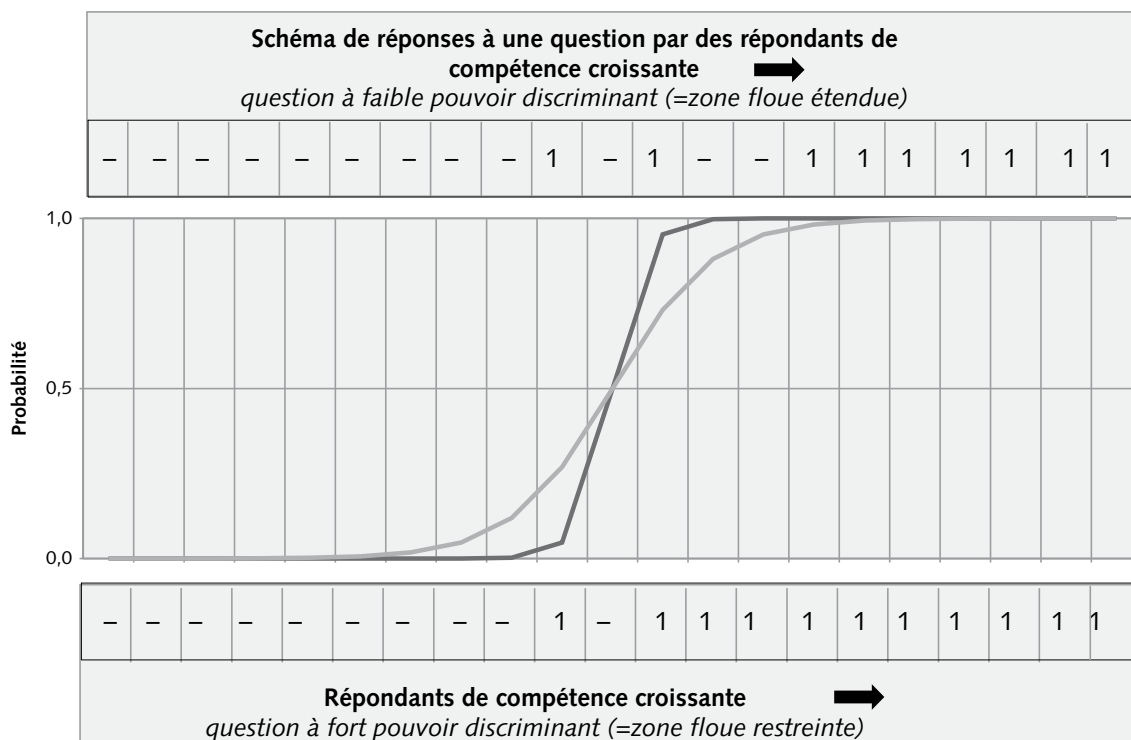
Une coupe verticale dans le tableau 4, de bas en haut, révèle une image du genre de celle ci-dessous. Cette image sera modélisée par la même équation, mais où cette fois P sera fonction de  $\theta$  et  $\alpha$  pour une valeur donnée de  $\delta$ .

Nous avons fait deux coupes, correspondantes à deux questions de même difficulté mais de pouvoirs discriminants différents.

Modèle IRT à deux paramètres	$\theta =$ paramètre du répondant «n»
$P(X_{nj} = 1) = \frac{1}{1 + e^{\alpha_i(\delta_i - \theta_n)}}$	$\delta_i =$ paramètre de la question «i»
	$\alpha_i =$ paramètre de discrimination

**Schéma observé des réponses à une question donnée par des répondants de compétence croissantes et courbes de probabilités représentatives de la question**

G 2



© Office fédéral de la statistique (OFS)

Ces deux courbes caractérisent chacune une question de même difficulté. La moins pentue pour  $\alpha$  petit (faible pouvoir discriminant), la plus pentue pour  $\alpha$  grand (fort pouvoir discriminant).

Le modèle à trois paramètres ajoute au modèle ci-dessus un paramètre qui détermine un seuil minimal de probabilité (une asymptote inférieure) en deçà duquel la production d'une réponse correcte est négligée parce que suspecte d'être due à la chance. Ce modèle serait adapté à l'analyse des résultats du concours de tir. Si nous avons relevé, en début de chapitre, la très faible probabilité d'une réponse correcte par hasard dans ALL, tel ne serait pas le cas d'un coup bien centré de la part d'un mauvais tireur.

ALL a donc utilisé le modèle à deux paramètres.

En résumé, la théorie des réponses aux questions amène à appliquer un modèle mathématique à une collection de données. Le modèle établit une relation explicite entre la difficulté des questions et la capacité des répondants à leur donner des réponses correctes. Les scores d'un répondant sont significatifs non par rapport aux scores d'un groupe témoin, comme dans l'évaluation classique des tests, mais par rapport à la difficulté des questions. Les paramètres des questions ne dépendent pas du paquet de questions choisi et les paramètres des répondants ne dépendent pas de l'échantillon des répondants; ils sont tous deux invariants. Ainsi, si on dispose d'un stock de questions de difficultés connues, il est possible de comparer entre eux tout échantillon de population soumis à ces questions. Réciproquement, à tout répondant de compétence connue on pourra attribuer une probabilité unique de répondre correctement à chacune des questions du stock. Cette dernière remarque est particulièrement importante comme on le verra plus loin.

## L'estimation des paramètres

### Maximum de vraisemblance<sup>24</sup>

Si dans le tableau 3 ci-dessus il n'y avait pas de zone floue on en tirerait un classement sans équivoque des questions et des répondants. On pourrait alors attribuer à chaque répondant un paramètre de compétences égal au classement de la réponse la plus difficile à laquelle il a

répondu. La zone floue, on l'a dit, traduit l'incertitude du classement par fréquence de réponses justes, tant pour les paramètres des questions (le degré de difficulté des questions) que pour celui des répondants (l'intensité de la compétence). Pour lever cette incertitude on applique et substitue aux schémas de réponses un modèle probabiliste IRT qui met en relation ces paramètres. L'estimation des paramètres va permettre de classer avec une grande précision les répondants d'un côté et les questions de l'autre.

Si l'on connaissait les valeurs des paramètres des répondants on pourrait calculer celles des paramètres des questions, et réciproquement. Mais, comme on ne connaît ni les unes ni les autres, on va procéder par approximation successive en commençant par attribuer des valeurs *a priori*<sup>25</sup> aux paramètres des questions pour calculer les valeurs des paramètres des répondants, puis à recalculer les valeurs des paramètres des questions compte tenu des valeurs des paramètres des répondants issus du calcul précédant, puis à enchaîner les cycles successifs jusqu'à ce que les valeurs soient stables d'un cycle à l'autre<sup>26</sup>.

A chaque cycle, l'estimation des paramètres se déduit de la meilleure adaptation possible du modèle au schéma de réponses effectivement obtenues (les observations), ce que l'on peut illustrer par la recherche de la position optimale des courbes de probabilité (voir figures 1 et 2) par rapport aux schémas de réponses. Cette position optimale correspond au maximum de la fonction de vraisemblance, une fonction qui exprime la probabilité de réalisation du schéma observé étant donnée la valeur du paramètre de la question (dans la première phase du cycle), ou du répondant (dans la deuxième phase du cycle). Cette probabilité est le produit des probabilités élémentaires de chacune des réalisations des réponses observées. Le maximum de la fonction sera recherché par approximation en faisant varier le paramètre, soit à, graphiquement, translater la courbe.

La figure 3 ci-dessous illustre de manière intuitive le mécanisme du procédé.

Les paramètres des questions étant connus (ou supposés tels) considérons les réponses d'un répondant donné (une ligne horizontale de notre tableau 3) et exprimons les probabilités réalisées de réponses justes et fausses à l'aide du modèle IRT. Les probabilités réalisées

<sup>24</sup> La discussion qui suit ne considère le traitement que d'un seul domaine de compétence à la fois. On fera plus loin une remarque sur la généralisation du procédé exposé ci-dessous appliqué au traitement simultané de plusieurs paramètres.

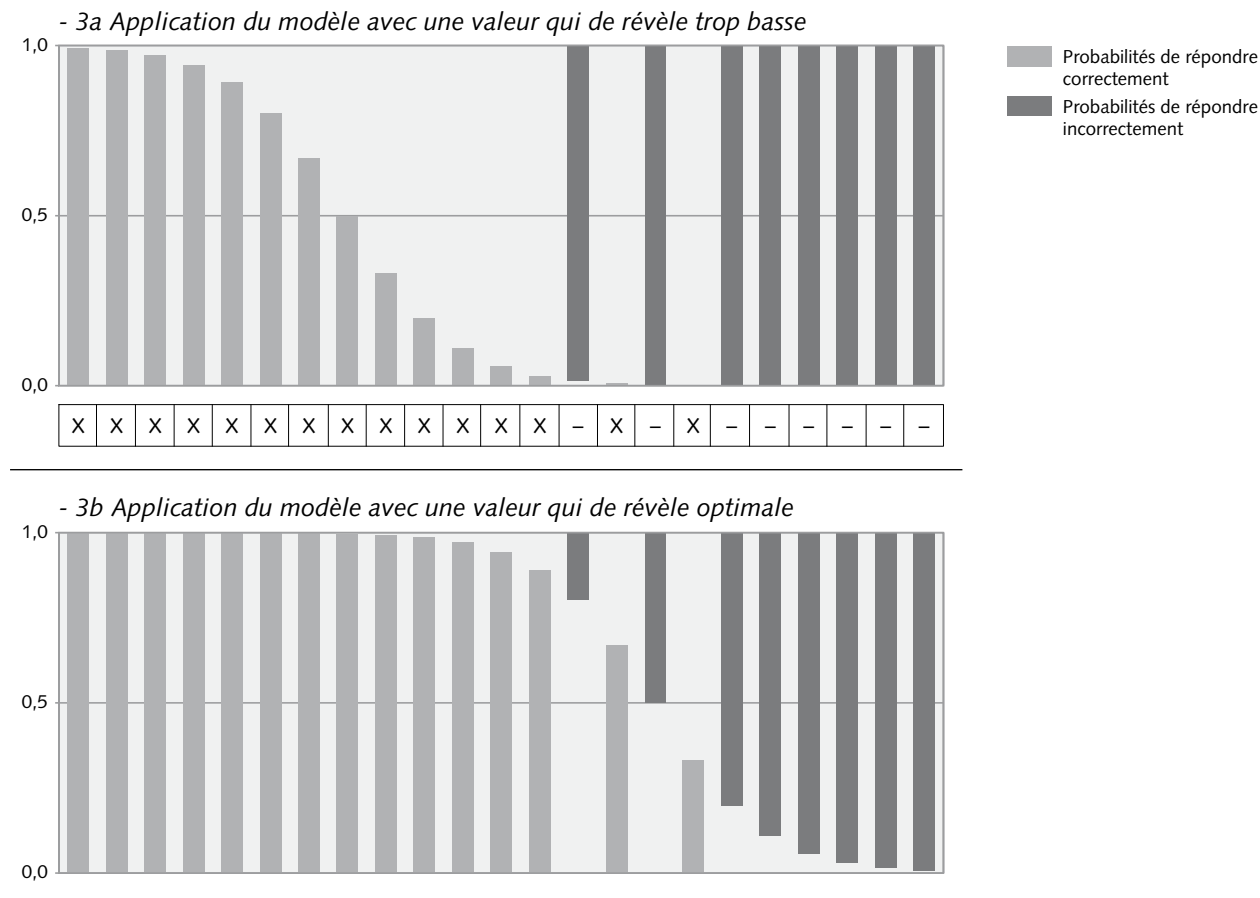
<sup>25</sup> On use des connaissances préalables sur la difficulté des questions pour procéder à cette première attribution.

<sup>26</sup> La méthode converge en effet mais il n'est pas lieu de le démontrer



**Probabilité que le répondant «n» a de répondre conformément au schéma observé si on lui attribue une valeur  $\theta$  de compétence**

G 3



sont représentées par les barres verticales dans le graphique ci-dessous: les bâtons noirs expriment les probabilités de répondre correctement aux questions auxquelles il a été effectivement correctement répondu, les bâtons grisés de répondre incorrectement aux questions qui n'ont effectivement pas reçu de réponses correctes.

Etant donnée un valeur de  $\theta$ , la fonction de vraisemblance, soit la probabilité d'observer le schéma de réponses ci-dessus, est le produit de chacune des probabilités d'observer les réponses effectives ou, graphiquement, le produit de tous les bâtons verticaux (noirs et grisés). L'emprise minimale des espaces vides dans le graphique 3b donne une interprétation visuelle du maximum de la fonction.

La procédure qui attribue des valeurs croissantes de  $\theta$  jusqu'à obtention du maximum de la fonction est dite de la recherche du maximum de vraisemblance (Maximum Likelihood).

**Le procédé de recouvrement des questions**

Les promoteurs d'une enquête souhaitent généralement récolter le plus grand nombre possible d'informations, un vœu que contrecarrent les limites pratiques de la récolte.

Les interviews de ALL ont duré en moyenne une heure et demie et c'est probablement la plus longue durée que l'on puisse imposer à un répondant sans risquer d'effets sensibles de fatigue ou de lassitude, voire un taux important d'abandons. La première moitié de l'interview était consacrée au questionnaire et la deuxième aux tests. Le nombre de questions de tests auquel il était alors possible de soumettre un répondant en trois-quarts d'heure environ était insuffisant à couvrir les quatre domaines. Cependant à tous les répondants ont été attribués des scores de compétences dans tous les quatre domaines<sup>27</sup> et ceci a été possible parce que l'on peut, sous certaines conditions, imputer des valeurs à des données non observées.

<sup>27</sup> Pour tous les domaines traités dans l'échantillon dont est issu le répondant. Il n'a pas été attribué de score en résolution de problèmes aux répondants d'échantillons qui n'ont pas traité ce domaine.

Ces conditions sont les suivantes

1. Il faut que les données non observées (dites plus généralement données manquantes) soient aléatoires, au sens que leur absence ne doit pas dépendre des valeurs qu'elles auraient pris si on avait pu les observer<sup>28</sup>. Dans la littérature les données manquantes de ce type sont dites «MAR» (Missing At Random).
2. Il faut distribuer les questions dans l'échantillon de façon à obtenir un nombre de recouvrements entre questions suffisant pour en déduire des corrélations
3. Il faut appliquer aux données un modèle IRT qui, distinguant la compétence de la difficulté des questions, permet d'établir la probabilité de réponse correcte d'un répondant donné face à une question donnée.

Ces conditions réunies, il devient possible de ne pas poser toutes les questions à tous les répondants. Ce faisant on peut augmenter le nombre total de questions bien au-delà de ce que l'on peut poser à un seul répondant et couvrir ainsi plusieurs domaines, ce qui aurait été autrement impossible<sup>29</sup>. La troisième condition ayant été abordée plus haut, voyons ci-dessous comment ont été remplies les deux premières.

La totalité des questions (ou items) s'est monté à 160<sup>30</sup>: 52 de littératie de textes suivis, 48 de littératie de textes schématiques, 41 de numératie et 19 de résolution de problème. Les 106 questions de littératie ont été réparties dans quatre blocs (Bl. 1 à 4), les 41 questions de numératie dans deux blocs (Bl. 5 et 6) et les 19 questions le résolution de problème dans deux blocs également (Bl. 7 et 8). Combinés deux à deux, les 8 blocs ont constitué les 28 livrets ci-dessous. Chaque répondant a dû répondre à un seul livret et chaque livret a été traité par le même nombre de répondants, l'attribution des livrets se faisant aléatoirement, ce qui satisfait la première condition.

Toutes les combinaisons possibles des blocs deux à deux sont réalisées entre les blocs de littératie (6 livrets), entre les blocs de littératie et de numératie (8 livrets), entre les blocs de littératie et de résolution de problèmes (8 livrets). Ces combinaisons totalisent 22 livrets auxquels s'ajoutent les livrets 5 et 8 qui doublent les livrets 2 et 4 avec inversion de l'ordre des blocs, les livrets 17 et 18 composés des deux blocs de numératie avec inversion des blocs et de même avec les livrets 27 et 28 en résolution de problèmes.

**Répartition des blocs dans les livrets. Littératie = blocs 1 à 4, numératie = blocs 5 et 6, résolution de problèmes = livrets 7 et 8.**

**G 4**

		Livrets																											
Bl		01	02	03	04	05	06	07	08	09	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
1		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
2		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
3		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
4		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
5	Numératie	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
6		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
7	Résolution de problèmes	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	
8		■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	

<sup>28</sup> Elles manquent alors pour une raison précise caractérisant un ou plusieurs sous-groupes.

<sup>29</sup> Au prix toutefois d'une augmentation du nombre d'interviews.

<sup>30</sup> Sans compter les 6 questions élémentaires posées à tous les répondants pour écarter ceux d'entre eux potentiellement incapables de réussir les plus simples des tests de l'épreuve complète.

Tout ceci a pour conséquence que, à l'intérieur d'un échantillon d'environ 2000 répondants (Cas de la Suisse alémanique et de la Suisse romande), chaque question a été posée environ 250 fois, chaque couple de question à l'intérieur d'un domaine 120 fois<sup>31</sup> et chaque couple de question entre littératie et numératie et entre littératie et résolution de problèmes environ 60 fois. Cette configuration des livrets entraîne un recouvrement des questions qui permet une validation croisée de leurs difficultés relatives, à l'intérieur d'un domaine et d'un domaine à l'autre. La deuxième condition est ainsi satisfaite.

### Valeurs manquantes et imputation multiple

Chaque répondant a dû remplir un livret de test et un seul, donc répondre aux questions de deux blocs sur les huit existants. Cependant chaque répondant s'est vu attribuer des scores de performance dans les quatre domaines testés. C'est-à-dire que les carrés blancs du schéma de composition des livrets ont été comblés à l'aide de méthodes d'imputation de données manquantes (voir l'annexe ci-après). L'attribution des livrets aux répondants s'est faite de manière aléatoire, satisfaisant l'hypothèse MAR évoquée plus haut.

Toutes les méthodes qui imputent une seule valeur à chaque donnée manquante ont le défaut d'aboutir à sous-estimer la variance des résultats, parce qu'à la valeur imputée, obligatoirement incertaine, on accorde une même confiance qu'à une valeur observée. Une méthode particulière, l'imputation de valeurs multiples pallie ce défaut, la multiplicité des valeurs reflétant l'incertitude inhérente à l'imputation.

Dans ALL, l'imputation des valeurs multiples s'est basée sur les résultats de l'application de la méthode dite **Marginal Maximum Likelihood**. La procédure a consisté à estimer une distribution des valeurs probables des paramètres des répondants tour à tour avec une distribution des valeurs probables des paramètres des questions en ajustant chaque fois les paramètres des répondants et des questions de façon à obtenir la vraisemblance maximale (dans les quatre domaines simultanément), relativement aux réponses observées et tenant compte des relations entre réponses observées et non observées. Une fois obtenue pour chaque répondant une distribution jugée suffisante de valeurs probables des scores de compétences, il en a été tiré aléatoirement<sup>32</sup> cinq valeurs, dites valeurs plausibles, que l'on a attribué au répondant. C'est de la combinaison de ces cinq valeurs que l'on tirera les résultats d'enquête les plus précis et dont la variance rendra compte de l'incertitude des valeurs imputées.

Il faut enfin remarquer que cette méthode délivre des résultats très fiables à propos de tout groupe de répondants (populations nationales, linguistique, urbaines, etc.) mais que cette fiabilité s'obtient au détriment de celle concernant les individus.

<sup>31</sup> Cela n'est pas tout à fait exact à propos de la littératie. Tout d'abord on ne distingue pas les deux types de littératie et ensuite la moitié des couples de questions sont traités deux fois moins souvent. Cette dissymétrie de traitement est compensée par la plus grande proportion des questions de littératie dans l'ensemble des questions.

<sup>32</sup> C'est là une «règle de l'art statistique», les imputations supplémentaires n'apportant qu'un faible gain de précision au prix d'une augmentation exponentielle du volume de calcul.

# ANNEXE B

## Généralités sur l'imputation de valeurs aux données manquantes

Dans tout relevé, des données manquent, pour toutes sortes de raisons. Une source fait défaut, les gens interrogés ne peuvent ou veulent pas répondre à certaines questions. Ou bien des données ont été perdues en cours de transmission, ou encore doivent être supprimées parce qu'aberrantes, à cause d'interprétations erronées, d'erreurs de relevé ou de transcription, etc. Dans les études longitudinales des observations entières peuvent manquer parce qu'une source d'information a disparu, parce qu'un individu a quitté l'échantillon. L'analyse des données s'accommode difficilement des valeurs manquantes qui sont comme des trous dans le système.

Pour exploiter le relevé incomplet une solution consiste à supprimer tous les enregistrements dont une valeur manque (et donc toutes les valeurs observées de l'enregistrement). Une autre, moins radicale, à ne supprimer que les enregistrements dont la valeur de la variable étudiée est manquante. Dans ce dernier cas la taille de l'échantillon diffère selon la variable étudiée. Toutefois ces deux solutions sont dommageables. Elles diminuent la taille du relevé et déprécient l'ensemble des données en rendant inutiles de nombreuses observations chèrement acquises. Enfin elles introduisent un risque de biais important, car les valeurs manquantes sont souvent le fait d'une sous population caractéristique.

C'est pourquoi l'imputation de valeurs aux données manquantes est payante et légitime dès lors que l'on peut montrer qu'on ne prend qu'un risque raisonnable de fausser la distribution de l'ensemble des données. Bien que les valeurs imputées ne contiennent en soi aucune information il est généralement avantageux de disposer d'un relevé complété, sans trous, apte à se prêter dans son entier aux travaux d'analyses. Enfin, rappelons le, «le comblement des trous» réduit le risque de biais.

La littérature propose plusieurs méthodes d'imputations et il faut se fier aux statisticiens expérimentés qui

savent appliquer les techniques les plus judicieuses, celles qui ont le plus de chance de renforcer la validité des analyses ultérieures.

En résumé, une imputation judicieuse de valeurs aux données manquantes apporte les gains suivants:

- une exploitation de tous les enregistrements, mêmes de ceux incomplets.
- un traitement standard facilité des données
- une réduction vraisemblable des biais.

De nombreuses méthodes d'imputation sont proposées. Entre autres, on peut:

- Imputer des valeurs observées dans d'autres relevés similaires,
- Imputer la moyenne des valeurs observées, dans la totalité de l'échantillon ou dans des sous échantillons spécifiques (Cette méthode peut conduire à sous-estimer la variance parce que les valeurs manquantes sont souvent plus fréquentes aux extrémités du spectre).
- Imputer des valeurs observées dans des enregistrements, du même relevé, de profils très semblables.

Une classe de méthodes, dites déductives, admet que l'ensemble des données est globalement redondant, c'est-à-dire que certaines données y sont interdépendantes.

Les méthodes déductives donnent des résultats valides à la condition que les valeurs manquantes soient distribuées aléatoirement dans la collection des données. Les valeurs manquantes sont alors dites «MAR (Missing At Random)». C'est-à-dire que leur probabilité d'absence est indépendante des valeurs qui devraient être observées.

Il faut remarquer que cette hypothèse est impossible à valider, puisqu'il faudrait pour cela connaître les valeurs manquantes. Il est un cas cependant où l'hypothèse n'est pas à valider, c'est lorsque les données manquantes sont planifiées, et planifiées de telle façon à ce qu'elle soient effectivement distribuées aléatoirement. C'est le cas de ALL.

# Bibliographie

- Murray, Scott T., Clermont, Yvan et Binkley, Marilyn, (2005). *Measuring Adult Literacy and Life Skills: New Frameworks for Assessment*. Statistics Canada, Ottawa
- Yamamoto Kentaro et Kirsch Irwin S. (1998). *Scaling and Scale Linking, Proficiency Estimation*. In Adult Literacy in OECD Countries, Technical Report on the First International Adult Literacy Survey, 161–190, U.S. Department of Education, Office of Educational Research and Improvement, Washington.
- Philipp Notter et all. (2006). *Lire et calculer au quotidien, compétences des adultes en Suisse*. Office fédéral de la statistique, Neuchâtel.
- Rychen, Dominique Simone et Salganik, Laura. (2003). *Key Competencies for a Successful Life and a Well-Functioning Society*. Hogrefe & Huber, Göttingen.
- Embretson, Susan E. et Reise, Steven P. (2000). *Item Response Theory for Psychologists*. Laurence Erlbaum Associates, London.
- Schafer, Joseph L. et Olsen, Maren K. (1998). *Multiple Imputation for Multivariate Missing-Data Problems: A Data Analyst's Perspective*. In *Multivariate Behavioral Research* 33(4), 545–571, Laurence Erlbaum Associates. Inc.



# Programme des publications de l'OFS

En sa qualité de service central de statistique de la Confédération, l'Office fédéral de la statistique (OFS) a pour tâche de rendre les informations statistiques accessibles à un large public.

L'information statistique est diffusée par domaine (cf. verso de la première page de couverture); elle emprunte diverses voies:

<i>Moyen de diffusion</i>	<i>Contact</i>
Service de renseignements individuels	032 713 60 11 info@bfs.admin.ch
L'OFS sur Internet	www.statistique.admin.ch
Communiqués de presse: information rapide concernant les résultats les plus récents	www.news-stat.admin.ch
Publications: information approfondie (certaines sont disponibles sur disquette/CD-Rom)	032 713 60 60 order@bfs.admin.ch
Banque de données (accessible en ligne)	032 713 60 86 www.statweb.admin.ch

Informations sur les divers moyens de diffusion sur Internet à l'adresse [www.statistique.admin.ch](http://www.statistique.admin.ch) → Services → Les publications de Statistique suisse.

## Education et science

Dans le domaine de l'éducation et de la science, trois sections de l'Office fédéral de la statistique traitent les thèmes suivants:

### **Section Systèmes d'éducation et science (BWT)**

- Système d'éducation (indicateurs du système de la formation)
- Formation et marché du travail (compétences des adultes, transition de l'éducation vers le marché du travail, indicateurs de la formation professionnelle, formation continue)
- Hautes écoles (indicateurs des hautes écoles, situation sociale des étudiants)

### **Section Formation scolaire et professionnelle (SCHUL)**

- Elèves et diplômés (élèves et étudiants, formation professionnelle et examens finals)
- Ressources et infrastructure (enseignants, finances et coûts, écoles)
- PISA (mesure des compétences des jeunes de 15 ans)

### **Section Hautes écoles (HSW)**

- Etudiants et diplômés des hautes écoles (universitaires et spécialisées)
- Personnel et finances des hautes écoles (universitaires et spécialisées)
- Perspectives de la formation (Elèves, étudiants, diplômés et corps enseignant de tous les niveaux de la formation)

Ces trois sections diffusent des publications régulières et des études thématiques. Nous vous invitons à consulter notre site Internet. Vous y trouverez également des informations sur les personnes de contact pour vos éventuelles questions.

[www.education-stat.admin.ch](http://www.education-stat.admin.ch)

La présente publication est destinée aux profanes curieux du contenu de l'enquête Adult Literacy and Lifeskills – ALL. Un premier chapitre présente une synthèse des domaines de mesure de l'enquête avec le principe de leurs tests. Un deuxième chapitre expose les rudiments des principes de psychométrie qui fondent le traitement des réponses aux tests aboutissant à l'attribution, à chaque personne soumise aux tests, de cinq valeurs de performances (dites valeurs plausibles) dans les quatre domaines de compétence

**N° de commande**

938-0800-05

**Commandes**

Tél.: 032 713 60 60

Fax: 032 713 60 61

E-mail: [order@bfs.admin.ch](mailto:order@bfs.admin.ch)

**Prix**

6 francs (TVA excl.), impression à la demande

ISBN 978-3-303-15446-5